ZEROS OF PRINCIPAL L-FUNCTIONS AND RANDOM MATRIX THEORY

ZEÉV RUDNICK AND PETER SARNAK

1. Introduction. Our goal in this paper is to study the distribution of zeros of the Riemann zeta function as well as of more general L-functions.[1] According to conjectures of Langlands [14], the most general L-function is that attached to an automorphic representation of GL_N over a number field, and these in turn should be expressible as products of the "standard" L-functions $L(s, \pi)$ attached to *cuspidal* automorphic representations of GL_m over the *rationals*. Such L-functions are therefore believed to be the building blocks for general L-functions, and we call them (principal) *primitive L-functions* of degree m. (They do not factor as products of such L-functions.)[2] For $m = 1$ these are the Riemann zeta function $\zeta(s)$ and Dirichlet L-functions $L(s, \chi)$ with χ primitive. For $m = 2$ the analytic properties and functional equation of such L-functions were investigated by Hecke and Maass, and for $m \geqslant 3$ by Godement and Jacquet [5]. We are interested in the fine structure of the distribution of the nontrivial zeros of such primitive $L(s, \pi)$. Let $\rho^{(\pi)} = (1/2) + i\gamma^{(\pi)}$ denote these zeros. To motivate the formulation of our results, we begin by assuming the Riemann hypothesis (RH) for $L(s, \pi)$, that is, that $\gamma^{(\pi)} \in \mathbf{R}$. We order the $\gamma^{(\pi)}$'s (with multiplicities)

$$\cdots \leqslant \gamma_{-2}^{(\pi)} \leqslant \gamma_{-1}^{(\pi)} < 0 \leqslant \gamma_1^{(\pi)} \leqslant \gamma_2^{(\pi)} \cdots .$$

The number of γ's in an interval $[T, T + 1]$ is asymptotic to $(m/2\pi) \log T$ as $T \to \infty$ (see (2.11)). It follows that the numbers $\tilde{\gamma}_j^{(\pi)} = (m/2\pi)\gamma_j \log|\gamma_j|$ have unit mean spacing. The problem is to understand the statistical nature of the sequence $\tilde{\gamma}_j^{(\pi)}$: Do they come down randomly (Poisson process) or do they follow a more revealing distribution?

In the case of the Riemann zeta function, following the original calculation by Montgomery [20] of the pair correlation (see below) and the extensive numerical calculations of Odlyzko [21], [22], it is now well accepted (but far from proven) that the consecutive spacings follow the Gaussian unitary ensemble (GUE) distribution from random matrix theory. That is, if $\delta_n = \tilde{\gamma}_{n+1} - \tilde{\gamma}_n$ are the normalized

[1] The reader interested only in the Riemann zeta function $\zeta(s)$ should read the paper with $L(s, \pi)$ replaced by $\zeta(s)$ and $m = 1$ everywhere, in which case the results were announced in [26].

[2] It is quite plausible that these coincide with the primitive Dirichlet series introduced by Selberg [29] or the "arithmetic Dirichlet series" in Piatetski-Shapiro [24].

Received 16 September 1994. Revision received 29 June 1995.

Authors partially supported by National Science Foundation grants DMS-9400163 and DMS-9102082.

spacings, then for any nice function on $(0, \infty)$, one expects

$$\frac{1}{N}\sum_{n\leqslant N} f(\delta_n) \to \int_0^\infty f(s)P(s)\,ds \tag{1.1}$$

where $P(s)$ is the distribution of consecutive spacings of the eigenvalues of a large random Hermitian matrix. This distribution was determined by Gaudin and Mehta [18] and is given as follows:

$$P(s) = \frac{d^2E}{ds^2}(s)$$

where $E(s) = \det(I - Q_s)$, Q_s being the trace class operator on $L^2(-1, 1)$ whose kernel is

$$Q_s(\xi, \eta) = \frac{\sin \pi s(\xi - \eta)/2}{\pi(\xi - \eta)}.$$

The Fredholm determinant defining $E(s)$ converges, and it is easy to check that $P(s)$ vanishes to second order at $s = 0$. This means that, unlike a Poisson process, the numbers $\tilde{\gamma}_n$ tend to "repel" each other (this is often referred to as "level repulsion"). The graph of $P(s)$ and its comparison with Odlyzko's computation for zeros near the 10^{20}th are depicted in Figure 1.

For Dirichlet L-functions, the picture is similar, and Rumely [27] has carried out analogous numerical experiments. Recently, Hejhal [6] has succeeded in computing the three-level correlation function for zeros of $\zeta(s)$, assuming RH and in a restricted range, similar to the assumption and restriction used by Montgomery.

The consecutive level spacing distribution is determined by the n-level correlation functions for all $n \geqslant 2$ [28]. The main result of this paper is the computation of the general n-level correlation function for the zeros of a primitive principal L-function (also in a restricted range). We show that the answer is universal and is precisely the one predicted by Dyson's computations for the GUE model [3]. To define the n-level correlations, suppose that as above we have a set B_N of N numbers $\tilde{\gamma}_1 \leqslant \cdots \leqslant \tilde{\gamma}_N$. The n-level correlation function measures the correlation between differences of n elements of B_N. That is, for a box $Q \subset \mathbf{R}^{n-1}$, set

$$R_n(B_N, Q) = \frac{1}{N}\#\{j_1, \ldots, j_n \leqslant N \text{ distinct}: (\tilde{\gamma}_{j_1} - \tilde{\gamma}_{j_2}, \ldots, \tilde{\gamma}_{j_{n-1}} - \tilde{\gamma}_{j_n}) \in Q\}. \tag{1.2}$$

A technically more convenient way to measure this distribution is to use smooth test functions $f(x_1, \ldots, x_n)$ satisfying the following.

Condition **TF 1**. *$f(x_1, \ldots, x_n)$ is symmetric.*

Condition **TF 2**. *$f(x + t(1, \ldots, 1)) = f(x)$ for $t \in \mathbf{R}$.*

Condition **TF 3**. *$f(x) \to 0$ rapidly as $|x| \to \infty$ in the hyperplane $\sum_j x_j = 0$.*

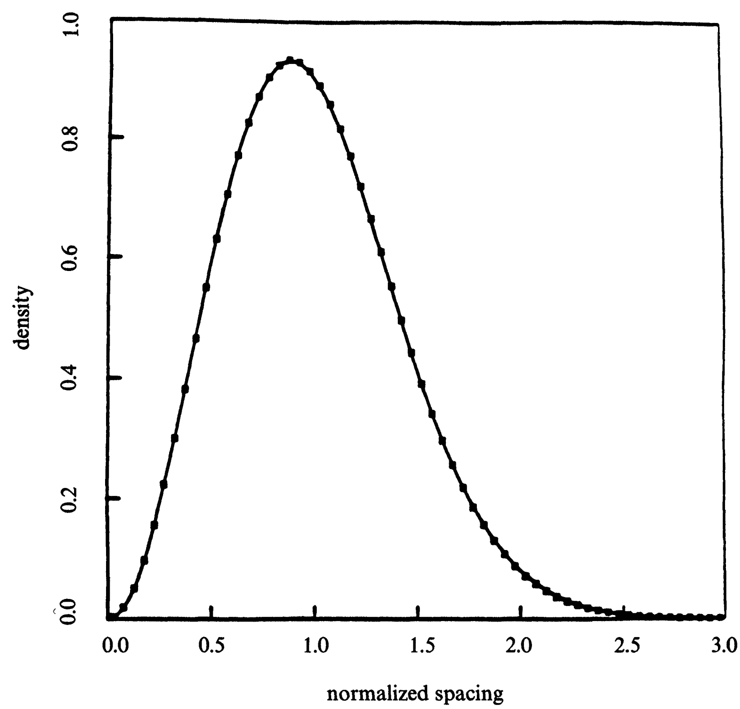

FIGURE 1. Nearest neighbor spacings of zeros of $\zeta(s)$

Probability density of the normalized spacings δ_n. Solid lines: GUE prediction. Scatterplot: empirical data based on 78,893,234 zeros near zero number 10^{20}. Reprinted from Odlyzko [21].

The n-level correlation sum $R_n(B_N, f)$ is defined by

$$R_n(B_N, f) = \frac{n!}{N} \sum_{\substack{S \subset B_N \\ |S| = n}} f(S). \tag{1.3}$$

Here $f(S) = f(a_1, \ldots, a_n)$ if $S = \{a_1, \ldots, a_n\}$. Since f is symmetric, this is well defined. Condition **TF 2** asserts that f is a function of the successive differences so that we recover what (1.2) seeks to measure. Condition **TF 2**, together with the localization **TF 3**, means that we can think of $R_n(B_N, f)$ as counting clusters of size n in B_N. It turns out that knowing the asymptotic behaviour of $R_n(B_N, f)$ as $N \to \infty$ is equivalent to knowing that of the smoothed correlations

$$R_n(T, f, h) = \sum_{j_1, \ldots, j_n}' h\left(\frac{\gamma_{j_1}}{T}\right) \cdots h\left(\frac{\gamma_{j_n}}{T}\right) f\left(\frac{L}{2\pi}\gamma_{j_1}, \ldots, \frac{L}{2\pi}\gamma_{j_n}\right) \tag{1.4}$$

for a sufficiently rich family of localized cutoff functions h (e.g., of rapid decrease). Here $L = m \log T$, and $\sum'$ means sum over distinct indices. Note that since h localizes γ to be of order T, the normalization $(L/2\pi)\gamma$ is the same as $\tilde{\gamma}$.

As mentioned above, Dyson [3] determined the limiting n-level correlation density $W_n(x_1, \dots, x_n)$ for the GUE model. He showed it is given by

$$W_n(x_1, \dots, x_n) = \det(K(x_i - x_j)), \qquad K(x) = \frac{\sin \pi x}{\pi x}. \tag{1.5}$$

$W_n(x)$ is a density (though not a probability density) satisfying $0 \leqslant W_n(x) \leqslant 1$ with $W_n(x) = 0$ if and only if $x_i = x_j$ for some $i \neq j$, and $W_n(x) = 1$ if and only if $x_i - x_j \in \mathbf{Z}$ and $x_i \neq x_j$ for all $i \neq j$.

Before stating our results, we need a technical hypothesis concerning the coefficients of $L(s, \pi)$. For $\mathrm{Re}(s)$ large, write

$$\frac{L'}{L}(s, \pi) = -\sum_{n=1}^{\infty} \frac{\Lambda(n) a_\pi(n)}{n^s} \tag{1.6}$$

where $\Lambda(n) = \log p$ if $n = p^e$ is a prime power, and is zero otherwise. The hypothesis asserts that for any $k \geqslant 2$,

$$\sum_p \frac{|a_\pi(p^k) \log p|^2}{p^k} < \infty. \tag{1.7}$$

This is a very mild hypothesis. Firstly, the general "Ramanujan conjectures" for cusp forms on GL_m asserts that $|a_\pi(p^k)| \leqslant m$, which yields (1.7) with a lot to spare. Secondly, we show in Section 2 that (1.7) is valid for $m \leqslant 3$.

Returning to the n-level correlations, we note that if h and f are defined for complex argument and are localized, then the sums (1.4) make sense even if we do not assume RH, and we still refer to these as the n-level correlations. As explained above, RH and the GUE model (if it applies) can be used to predict their asymptotic behaviour as $T \to \infty$. Our first result proves that this prediction is correct at least for a restricted class of f's.

THEOREM 1.1. *Let π be a cuspidal automorphic representation of $GL_m/\mathbf{Q}$. Assume $m \leqslant 3$ or the hypothesis (1.7). Let f satisfy* **TF 1, 2, 3** *and in addition assume that $\hat{f}(\xi)$ is supported in $\sum_j |\xi_j| < 2/m$. Let $g \in C_c^\infty(\mathbf{R})$ and $h(r) = \int_{-\infty}^{\infty} g(u) e^{iru}\, du$ (so that h and f are entire). Then as $T \to \infty$,*

$$R_n(T, f, h) \sim \frac{m}{2\pi} T \log T \int_{-\infty}^{\infty} h(r)^n\, dr \int_{\mathbf{R}^n} f(x) W_n(x) \delta\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n$$

where $\delta(x)$ is the Dirac mass at zero.

If we assume RH for $L(s, \pi)$, we can relax the smoothness condition on h and in fact choose it to be the characteristic function of an interval. In this way we can prove that the n-level correlations of the zeros are GUE at least for f's with restricted Fourier transforms. Precisely, we deduce the following from Theorem 1.1.

THEOREM 1.2. *With the assumptions of Theorem 1.1 and also RH for $L(s, \pi)$,*

$$R_n(B_N, f) \to \int_{\mathbf{R}^n} f(x) W_n(x) \delta\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n$$

as $N \to \infty$.

Remark 1. The restriction $\sum_j |\xi_j| < 2/m$ is a natural one when $m = 1$, since in this case it is exactly the region in which the asymptotic behaviour of $R_n(B_N, f)$ is dominated by the contributions from all the multidiagonals (see Section 3). Beyond this region, a saturation takes effect and the diagonals no longer dominate. For $\zeta(s)$, this region is also distinguished by being the range in which the pole at $s = 1$ contributes only terms of lower order to $R_n(B_N, f)$. In the case of $\zeta(s)$ and $n = 2$, Theorem 1.2 coincides with the result of Montgomery [20]. For $m > 1$, the restriction $\sum_{j=1}^n |\xi_j| < 2/m$ is no longer natural in the sense that we expect the diagonals to continue being the dominant terms as long as $\sum_{j=1}^n |\xi_j| < 2$. The difference and heuristic reasoning leading to this is given at the end of Section 3. In all cases we conjecture the complete universality of the n-level correlations—that is to say that Theorems 1.1 and 1.2 hold without any restrictions on the support of $\hat{f}$. It would be very interesting to check numerically the level spacing distribution for the various types of primitive L-functions of degree $m = 2$ (e.g., of CM type, general type, holomorphic, and nonholomorphic).

Remark 2. The condition that $L(s, \pi)$ be primitive (i.e., coming from a cuspidal π over $\mathbf{Q}$) is crucial. Firstly, if, for example, we look at $L(s) = \zeta(s)^2$, then clearly the distribution of the zeros of $L(s)$ will be GUE with multiplicity two. However, also in the case $L(s) = L(s, \pi_1)L(s, \pi_2)$, with $\pi_1 \neq \pi_2$ (e.g., the Dedekind zeta function of a quadratic extension of $\mathbf{Q}$), the distribution will not be GUE. The reason is that one can by these methods easily see that the zeros of distinct primitive L-functions are *uncorrelated*—so to speak are unaware of each others' existence. As a consequence, the zeros of $L(s)$ will not exhibit the "level repulsion" characteristic of the GUE distribution. Indeed, the natural conjecture here is that the zeros of $L(s) = L(s, \pi_1)L(s, \pi_2)$ will follow the distribution of the *superposition* of two GUEs [18]. This clarifies the role played by the primitive L-functions in understanding the distribution of zeros of the general L-function.

Remark 3. The universality (in π) of the distribution of zeros of $L(s, \pi)$ is somewhat surprising, the reason being that the distribution of the coefficients $a_\pi(p)$ in (1.6), as p runs over primes, is not universal. For example, for degree-two primitive L-functions, there are two conjectured possible limiting distributions for the $a_\pi(p)$'s: Sato-Tate or uniform distribution (with a Dirac mass term) [30]. As the degree increases, the number of possible limit distributions increases rapidly. However, it is a consequence of the theory of the Rankin-Selberg L-functions (developed by Jacquet, Piatetski-Shapiro, and Shalika [8] for $m \geqslant 3$) that all these limiting distributions have the same second moment (at least under hypothesis (1.7)). It is the universality of the second moment that is eventually responsible for the universality in Theorems 1.1 and 1.2. For the case of pair correlation ($n = 2$), this is reasonably evident; for $n > 2$ it was (at least for us) unexpected, and it has its roots in a key feature of "diagonal pairings" that emerges as the main term in the asymptotics of $R_n(T, f, h)$ (see Section 3).

To end the introduction, we outline our proof of Theorems 1.1 and 1.2. As in [20] (and indeed in all work on zeros of L-functions), we use some version of Riemann's "explicit formula" relating sums over zeros to sums over primes. One technical novelty lies in our means of using the explicit formula, which results in an integrated and smoothed version in Theorem 1.1. This allows us to avoid appealing to RH and also considerably facilitates the computation of the n-level correlation functions. Theorem 1.2 is easily recovered from the smooth version. The advantage is that the multidimensional sums over primes that arise can be analyzed by rather "soft" means (e.g., no large sieve inequalities are needed). What emerges as the main term are contributions from diagonal pairs. The combinatorics relating this to what is predicted by GUE, viz. the determinant (1.5), are nontrivial; after all, at some point we have to see this determinant emerge from the theory of primes. The marriage comes from the pairing structure mentioned earlier (which in turn stems from unique factorization) and the cycle structure of the determinant (1.5)—this being encoded in the identity of Theorem 4.1 and Proposition 4.3. This combinatorial analysis is described in Section 4. A crucial ingredient is the combinatorial method of Spitzer [34]. In Section 2 we collect various facts about principal L-functions, local factors, and Rankin-Selberg L-functions, as well as Ramanujan-type bounds that will be needed. The use of the explicit formula to convert the problem to sums over primes is carried out in Section 3. The appendix contains the calculations of the Rankin-Selberg local factors at the ramified places.

2. Background on L-functions

2.1. Principal L-functions. This section is devoted to reviewing some more or less standard facts about automorphic L-functions on GL_m. Our emphasis, in places, will be on the higher-rank theory ($m \geqslant 3$), the results for Dirichlet L-functions being well known and the case $m = 2$ being by now also classical. For definitions and proofs of various statements below, see Jacquet's article [7]. We have also included an appendix in which proofs are given of some facts that we could not find either explicitly stated or proved in the literature.

Let $\pi = \otimes_p \pi_p$ be an irreducible cuspidal automorphic representation of $GL_m/\mathbf{Q}$.

For normalization purposes, we assume that π is unitary, by which we mean that the central character ω_π of π is unitary. To π, one associates an Euler product $L(s, \pi) = \prod_p L(s, \pi_p)$ given by a product of local factors. Outside of a finite set of primes S_π, π_p is unramified and we can associate to π_p a semisimple conjugacy class $\{A_\pi(p)\} \in GL_m(\mathbf{C})$. Such a conjugacy class is parametrized by its eigenvalues $\alpha_\pi(j, p)$, $j = 1, \ldots, m$. The local factors $L(s, \pi_p)$ for the unramified primes are given by

$$L(s, \pi_p) = \det(I - p^{-s}A_\pi(p))^{-1} = \prod_{j=1}^{m} (1 - \alpha_\pi(j, p)p^{-s})^{-1}. \tag{2.1}$$

At the ramified finite primes, the local factors are best described by the Langlands parameters of π_p (see the appendix). They are of the form $L(s, \pi_p) = P_p(p^{-s})^{-1}$, where $P_p(x)$ is a polynomial of degree at most m, and $P_p(0) = 1$. We will find it convenient in this case, too, to write the local factors in the form (2.1), with the convention that we now allow some of the α's to be zero.

The local constituents π_p of a cuspidal π as above are generic [23], [33]. Using local methods, Jacquet and Shalika [10] show that a generic π_p satisfies

$$|\alpha_\pi(j, p)| < p^{1/2}. \tag{2.2}$$

The general "Ramanujan conjectures" for cuspidal automorphic π on GL_m assert that for p unramified, $|\alpha_\pi(j, p)| = 1$. This is known for certain π (e.g., on GL_2 corresponding to holomorphic forms, due to Deligne) but certainly not in general. We will derive a slightly sharper estimate than (2.2) for all $p < \infty$. In the appendix it is shown by a well-known global argument that for any $p < \infty$,

$$|\alpha_\pi(j, p)| \leqslant p^{(1/2)-1/(m^2+1)}. \tag{2.3}$$

There is also an archimedean local factor $L(s, \pi_\infty)$. Again, it is best described in terms of the Langlands parameters of π_∞ (see the appendix). For now it suffices to note that $L(s, \pi_\infty)$ can be written as a product of m Gamma factors:

$$L(s, \pi_\infty) = \prod_{j=1}^{m} \Gamma_{\mathbf{R}}(s + \mu_\pi(j)) \tag{2.4}$$

where $\Gamma_{\mathbf{R}}(s) = \pi^{-s/2}\Gamma(s/2)$ and $\{\mu_\pi(j)\}$ is a set of m numbers associated to π_∞. They satisfy the analogue of (2.2),

$$\operatorname{Re}(\mu_\pi(j)) > -\frac{1}{2}. \tag{2.5}$$

We refer to the appendix for a discussion of (2.5) and of the analogue of (2.3).

2.2. The functional equation. With all the local factors defined, we can turn to the functional equation. Firstly, from (2.2) it is clear that

$$L(s, \pi) = \prod_{p<\infty} L(s, \pi_p) \tag{2.6}$$

converges absolutely, at least for Re $s > 3/2$. Set

$$\Phi(s, \pi) = L(s, \pi_\infty)L(s, \pi). \tag{2.7}$$

Associated to π is its *contragredient* $\tilde{\pi}$, which is itself an irreducible cuspidal automorphic representation. For any $p \leqslant \infty$, $\tilde{\pi}_p$ is equivalent to the complex conjugate $\overline{\pi_p}$ [4], and hence

$$\begin{gathered} \{\alpha_{\tilde{\pi}}(j, p)\} = \{\overline{\alpha_\pi(k, p)}\} \\ \{\mu_{\tilde{\pi}}(j)\} = \{\overline{\mu_\pi(j)}\}. \end{gathered} \tag{2.8}$$

The basic analytic result, proven by Godement-Jacquet [5], [7] is that $\Phi(s, \pi)$ extends to an entire function (except in the case of $\zeta(s)$, which has a simple pole at $s = 1$). Moreover, $\Phi(s, \pi)$ is bounded in vertical strips and satisfies a functional equation

$$\begin{gathered} \Phi(s, \pi) = \varepsilon(s, \pi)\Phi(1 - s, \tilde{\pi}) \\ \varepsilon(s, \pi) = \tau(\pi)Q_\pi^{-s} \end{gathered} \tag{2.9}$$

where $Q_\pi > 0$ is the conductor of π. It is a positive integer with prime factors in S_π [9], and $\tau(\pi) \in \mathbf{C}^*$. We note that $Q_{\tilde{\pi}} = Q_\pi$ and $\tau(\pi)\tau(\tilde{\pi}) = Q_\pi$.

The zeros of $\Phi(s, \pi)$ will be denoted by ρ_π, and by definition are the "nontrivial" zeros of $L(s, \pi)$. The nontrivial zeros of $L(s, \tilde{\pi})$ are related to those of $L(s, \pi)$ via $s \mapsto 1 - s$. The analogue of the Riemann hypothesis for $L(s, \pi)$ is that $\operatorname{Re}(\rho_\pi) = 1/2$. Inasmuch as $\Phi(s, \pi)$ is of order one and the real parts of the zeros are constrained to lie in a strip, it follows that the counting function

$$N_\pi(T) := \#\{\rho_\pi : |\operatorname{Im} \rho_\pi| < T\} \tag{2.10}$$

satisfies $N(T) = O(T^{1+\varepsilon})$ for all $\varepsilon > 0$. A standard winding number argument [2] shows that the Gamma factors in $L(s, \pi_\infty)$ control the number of zeros; in fact,

$$N_\pi(T) \sim \frac{m}{\pi} T \log T. \tag{2.11}$$

2.3. An explicit formula. For Re $s > 3/2$, we may take the logarithmic derivative of (2.6). This yields

$$\frac{L'}{L}(s, \pi) = -\sum_{n=1}^{\infty} \frac{\Lambda(n)a_\pi(n)}{n^s} \tag{2.12}$$

where $\Lambda(n) = \log p$ if $n = p^k$ and zero otherwise, while

$$a_\pi(p^k) = \sum_{j=1}^{m} \alpha_\pi(p, j)^k. \tag{2.13}$$

Note that

$$a_{\tilde{\pi}}(n) = \overline{a_\pi(n)}. \tag{2.14}$$

It will be convenient to set

$$c_\pi(n) = \Lambda(n) a_\pi(n). \tag{2.15}$$

We recast the information in the Euler product and functional equation in terms of an explicit relation between the zeros ρ_π and the $a_\pi(p^k)$. Such relations go by the name of "explicit formulae"; the one we use is a smooth version of Riemann's original formula [25].

PROPOSITION 2.1 (The explicit formula). *Let $g \in C_c^\infty(\mathbf{R})$ be a smooth compactly supported function, and let $h(r) = \int_{-\infty}^{\infty} g(u) e^{iru}\,du$. Write $\rho_\pi = 1/2 + i\gamma_\pi$. Then*

$$\begin{aligned} \sum h(\gamma_\pi) = {} & \delta(\pi)\left\{h\left(-\frac{i}{2}\right) + h\left(\frac{i}{2}\right)\right\} \\ & + \frac{1}{2\pi}\int_{-\infty}^{\infty} h(r)\left(\log Q_\pi + \sum_{j=1}^{m} \frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + \mu_\pi(j) + ir\right)\right. \\ & \left.\left. + \frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + \overline{\mu_\pi(j)} - ir\right)\right)\right) dr \\ & - \sum_{n=1}^{\infty}\left(\frac{c_\pi(n)}{\sqrt{n}} g(\log n) + \frac{\overline{c_\pi(n)}}{\sqrt{n}} g(-\log n)\right) \end{aligned} \tag{2.16}$$

where $\delta(\pi) = 1$ if π corresponds to $\zeta(s)$, and is zero otherwise.

Proof. Set $H(s) := h((s - 1/2)/i)$, and consider the integral

$$\mathscr{I} = \frac{1}{2\pi i}\int_{\operatorname{Re} s = 2} \frac{\Phi'}{\Phi}(s, \pi) H(s)\,ds. \tag{2.17}$$

Now $H(s)$ is rapidly decreasing in $\operatorname{Im} s$ and is entire, so that the integral converges absolutely, and all contour shifts below are legitimate. Φ'/Φ has simple poles at the zeros of $\Phi(s, \pi)$ with residues the multiplicity of the zero (and in the case of $\zeta(s)$ a simple pole with residue -1 at the poles $s = 0, 1$). Shifting the contour in (2.17) to $\operatorname{Re} s = -1$, we have

$$\mathscr{I} = -\delta(\pi)\left\{h\left(-\frac{i}{2}\right) + h\left(\frac{i}{2}\right)\right\} + \sum h(\gamma_\pi) + \frac{1}{2\pi i}\int_{\operatorname{Re} s = -1} \frac{\Phi'}{\Phi}(s, \pi) H(s)\,ds$$

where the sum is over the zeros, each counted with its multiplicity. The functional equation (2.9) gives

$$\frac{\Phi'}{\Phi}(s, \pi) = -\log Q_\pi - \frac{\Phi'}{\Phi}(1 - s, \tilde{\pi}).$$

Using this and changing variables gives

$$\mathscr{I} = -\delta(\pi)\left\{h\left(-\frac{i}{2}\right) + h\left(\frac{i}{2}\right)\right\} + \sum h(\gamma_\pi) - \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=2} \log Q_\pi H(s)\, ds$$
$$- \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=2} \frac{\Phi'}{\Phi}(s, \tilde{\pi}) H(1 - s)\, ds$$

or

$$\sum h(\gamma_\pi) = \delta(\pi)\left\{h\left(-\frac{i}{2}\right) + h\left(\frac{i}{2}\right)\right\} - \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=2} \log Q_\pi H(s)\, ds$$
$$+ \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=2} \frac{\Phi'}{\Phi}(s, \pi) H(s)\, ds + \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=2} \frac{\Phi'}{\Phi}(s, \tilde{\pi}) H(1 - s)\, ds.$$

Using $\Phi(s, \pi) = L(s, \pi_\infty)L(s, \pi)$, we get

$$\frac{1}{2\pi i}\int_{\mathrm{Re}\, s=2} \frac{\Phi'}{\Phi}(s, \pi) H(s)\, ds = \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=2} \left(\sum_{j=1}^{m} \frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}(s + \mu_\pi(j)) + \frac{L'}{L}(s, \pi)\right) H(s)\, ds.$$

Now, shifting the contour of integration to Re $s = 1/2$,

$$\frac{1}{2\pi i}\int_{\mathrm{Re}\, s=2} \sum_{j=1}^{m} \frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}(s + \mu_\pi(j)) H(s)\, ds = \frac{1}{2\pi}\int_{-\infty}^{\infty} \sum_{j=1}^{m} \frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + ir + \mu_\pi(j)\right) h(r)\, dr.$$

That no poles are picked up on shifting the contour from Re $s = 2$ to Re $s = 1/2$ is equivalent to the inequality (2.5).

Thirdly, using (2.12) we have

$$\frac{1}{2\pi i}\int_{\mathrm{Re}\, s=2} \frac{L'}{L}(s, \pi) H(s)\, ds = -\sum_{n=1}^{\infty} \frac{c_\pi(n)}{\sqrt{n}} \frac{1}{2\pi}\int_{-\infty}^{\infty} h(r) e^{-ir\log n}\, dr$$
$$= -\sum_{n=1}^{\infty} \frac{c_\pi(n)}{\sqrt{n}} g(\log n).$$

We do the same for the integral involving $\Phi'/\Phi(s, \tilde{\pi})$, and use (2.14) and (2.8). Collecting the terms gives the explicit formula of the proposition. □

2.4. Rankin-Selberg convolutions. A crucial ingredient in Section 3 is the asymptotic behaviour of the mean square of $c_\pi(n) = \Lambda(n) a_\pi(n)$. To determine the asymptotics, we will need the Rankin-Selberg L-function. Its general theory has been developed by Jacquet, Piatetski-Shapiro, and Shalika [8], and more recently by Shahidi [32] and Mœglin-Waldspurger [19]. For cuspidal automorphic representations π on GL_m, π' on $GL_{m'}$, the Rankin-Selberg L-function $L(s, \pi \times \pi')$ is defined as a product of local factors $L(s, \pi \times \pi') = \prod_p L(s, \pi_p \times \pi'_p)$. Initially, it is seen to be absolutely convergent for $\operatorname{Re} s \gg 1$, but in the end one finds this to be so in $\operatorname{Re} s > 1$. For primes p where both π_p and π'_p are unramified, the local factor is given in terms of the corresponding semisimple conjugacy classes $A_\pi(p)$, $A'_\pi(p)$ (2.1) by

(2.18)
$$L(s, \pi_p \times \pi'_p) = \det(I - p^{-s}A_\pi(p) \otimes A_{\pi'}(p))^{-1} = \prod_{j,k} (1 - \alpha_\pi(p, j)\alpha_{\pi'}(p, k)p^{-s})^{-1}.$$

The local factors for ramified primes will be described in the appendix. They are of the form $P_p(p^{-s})^{-1}$, where $P_p(x)$ is a polynomial of degree at most mm' with $P(0) = 1$. At infinity the local factor is of the form $\prod_{j,k} \Gamma_{\mathbf{R}}(s + \mu_{\pi\times\pi'}(j, k))$. If π_∞ and π'_∞ are unramified, then

$$\{\mu_{\pi\times\pi'}(j, k)\} = \{\mu_\pi(j) + \mu_{\pi'}(k)\}.$$

See the appendix for a description of the general case.

For us, the case of most interest is $\pi' = \tilde{\pi}$. In this case we see from (2.8) and (2.18) that for π_p unramified,

$$\log L(s, \pi_p \times \tilde{\pi}_p) = \sum_{j,k} \sum_{\nu=1}^{\infty} \frac{(\alpha_\pi(p, j)\overline{\alpha_\pi(p, k)})^\nu}{\nu p^{\nu s}} \tag{2.19}$$
$$= \sum_{\nu=1}^{\infty} \frac{|a_\pi(p^\nu)|^2}{\nu p^{\nu s}}.$$

With the local factors, one can define the completed Rankin-Selberg L-function

$$\Phi(s, \pi \times \tilde{\pi}) = L(s, \pi_\infty \times \tilde{\pi}_\infty)L(s, \pi \times \tilde{\pi}). \tag{2.20}$$

Some of the basic analytic properties of $L(s, \pi \times \tilde{\pi})$ which we will use are as follows.

Property **RS 1** [10]. *The Euler product for $L(s, \pi \times \tilde{\pi})$ converges absolutely for* $\operatorname{Re} s > 1$, *and $L(s, \pi \times \tilde{\pi})$ has a simple pole at $s = 1$.*

Property **RS 2**. *$\Phi(s, \pi \times \tilde{\pi})$ has a meromorphic continuation to the entire complex plane and satisfies a functional equation*

$$\Phi(s, \pi \times \tilde{\pi}) = \varepsilon(s, \pi \times \tilde{\pi})\Phi(1 - s, \pi \times \tilde{\pi})$$
$$\varepsilon(s, \pi \times \tilde{\pi}) = \tau(\pi \times \tilde{\pi})Q_{\pi\times\tilde{\pi}}^{-s}$$

where $Q_{\pi\times\tilde{\pi}} > 0$ and $\tau(\pi \times \tilde{\pi}) = \pm Q_{\pi\times\tilde{\pi}}^{1/2}$.

PROPERTY **RS 3.** *$\Phi(s, \pi \times \tilde{\pi})$ is bounded in vertical strips, and is holomorphic except for simple poles at $s = 0, 1$.*

There are two approaches to proving analytic properties of $\Phi(s, \pi \times \tilde{\pi})$. The first is via Rankin-Selberg integrals as developed by Jacquet, Piatetski-Shapiro, and Shalika, and the second uses the constant term of general Eisenstein series, as is done by Shahidi and by Mœglin and Waldspurger. The first approach yields **RS 1** and **RS 2**, but the complicated nature of the archimedean integrals [11] makes **RS 3** much more elusive by this method. On the other hand, the second method (which avoids such integrals) yields [19] that $\Phi(s, \pi \times \tilde{\pi})$ is entire except for simple poles at $s = 0, 1$. To see that $s(1 - s)\Phi(s, \pi \times \tilde{\pi})$ is of order one and bounded in vertical strips, we can proceed as follows: As in the first part of [11] choose Whittaker functions W_∞ and W'_∞ for π_∞ and $\tilde{\pi}_\infty$, and $\phi \in \mathscr{S}(\mathbf{R}^m)$. The archimedean integrals $\Psi(s, W_\infty, W'_\infty, \phi)$

$$g(s, W_\infty, W'_\infty, \phi) := \frac{\Psi(s, W_\infty, W'_\infty, \phi)}{L(s, \pi_\infty, \times \tilde{\pi}_\infty)}$$

are entire and satisfy a functional equation

$$g(1 - s, \tilde{W}_\infty, \tilde{W}'_\infty, \hat{\phi}) = \varepsilon(s) g(s, W_\infty, W'_\infty, \phi)$$

with $\varepsilon(s)$ of the form ab^s. Moreover, note that $\Psi(s, W_\infty, W'_\infty, \phi)$ is bounded in vertical strips (except for a finite number of poles in the strip in question) and is uniformly bounded for $\operatorname{Re} s \gg 0$. It follows that $g(s, W_\infty, W'_\infty, \phi)$ is of order one. Now, using the global Rankin-Selberg integral, one checks that

$$s(1 - s)\Phi(s, \pi \times \tilde{\pi}) = \frac{B(s, W_\infty, W'_\infty, \phi)}{g(s, W_\infty, W'_\infty, \phi)}$$

where $B(s)$ is entire of order one (it comes from an integral against a standard Eisenstein series). On the other hand, by [19] we know that $s(1 - s)\Phi(s, \pi \times \tilde{\pi})$ is entire, and so it must be of order one. Moreover, $\Phi(s, \pi \times \tilde{\pi})$ is bounded for $\operatorname{Re} s \gg 1$, and hence by the functional equation **RS 2** this is also so for $\operatorname{Re} s \ll 0$. By an application of the Phragmen-Lindelöf principle, the claim follows.

As an application of **RS 1**, we obtain the asymptotics of

$$\sigma(x) := \sum_{\log n \leqslant x} \frac{|c_\pi(n)|^2}{n}. \tag{2.21}$$

We note that the bound (2.3) ensures that the contribution to $\sigma(x)$ of $n = p^e$ for ramified primes $p \in S_\pi$ is bounded independently of x. As for the unramified primes, set $L_S(s, \pi \times \tilde{\pi}) = \prod_{p \notin S} L(s, \pi_p \times \tilde{\pi}_p)$ (sometimes called the partial L-

function). Differentiating (2.19), we see

$$\frac{L_S'}{L_S}(s, \pi \times \tilde{\pi}) = -\sum_{(n,S_\pi)=1} \frac{\Lambda(n)|a_\pi(n)|^2}{n^s}. \tag{2.22}$$

Differentiating (2.22), we have

$$G(s) := \left(\frac{L_S'}{L_S}\right)'(s+1, \pi \times \tilde{\pi}) = \sum_{(n,S_\pi)=1} \frac{(\log n)\Lambda(n)|a_\pi(n)|^2}{n^{s+1}}. \tag{2.23}$$

Since $L(s, \pi \times \tilde{\pi})$ has a simple pole at $s = 1$, it follows from (2.23) that

$$\sum_{n=1}^{\infty} \frac{(\log n)\Lambda(n)|a_\pi(n)|^2}{n} n^{-s} \sim \frac{1}{s^2} \qquad \text{as } s \searrow 0 \ (s \text{ real}).$$

Hence, by a standard Tauberian argument, we conclude that

$$\sigma_1(x) := \sum_{n \leqslant x} \frac{(\log n)\Lambda(n)|a_\pi(n)|^2}{n} \sim \frac{\log^2 x}{2}. \tag{2.24}$$

To relate $\sigma_1(x)$ and $\sigma(x)$, we need to make a technical hypothesis. (This is the hypothesis (1.7).)

Hypothesis **H**. *For any fixed $k \geqslant 2$,*

$$\sum_p \frac{|(\log p)a_\pi(p^k)|^2}{p^k} < \infty.$$

We will establish **H** in many cases below. Note that it is an immediate consequence of the "Ramanujan conjectures" mentioned after (2.2). Indeed, these assert that $|a_\pi(p^k)| \leqslant m$, which implies **H** with lots to spare. In view of (2.3), we see that if $k > (m^2+1)/2$, then

$$\sum_p \frac{|(\log p)a_\pi(p^k)|^2}{p^k} < \infty.$$

Hence, assuming **H**, we have

$$\sum_{k \geqslant 2} \sum_p \frac{|(\log p)a_\pi(p^k)|^2}{p^k} < \infty, \tag{2.25}$$

so that

$$\sigma_1(x) = \sigma(x) + O(1). \tag{2.26}$$

PROPOSITION 2.2. *Assuming* **H**, *we have* $\sigma(x) \sim (\log x)^2/2$.

That this asymptotic is independent of π is at the root of the universality of GUE. Using **RS 3**, we can sharpen Proposition 2.2 somewhat.

PROPOSITION 2.3. *Assuming* **H**, *we have*

$$\sum_{n \leqslant x} \frac{|c_\pi(n)|^2}{n} = \frac{\log^2 x}{2} + O(\log x).$$

Proof. Since the ramified primes contribute only a bounded quantity, we need only estimate the sum over $(n, S_\pi) = 1$. The function $G(s)$ in (2.23) is holomorphic for $\operatorname{Re} s > 0$, with Taylor expansion at $s = 0$ of the form $1/s^2$ + holomorphic. $G(s)$ is meromorphic and has at most double poles in $\operatorname{Re} s \leqslant 0$. A term-by-term integration yields the familiar identity

$$\sum_{n \leqslant x} \left(1 - \frac{n}{x}\right) \frac{|(\log n)\Lambda(n) a_\pi(n)|^2}{n} = \frac{1}{2\pi i} \int_{\operatorname{Re} s = 1} G(s) \frac{x^s}{s(s+1)} \, ds + O(1) \tag{2.27}$$

(the $O(1)$ term coming from the ramified primes and from (2.25)). Now **RS 3** allows us to give standard bounds for $G(s)$ in $\operatorname{Re} s < 0$ and also to bound the number of poles of $G(s)$ in $|s| < T$ by $O_\varepsilon(T^{1+\varepsilon})$. In particular, $\sum 1/|\rho|(|\rho| + 1) < \infty$, where the sum is over the zeros of $L(s, \pi \times \tilde{\pi})$. So shifting the contour in (2.27) to the left of $\operatorname{Re} s = 0$ yields

$$\begin{aligned} \sum_{n \leqslant x} \left(1 - \frac{n}{x}\right) \frac{|c_\pi(n)|^2}{n} &= \operatorname*{Res}_{s=0} \frac{G(s) x^s}{s(s+1)} + O\left(\sum_{\rho \neq 0} \frac{\log x}{|\rho|(|\rho| + 1)}\right) \\ &= \frac{\log^2 x}{2} + O(\log x). \end{aligned} \tag{2.28}$$

If $f(x) = \sigma(\log x)$, σ as in (2.21), then

$$\frac{1}{x} \int_1^x f(t) \, dt = \sum_{n \leqslant x} \left(1 - \frac{n}{x}\right) \frac{|c_\pi(n)|^2}{n} + O(1),$$

and hence

$$\int_1^x f(t) \, dt = \frac{x \log^2 x}{2} + O(x \log x). \tag{2.29}$$

Since f is increasing, we have for any $h \leqslant x$

$$\frac{1}{h} \int_{x-h}^x f(t) \, dt \leqslant f(x) \leqslant \frac{1}{h} \int_x^{x+h} f(t) \, dt.$$

Applying (2.29) with $h = x/4$ yields Proposition 2.3. $\square$

We turn to the "technical hypothesis" **H**. There is little doubt about its truth, since as was pointed out it follows from very modest bounds towards the "Ramanujan conjectures" (which are proven for some of the known π's on GL_m). Even so, we have not been able to establish it in general.

PROPOSITION 2.4. *Hypothesis* **H** *holds for* $1 \leqslant m \leqslant 3$.

Proof. For $m = 1$ this is trivial. We give the proof for $m = 3$. For $m = 2$ it is proven in the same way (or follows from known bounds in that case). Write $A_\pi(p) = \mathrm{diag}(\alpha_1, \alpha_2, \alpha_3)$ so that $a_\pi(p^k) = \mathrm{tr}\, A_\pi(p)^k$. Assume that $|\alpha_1| \geqslant |\alpha_2| \geqslant |\alpha_3|$. Since $\omega_\pi(p) = \det A_\pi(p)$ has absolute value one and $\{\overline{\alpha}_j\} = \{\alpha_j^{-1}\}$ by (2.8), we must have $|\alpha_2| = 1$ and $|\alpha_3| = 1/|\alpha_1|$. Therefore,

$$|\alpha_1| + |\alpha_2| + |\alpha_3| \leqslant |\alpha_1 + \alpha_2 + \alpha_3| + 4$$

and

$$|a_\pi(p^k)|^2 \ll 1 + |a_\pi(p)|^{2k}.$$

Together with $|a_\pi(p)| \ll p^{1/2-1/10}$ (see (2.3)), we get

$$\sum_p \frac{|(\log p)^2 |a_\pi(p^k)|^2}{p^k} \ll \sum_p \frac{\log^2 p}{p^k} + \sum_p \frac{\log^2 p |a_\pi(p)|^2}{p^{1+2(k-1)/5}}.$$

Since $k > 1$ we can apply **RS 1** (in particular, the convergence in $\mathrm{Re}\, s > 1$) to conclude that

$$\sum_p \frac{|(\log p)^2 |a_\pi(p^k)|^2}{p^k} < \infty. \quad \square$$

For the rest of the paper we will assume that either $m \leqslant 3$ or that Hypothesis **H** is valid.

3. Sums over primes. We wish to study the asymptotic behaviour of the n-level correlation function

$$R_n(f, T) = \frac{1}{N} \sideset{}{^*}\sum_{i_1, \ldots, i_n \leqslant N} f(\tilde{\gamma}_{i_1}, \ldots, \tilde{\gamma}_{i_n}) \tag{3.1}$$

where $N = N(T)$ and $\sum^*$ means we sum over distinct indices i_j. Instead of looking directly at $R_n(f, T)$, we instead look at the sums

$$C_n(f, T) = \sum_{i_1, \ldots, i_n \leqslant N} f(\tilde{\gamma}_{i_1}, \ldots, \tilde{\gamma}_{i_n}). \tag{3.2}$$

It is important to note that the sum in (3.2) is no longer over *distinct* ordered zeros as in the definition of the n-level correlation function (3.1). We will recover (3.1) from (3.2) by combinatorial sieving in Section 4. In order to determine the asymptotics of $C_n(f, T)$, we look at smoothed sums

$$C_n(f, h, T) := \sum_{\gamma_1, \ldots, \gamma_n} h_1\left(\frac{\gamma_1}{T}\right) \cdots h_n\left(\frac{\gamma_n}{T}\right) f\left(\frac{L}{2\pi}\gamma_1, \ldots, \frac{L}{2\pi}\gamma_n\right) \tag{3.3}$$

where we have set

$$L = m \log T \tag{3.4}$$

and $h_j(r)$ is a smooth "cutoff"—we take

$$h_j(r) = \int_{-\infty}^{\infty} g_j(u) e^{iru}\, du \tag{3.5}$$

with $g_j \in C_c^\infty(\mathbf{R})$. Our main result in this section is Theorem 3.1, which gives the asymptotics of $C_n(f, h, T)$. We prove it for f satisfying **TF 2**, **3**, though in the end one is only interested in looking at symmetric f. Our reason for considering these more general f's is to carry out the induction in Section 4. In fact, it will be convenient to work with the Fourier transform of f; thus for Φ, a compactly supported C^1 function on $\mathbf{R}^n$, we get an f satisfying **TF 2** by setting

$$f(x) = \int_{\mathbf{R}^n} \Phi(\xi) \delta(\xi_1 + \cdots + \xi_n) e(-x \cdot \xi)\, d\xi. \tag{3.6}$$

In the sequel, we set for $\mathbf{h} = (h_1, \ldots, h_n)$

$$\kappa(\mathbf{h}) = \int_{-\infty}^{\infty} h_1(r) \cdots h_n(r)\, dr. \tag{3.7}$$

THEOREM 3.1. *Let $\Phi \in C^1(\mathbf{R}^n)$ be supported in $\sum_{j=1}^n |\xi_j| < 2/m$, and let $f(x) = \int_{\mathbf{R}^n} \Phi(\xi) \delta(\xi_1 + \cdots + \xi_n) e(-x \cdot \xi)\, d\xi$. Then, for h_j as in (3.5), we have*

$$\sum h_1\left(\frac{\gamma_1}{T}\right) \cdots h_n\left(\frac{\gamma_n}{T}\right) f\left(\frac{L}{2\pi}\gamma_1, \ldots, \frac{L}{2\pi}\gamma_n\right) = \kappa(\mathbf{h}) \frac{TL}{2\pi} \int_{\mathbf{R}^n} C_{\underline{O}}(v) \Phi(v)\, dv + O(T) \tag{3.8}$$

with

$$\int_{\mathbf{R}^n} \Phi(v) C_{\underline{O}}(v)\, dv \tag{3.9}$$
$$= \Phi(0) + \sum_{r=1}^{[n/2]} \sum \int \cdots \int |v_1| \cdots |v_r| \Phi(v_1 \mathbf{e}_{i(1), j(1)} + \cdots + v_r \mathbf{e}_{i(r), j(r)})\, dv_1 \cdots dv_r$$

where the sum is over all choices of r disjoint *pairs of indices $i(t) < j(t)$ in $\{1, \ldots, n\}$ and for $i < j$ we set*

$$\begin{cases} \mathbf{e}_{i,j} &= \mathbf{e}_i - \mathbf{e}_j \\ \mathbf{e}_i &= (0, \ldots, 1, 0, \ldots) \text{ the } i\text{th standard basis vector}. \end{cases} \tag{3.10}$$

From Theorem 3.1, we will deduce the asymptotics of the unsmoothed sums $C_n(f, T)$; for this we will assume the Riemann hypothesis for $L(s, \pi)$.

THEOREM 3.2. *Let $\Phi \in C^2(\mathbf{R}^n)$ be supported in $\sum |\xi_j| < 2/m$, and f be given by (3.6). Assume the Riemann hypothesis for $L(s, \pi)$; then*

$$C_n(f, T) \sim N(T) \int_{\mathbf{R}^n} \Phi(u) C_{\underline{Q}}(u)\, du + O(T).$$

Proof of Theorem 3.1. To begin the proof, we rewrite the sum $C_n(f, h, T)$ using the Fourier transform as

$$C_n(f, h, T) = \int_{\mathbf{R}^n} \prod_{j=1}^{n} \left\{ \sum_{\gamma_j} h_j\left(\frac{\gamma_j}{T}\right) e^{-iL\gamma_j\xi_j} \right\} \Phi(\xi) \delta(\xi_1 + \cdots + \xi_n)\, d\xi. \tag{3.11}$$

We can convert (3.11) into a sum over primes by use of the explicit formula (2.16), with the test functions

$$H_T(r) = h_j\left(\frac{r}{T}\right) e^{-irL\xi}, \qquad G_T(u) = Tg_j(T(L\xi + u)) \tag{3.12}$$

where $g_j(u)$, $h_j(r)$ are as in (3.5), and $\xi \in \mathbf{R}$.

The explicit formula (2.16) with this choice reads

(3.13)
$$\begin{aligned} \sum h\left(\frac{\gamma}{T}\right) e^{-iL\gamma\xi} &= \delta(\pi) \left\{ h\left(-\frac{i}{2T}\right) T^{-m\xi/2} + h\left(\frac{i}{2T}\right) T^{m\xi/2} \right\} + \log Q_\pi \cdot Tg(TL\xi) \\ &\quad + \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{j=1}^{m} \left(\frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + \mu_\pi(j) + ir\right) \right. \\ &\qquad \left. + \frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + \overline{\mu_\pi(j)} - ir\right) \right) h\left(\frac{r}{T}\right) e^{-irL\xi}\, dr \\ &\quad - T \sum_{n=1}^{\infty} \frac{\Lambda(n)}{\sqrt{n}} \{a(n) g(T(L\xi + \log n)) + \overline{a(n)} g(T(L\xi - \log n))\} \\ &= \text{polar} + Tg_T(TL\xi) + TS^+(\xi) + TS^-(\xi) \end{aligned}$$

where

$$\text{polar} = \delta(\pi)\left\{h\left(-\frac{i}{2T}\right)T^{-m\xi/2} + h\left(\frac{i}{2T}\right)T^{m\xi/2}\right\} \tag{3.14}$$

$$g_T(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(r)\Omega_\pi(rT)e^{-irx}\,dr \tag{3.15}$$

$$\Omega_\pi(r) = \log Q_\pi + \sum_{j=1}^{m} \frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + \mu_\pi(j) + ir\right) + \frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + \overline{\mu_\pi(j)} - ir\right) \tag{3.16}$$

$$\begin{aligned} S^+(\xi) &= -\sum_{n=1}^{\infty} \frac{\Lambda(n)a(n)}{\sqrt{n}} g(T(L\xi + \log n)) \\ S^-(\xi) &= -\sum_{n=1}^{\infty} \frac{\Lambda(n)\overline{a(n)}}{\sqrt{n}} g(T(L\xi - \log n)). \end{aligned} \tag{3.17}$$

(The term *polar* occurs only in the case of $\zeta(s)$; in the sequel we omit it.)

Inserting (3.13) into (3.11) with the different h_j, we find that we have expressed $C_n(f, h, T)$ as a sum over primes as desired:

(3.18)

$$C_n(f, h, T) = \int_{\mathbf{R}^n} \prod_{j=1}^{n} \{Tg_{j,T}(TL\xi_j) + TS_j^+(\xi_j) + TS_j^-(\xi_j)\}\Phi(\xi)\delta(\xi_1 + \cdots + \xi_n)\,d\xi.$$

Expanding the product in (3.18), we find that $C_n(f, h, T)$ is an alternating sum of terms of the form

$$C_{r,s}(T) = \sum \frac{c(n_1)\cdots c(n_r)\overline{c(n_{r+1})}\cdots\overline{c(n_{r+s})}}{\sqrt{n_1\cdot\,\cdots\,\cdot n_{r+s}}} A_{r,s}(\mathbf{n}, T) \tag{3.19}$$

where we have set

$$c(n) = \Lambda(n)a(n) \tag{3.20}$$

and

$$\begin{aligned} A_{r,s}(\mathbf{n}, T) = T^n \int \prod_{j=1}^{r} g_j(T(L\xi_j + \log n_j)) \prod_{j=r+1}^{r+s} g_j(T(L\xi_j - \log n_j)) \\ \cdot \prod_{j>r+s} g_{j,T}(TL\xi_j)\cdot\Phi(\xi)\delta(\xi_1 + \cdots + \xi_n)\,d\xi. \end{aligned} \tag{3.21}$$

In expressing $C_n(f, h, T)$ as a sum of various $C_{r,s}(T)$, we get terms from all possible choices of r of the factors to be $S_j^+(\xi_j)$, s of the factors to be $S_j^-(\xi_j)$, and the remaining $k = n - r - s$ of the factors to be $g_{j,T}(TL\xi_j)$.

LEMMA 3.1. (1) *We have*

$$g_T(x) \ll \begin{cases} \log T, & |x| \ll \log\log T \\ \dfrac{1}{|x|}, & |x| \gg \log\log T. \end{cases}$$

(2) *We have* $\int |g_T(x)|\, dx \ll \log T$.

Proof. Recall that by (3.16),

$$g_T(x) = \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=0} h\left(\frac{s}{i}\right)\omega_\pi(Ts)e^{-sx}\, ds$$

with

$$(3.22)\quad \omega_\pi(s) = \log Q_\pi + \sum_{j=1}^{m}\left(\frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + \mu_\pi(j) + s\right) + \frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + \overline{\mu_\pi(j)} - s\right)\right).$$

Assuming that $x > 0$, we shift the contour of integration to the right to $s = \sigma + ir$, with $\sigma > 0$; this we can do since, by Stirling's formula, $\Omega_\pi(r) \ll \log(r)$ and $h(r)$ is rapidly decreasing as $|\mathrm{Re}(r)| \to \infty$. Since $\mathrm{Re}(1/2 + \mu_\pi(j)) > 0$ (2.5), the first Γ-factor is holomorphic in $\mathrm{Re}\, s \geqslant 0$. The second factors contribute simple poles at

$$s_k^{(j)} = \frac{1/2 + \bar{\mu}_\pi(j) + 2k}{T}, \qquad 0 \leqslant k \leqslant \frac{\sigma}{2}T.$$

Thus

$$g_T(x) = \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=\sigma} h\left(\frac{s}{i}\right)\omega_\pi(sT)e^{-sx}\, ds + \frac{1}{T}\sum_{j=1}^{m}\sum_{0\leqslant k\leqslant(\sigma/2)T} h\left(\frac{s_k^{(j)}}{i}\right)e^{-s_k^{(j)}x}.$$

The double sum is majorized by

$$\frac{1}{T}\sum_{j=1}^{m}\sum_{0\leqslant k\leqslant(\sigma/2)T} e^{-(1/2+\bar{\mu}_\pi(j)+2k)x/T} \ll \frac{e^{-\alpha x/T}}{x}(1 - e^{-\sigma x/2})$$

where $\alpha = \min_{1\leqslant j\leqslant m}\{1/2 + \mathrm{Re}\, \mu_\pi(j)\} > 0$.

As for the integral, by Stirling's formula,

$$\frac{\Gamma'_{\mathbf{R}}}{\Gamma_{\mathbf{R}}}\left(\frac{1}{2} + \bar{\mu}_\pi(j) \pm sT\right) \ll \log((1 + |s|)T),$$

and since $h(s/i)$ is rapidly decreasing in vertical strips, we can bound the integral by

$$\frac{1}{2\pi i}\int_{\mathrm{Re}\, s=\sigma} h\left(\frac{s}{i}\right)\omega_\pi(sT)e^{-sx}\, ds \ll e^{-\sigma x}\log T.$$

Thus we see that

$$g_T(x) \ll \frac{e^{-\alpha x/T}}{x}(1 - e^{-\sigma x/2}) + e^{-\sigma x}\log T,$$

giving part (1). Part (2) follows from integrating this. □

From Lemma 3.1 we see that the integrals defining $A_{r,s}(\mathbf{n}, T)$ are rapidly convergent.

LEMMA 3.2. *Let Φ as in (3.6) be supported in $|\xi_1| + \cdots + |\xi_n| \leqslant (2 - \delta)/m$. Then $A_{r,s}(\mathbf{n}, T) = 0$ unless $|n_j| \ll T$ and $n_1 n_2 \cdot \cdots \cdot n_{r+s} \ll T^{2-\delta}$.*

Proof. The integrand in (3.21) is zero unless there is an $\eta \in \operatorname{Supp}\Phi$ (so $\sum_j \eta_j = 0$) such that

$$\begin{cases} |T(\eta_j L + \log n_j)| \ll 1, & j = 1, \ldots, r \\ |T(\eta_j L - \log n_j)| \ll 1, & j = r+1, \ldots, r+s. \end{cases} \tag{3.23}$$

Hence

$$\frac{n_j}{T^{m|\eta_j|}} = 1 + O\left(\frac{1}{T}\right) \tag{3.24}$$

so that $n_j \ll T^{m|\eta_j|} \ll T^{1-\delta/2}$ and $n_1 n_2 \cdot \cdots \cdot n_{r+s} \ll T^{m\sum|\eta_j|} \ll T^{2-\delta}$. □

What follows is a series of reductions which show that the main term in $C_{r,s}(T)$ comes from the "diagonal sums."

LEMMA 3.3. *Let $\tilde{A}_{r,s}(\mathbf{n}, T)$ be as in (3.21) with the region of integration in the variables ξ_j, $j > r + s$, restricted to $|TL\xi_j| \ll T^{\delta/3}$. Then, for T sufficiently large, $\tilde{A}_{r,s}(\mathbf{n}, T) = 0$ unless $\mathbf{n} = (n_1, \ldots, n_{r+s})$ satisfies the conclusion of Lemma 3.2 and, in addition,*

$$n_1 \cdots n_r = n_{r+1} \cdots n_{r+s}. \tag{3.25}$$

Proof. Since g has compact support, in order that the integrand not vanish we need some $\eta \in \operatorname{Supp} \Phi$, that is,

$$\begin{cases} |T(\eta_j L + \log n_j)| \ll 1, & j = 1, \ldots, r \\ |T(\eta_j L - \log n_j)| \ll 1, & j = r+1, \ldots, r+s \\ |TL\eta_j| \ll T^{\delta/3}, & j > r+s. \end{cases} \tag{3.26}$$

Furthermore, $\sum_{j=1}^{n} \eta_j = 0$, and hence

$$\left|\log \frac{n_1 \cdots n_r}{n_{r+1} \cdots n_{r+s}}\right| = \left|\sum_{j=1}^{r} (L\eta_j + \log n_j) + \sum_{j=r+1}^{r+s} (L\eta_j - \log n_j) + \sum_{j>r+s} L\eta_j\right| \ll \frac{1}{T} + T^{\delta/3-1}. \tag{3.27}$$

Thus

$$\left|\log \frac{n_1 \cdots n_r}{n_{r+1} \cdots n_{r+s}}\right| \ll T^{-1+\delta/3}. \tag{3.28}$$

Setting $M = n_1 \cdots n_r$, $N = n_{r+1} \cdots n_{r+s}$, we know that $MN \ll T^{2-\delta}$ and that $|\log M/N| \ll T^{-1+\delta/3}$. Assume that $M \neq N$, say $M = N + u$, $u \geqslant 1$. Then

$$T^{-1+\delta/3} \gg \left|\log \frac{M}{N}\right| = \log\left(1 + \frac{u}{N}\right) \geqslant \frac{1}{N} \geqslant \frac{1}{\sqrt{MN}} \gg T^{-1+\delta/2}. \tag{3.29}$$

Since $\delta > 0$, this gives a contradiction for T large. Thus $M = N$, establishing the lemma. □

Recall the sum $C_{r,s}(T)$ given in (3.19), and denote by $\tilde{C}_{r,s}(T)$ the corresponding sum with $A_{r,s}(\mathbf{n}, T)$ replaced by $\tilde{A}_{r,s}(\mathbf{n}, T)$.

PROPOSITION 3.1. $C_{r,s}(T) = \tilde{C}_{r,s}(T) + O(T^{1-\delta/3})$.

We begin by estimating the difference between $A_{r,s}(\mathbf{n}, T)$ and $\tilde{A}_{r,s}(\mathbf{n}, T)$. As in the proof of Lemma 3.3, we set $M = n_1 \cdots n_r$, $N = n_{r+1} \cdots n_{r+s}$.

LEMMA 3.4. *If $NM \ll T^{2-\delta}$, then*

$$A_{r,s}(\mathbf{n}, T) - \tilde{A}_{r,s}(\mathbf{n}, T) \ll \begin{cases} T^{1-\delta/3} L^{-r-s}, & |\log M/N| \ll T^{\delta/3-1} \\ \dfrac{L^{-r-s}}{|\log M/N|}, & |\log M/N| \gg T^{\delta/3-1}. \end{cases}$$

Proof. The region of integration for the difference $A_{r,s}(\mathbf{n}, T) - \tilde{A}_{r,s}(\mathbf{n}, T)$ is a union U of the sets $\mathscr{F}_k = \{\xi : |TL\xi_k| \gg T^{\delta/3}\}$, $k > r + s$. Without loss of generality, we estimate this integral over the region $\mathscr{F}_n$. For this purpose, set

$$x_j = \begin{cases} T(L\xi_j + \log n_j), & 1 \leqslant j \leqslant r \\ T(L\xi_j - \log n_j), & r+1 \leqslant j \leqslant r+s \\ TL\xi_j, & j > r+s. \end{cases}$$

We have, on changing variables,

$$\begin{aligned} A_{r,s}(\mathbf{n}, T) - \tilde{A}_{r,s}(\mathbf{n}, T) &= T^n \int_U \prod_{j=1}^{r} g_j(T(L\xi_j + \log n_j)) \prod_{j=r+1}^{r+s} g_j(T(L\xi_j - \log n_j)) \\ &\quad \cdot \prod_{j>r+s} g_{j,T}(TL\xi_j) \cdot \Phi(\xi)\delta(\xi_1 + \cdots + \xi_n)\, d\xi \\ &\ll \frac{T}{L^{n-1}} \int_{\{|x_j| \ll TL,\, |x_n| \gg T^{\delta/3}\}} \prod_{j \leqslant r+s} |g_j(x_j)| \prod_{r+s<j\leqslant n-1} |g_{j,T}(x_j)| \\ &\quad \cdot \left| g_{n,T}\left(T\log M - T\log N - \sum_{j=1}^{n-1} x_j \right) \right| dx_1 \cdots dx_{n-1}. \end{aligned}$$

We claim that

$$(3.30) \quad \begin{aligned} &\int_{\{|x_j| \ll TL,\ |x_n| \gg T^{\delta/3}\}} \\ &\quad \cdot \left| \prod_{j\leqslant r+s} g_j(x_j) \prod_{r+s<j\leqslant n-1} g_{j,T}(x_j) g_{n,T}\left(T\log M/N - \sum_{j=1}^{n-1} x_j \right) \right| dx_1 \cdots dx_{n-1} \\ &\ll \begin{cases} \dfrac{L^{n-1-r-s}}{T^{\delta/3}}, & |\log M/N| \ll T^{\delta/3-1} \\[2ex] \dfrac{L^{n-1-r-s}}{T|\log M/N|}, & |\log M/N| \gg T^{\delta/3-1}. \end{cases} \end{aligned}$$

To see this, first assume that $|\log M - \log N| \ll T^{\delta/3-1}$. As in the proof of Lemma 3.3, since $MN \ll T^{2-\delta}$, this implies that $M = N$, and so in this case, on using Lemma 3.1, the integral in (3.30) is bounded by

$$\int_{|x_j| \ll TL} \left| \prod_{j\leqslant r+s} g_j(x_j) \prod_{r+s<j\leqslant n-1} g_{j,T}(x_j) \right| T^{-\delta/3}\, dx_1 \cdots dx_{n-1} \ll \frac{L^{n-1-r-s}}{T^{\delta/3}}.$$

If $|\log M - \log N| \gg T^{\delta/3-1}$, then we write the integral as a sum $I_1 + I_2$ of integrals over regions $|\sum_{j=1}^{n-1} x_j| < (1/2)T|\log M/N|$ and $|\sum_{j=1}^{n-1} x_j| \geqslant (1/2)T|\log M/N|$. In the first case, we have $|x_n| = |T\log M/N - \sum_{j=1}^{n-1} x_j| \gg T|\log M/N|$, and then

$$I_1 \ll \frac{1}{|T\log M - T\log N|}\int_{|\sum_{j=1}^{n-1} x_j| < (1/2)T|\log M/N|}\left|\prod_{j\leqslant r+s} g_j(x_j)\prod_{r+s<j\leqslant n-1} g_{j,T}(x_j)\right| dx$$

$$\ll \frac{L^{n-1-r-s}}{|T\log M - T\log N|}.$$

For the integral I_2 over the region where $|\sum x_j| \gg T|\log M/N|$, we write the region as a union of domains where for some $j > r+s$ we have $|x_j| \gg T|\log M/N|$ (this is possible since $|x_j| \ll 1$ if $j \leqslant r+s$). On such a domain, we use

$$g_T(x_j) \ll \frac{1}{|x_j|} \ll \frac{1}{T|\log M/N|}$$

(Lemma 3.1) to see that the integral I_2 over, say, $|x_{n-1}| \gg T|\log M/N|$, is bounded by

$$\ll \frac{1}{T|\log M/N|}\int_{\{x\in\mathscr{F}_n:\,|x_{n-1}|\gg T|\log M/N|\}}\left|\sum_{j=1}^{r+s} g_j(x_j)\prod_{j>r+s}^{n-2} g_{j,T}(x_j)\right.$$

$$\left.\cdot\, g_{n,T}\left(T\log M/N - \sum_{j=1}^{n-1} x_j\right)\right| dx_1\cdots dx_{n-1}.$$

Now set $y = T\log M/N - \sum_{j=1}^{n-1} x_j$ and change variables in the integral over x_{n-1}; in the region of integration, it is bounded by

$$\int_{T^{\delta/3}\ll y\ll TL}\frac{dy}{y} \ll L.$$

We find that

$$\frac{1}{T|\log M/N|}\int_{\{x\in\mathscr{F}_n:\,|x_{n-1}|\gg T|\log M/N|\}} \ll \frac{L^{n-1-r-s}}{|T\log M - T\log N|}.$$

This proves our claim (3.30) and so Lemma 3.4. □

To prove Proposition 3.1, we divide the difference $C_{r,s}(T) - \tilde{C}_{r,s}(T)$ into two sums $\Sigma_{\text{diag}} + \Sigma_{\text{off}}$, the first sum Σ_{diag} over $\mathbf{n}$ such that $|\log M/N| \ll T^{\delta/3-1}$, which implies as before that $M = N$, and the second sum Σ_{off} over $\mathbf{n}$ for which $T|\log M/N| \gg T^{\delta/3}$.

For the diagonal sum Σ_{diag}, we need to note that

$$\sum_{\substack{M=N \\ n_1\cdots n_{r+s}\ll T^{2-\delta}}} \sum_{j=1}^{r+s} \frac{|c(n_j)|}{\sqrt{n_j}} \ll (\log T)^{r+s}.$$

We defer the proof to Lemma 3.9 where we will see a more precise result. Combining this with Lemma 3.4, we see that

$$\Sigma_{\text{diag}} \ll T^{1-\delta/3}. \tag{3.31}$$

Next we handle the off-diagonal sum Σ_{off}. We have

$$\Sigma_{\text{off}} \ll \frac{1}{L^{r+s}} \sum_{\substack{MN\ll T^{2-\delta} \\ M\neq N}} \frac{1}{|\log M/N|} \prod_{j=1}^{r+s} \frac{|c(n_j)|}{\sqrt{n_j}}.$$

Setting

$$a_k(M) := \sum_{n_1\cdots n_k=M} \prod_{j=1}^{k} |c(n_j)|,$$

we have

$$\Sigma_{\text{off}} \ll \frac{1}{L^{r+s}} \sum_{\substack{MN\ll T^{2-\delta} \\ M\neq N}} \frac{a_r(M)a_s(N)}{\sqrt{MN}|\log M/N|}. \tag{3.32}$$

LEMMA 3.5. *For $k \geqslant 1$ fixed, and any $\varepsilon > 0$,*

$$\sum_{m\leqslant X} a_k(m)^2 \ll_\varepsilon X^{1+\varepsilon}.$$

Proof. We begin by noting that, on using Cauchy-Schwartz and the fact that the number of ways of writing $m = m_1\cdot \cdots \cdot m_k$ is $O(m^\varepsilon)$ for any $\varepsilon > 0$, we have

$$a_k(m)^2 \ll m^\varepsilon \sum_{m_1\cdot \cdots \cdot m_k=m} |c(m_1)|^2\cdots|c(m_k)|^2,$$

and so

$$\sum_{m\leqslant X} a_k(m)^2 \ll X^\varepsilon \sum_{m_1\cdot \cdots \cdot m_k\leqslant X} |c(m_1)|^2\cdots|c(m_k)|^2. \tag{3.33}$$

To estimate the above sum, we first note that

$$\sum_{n\leqslant X} |c(n)|^2 \ll X^{1+\varepsilon} \tag{3.34}$$

for all $\varepsilon > 0$, which follows immediately from the absolute convergence of (2.22) in $\mathrm{Re}(s) > 1$ together with that series having nonnegative coefficients. Next, we make a dyadic decomposition of the sum (3.33) into $O((\log X)^k)$ terms of the form

$$\sum_{M_1 \leqslant m_1 \leqslant 2M_1} \cdots \sum_{M_k \leqslant m_k \leqslant 2M_k} |c(m_1)|^2 \cdots |c(m_k)|^2.$$

If $M_1 \cdots M_k \ll X$, then on using (3.34) we find

$$\sum_{M_1 \leqslant m_1 \leqslant 2M_1} \cdots \sum_{M_k \leqslant m_k \leqslant 2M_k} |c(m_1)|^2 \cdots |c(m_k)|^2 \ll M_1^{1+\varepsilon} \cdots M_k^{1+\varepsilon} \ll X^{1+\varepsilon},$$

which gives the desired estimate. □

Returning to (3.32), we may without loss of generality assume that $N < M$, which implies that $N \ll T^{1-\delta/2}$. We consider two ranges in (3.32): the sum $\mathscr{I}_1$ with $N < M < 2N$ and the sum $\mathscr{I}_2$ with $N \leqslant M/2$. For the first case,

$$\begin{aligned}
\mathscr{I}_1 &= \sum_{N \leqslant T^{1-\delta/2}} \sum_{N < M < 2N} \frac{a_r(N) a_s(M)}{\sqrt{NM} \log M/N} \\
&\ll \sum_{N \leqslant T^{1-\delta/2}} \sum_{k=1}^{N} \frac{a_r(N) a_s(N+k)}{\sqrt{N(N+k)} \log(1+k/N)} \\
&\ll \sum_{N \leqslant T^{1-\delta/2}} \sum_{k=1}^{N} \frac{a_r(N) a_s(N+k)}{k} \\
&\ll \sum_{k \leqslant T^{1-\delta/2}} \frac{1}{k} \sum_{N \leqslant 2T^{1-\delta/2}} (a_r(N)^2 + a_s(N)^2) \\
&\ll (T^{1-\delta/2})^{1+\varepsilon_1} \log T,
\end{aligned}$$

so that for all $\varepsilon > 0$

$$\mathscr{I}_1 \ll T^{1-\delta/2+\varepsilon}. \tag{3.35}$$

For the sum $\mathscr{I}_2$, when $2N < M$, then $\log M/N > \log 2$, so that

$$\mathscr{I}_2 = \sum_{N \leqslant T^{1-\delta/2}} \sum_{\substack{2N < M \\ MN < T^{2-\delta}}} \frac{a_r(N) a_s(M)}{\sqrt{MN} \log M/N} \ll \sum_{MN < T^{2-\delta}} \frac{a_r(N) a_s(M)}{\sqrt{MN}}.$$

Again by use of a dyadic decomposition, we will know

$$\mathscr{I}_2 \ll T^{1-\delta/2+\varepsilon} \tag{3.36}$$

once we can show that for $AB < T^{2-\delta}$

$$\sum_{\substack{A \leqslant N \leqslant 2A \\ B \leqslant M \leqslant 2B}} \frac{a_r(N)a_s(M)}{\sqrt{MN}} \ll (AB)^{1/2+\varepsilon}. \tag{3.37}$$

Now the left-hand side in (3.37) is

$$\left(\sum_{A \leqslant N \leqslant 2A} \frac{a_r(N)}{\sqrt{N}}\right)\left(\sum_{B \leqslant M \leqslant 2B} \frac{a_s(M)}{\sqrt{M}}\right)$$

$$\ll \left(\sum_{A \leqslant N \leqslant 2A} \frac{1}{N}\right)^{1/2}\left(\sum_{A \leqslant N \leqslant 2A} a_r(N)^2\right)^{1/2}\left(\sum_{B \leqslant M \leqslant 2B} \frac{1}{M}\right)^{1/2}\left(\sum_{B \leqslant M \leqslant 2B} a_s(M)^2\right)^{1/2}$$

$$\ll A^{1/2+\varepsilon_1}B^{1/2+\varepsilon_1},$$

by Lemma 3.5.

Combining the estimates (3.35) and (3.36), we conclude that

$$\Sigma_{\text{off}} \ll T^{1-\delta/2+\varepsilon} \tag{3.38}$$

for any $\varepsilon > 0$. From (3.31) and (3.38), we get Proposition 3.1. □

We have seen that for $T \gg 1$,

$$C_{r,s}(T) = \sum_{\substack{n_j \ll T \\ n_1 \cdots n_r = n_{r+1} \cdots n_{r+s} \\ n_1 \cdots n_{r+s} \ll T^2}} \prod_{j=1}^{r} \frac{c(n_j)}{n_j} \prod_{j=r+1}^{r+s} \overline{c(n_j)} \tilde{A}_{r,s}(\mathbf{n}, T) + O(T^{1-\delta/3}). \tag{3.39}$$

Now change variables in the integral (3.21) for $\tilde{A}_{r,s}(\mathbf{n}, T)$ (when $n_1 \cdots n_r = n_{r+1} \cdot n_{r+s}$) by setting

$$y_j = \begin{cases} T(L\xi_j + \log n_j), & 1 \leqslant j \leqslant r \\ T(L\xi_j - \log n_j), & r+1 \leqslant j \leqslant r+s \\ TL\xi_j, & j > r+s. \end{cases} \tag{3.40}$$

Note that we still have $\sum_j y_j = 0$, and the region of integration is

$$V = \begin{cases} \sum_j y_j = 0 \\ |y_j| \ll 1, & j \leqslant r+s \\ |y_j| \ll T^{\delta/3}, & j > r+s. \end{cases} \tag{3.41}$$

We then get

$$(3.42)\qquad \tilde{A}_{r,s}(\mathbf{n}, T) = \frac{T}{L^{n-1}} \int_V \prod_{j=1}^{r+s} g_j(y_j) \prod_{j>r+s} g_{j,T}(y_j)$$
$$\cdot \Phi\left(\frac{y_1}{TL} - \frac{\log n_1}{L}, \ldots, \frac{y_{r+s}}{TL} + \frac{\log n_{r+s}}{L}, \ldots, \frac{y_n}{TL}\right) dy.$$

Expanding Φ in a Taylor series about the point $(-(\log n_1)/L, \ldots, (\log n_{r+s})/L, \ldots, 0)$ (recall that Φ is C^1) and using the constraints (3.41), we see that

$$(3.43)\qquad \tilde{A}_{r,s}(\mathbf{n}, T) = \frac{T}{L^{n-1}} \int_V \prod_{j=1}^{r+s} g_j(y_j) \prod_{j>r+s} g_{j,T}(y_j)\, dy$$
$$\cdot \left\{\Phi\left(-\frac{\log n_1}{L}, \ldots, \frac{\log n_{r+s}}{L}, \ldots, 0\right) + O(T^{-1+\delta/3})\right\}$$

with the error term uniform in $\mathbf{n}$ subject to $n_j \ll T$.

LEMMA 3.6. *Setting $k = n - r - s$, we have*

$$(3.44)\qquad \int_V \prod_{j=1}^{r+s} g_j(y_j) \prod_{j>r+s} g_{j,T}(y_j)\, dy = \frac{1}{2\pi} \kappa(\mathbf{h}) L^k + O(L^{k-1}).$$

Proof. First we claim that

(3.45)
$$\int_V \prod_{j=1}^{r+s} g_j(y_j) \prod_{j>r+s} g_{j,T}(y_j)\, dy = \int_{\sum y_j = 0} \prod_{j=1}^{r+s} g_j(y_j) \prod_{j>r+s} g_{j,T}(y_j)\, dy + O\left(\frac{L^{n-1-r-s}}{T^{\delta/3}}\right).$$

Indeed, the difference between the two integrals in (3.45) is an integral over the union $\bigcup_{k>r+s} V_k$, where

$$V_k = \left\{\sum_{j=1}^{n} y_j = 0 \colon |y_k| \gg T^{\delta/3}\right\}.$$

It suffices to estimate the integral in (3.45) over such V_k, say $k = n$:

$$S_n := \int_{V_n} \left|\prod_{j=1}^{r+s} g_j(y_j) \prod_{j>r+s} g_{j,T}(y_j)\right| dy$$
$$= \iint_{|y_1 + \cdots + y_{n-1}| \gg T^{\delta/3}} \left|\prod_{j=1}^{r+s} g_j(y_j) \prod_{j=r+s+1}^{n-1} g_{j,T}(y_j)\right| \cdot \left|g_{n,T}\left(-\sum_{j=1}^{n-1} y_j\right)\right| dy_1 \cdots dy_{n-1}.$$

By Lemma 3.1, part (1), on V_n,

$$\left| g_{n,T}\left(-\sum_{j=1}^{n-1} y_j\right)\right| \ll \frac{1}{|\sum_{j=1}^{n-1} y_j|} \ll \frac{1}{T^{\delta/3}},$$

so that

$$S_n \ll \frac{1}{T^{\delta/3}} \iint \left| \prod_{j=1}^{r+s} g_j(y_j) \prod_{j=r+s+1}^{n-1} g_{j,T}(y_j)\right| dy_1 \cdots dy_{n-1} \ll \frac{L^{n-1-r-s}}{T^{\delta/3}}.$$

Applying Parseval's equality to (3.45) gives

$$\int_{\sum_j y_j=0} g_1(y_1)\cdots g_n(y_n)\, dy = \frac{1}{2\pi}\int_{-\infty}^{\infty} h_1(r)\cdots h_n(r)\, dr = \frac{1}{2\pi}\kappa(\mathbf{h}), \tag{3.46}$$

from which it follows that

$$\int_{\sum y_j=0} \prod_{j=1}^{r+s} g_j(y_j) \prod_{j>r+s} g_{j,T}(y_j)\, dy = \frac{1}{2\pi}\int_{-\infty}^{\infty} h_1(r)\cdots h_n(r)\Omega_\pi(Tr)^k\, dr. \tag{3.47}$$

On using Stirling's approximation, we get for $|r| \geqslant 1$ that $\Omega_\pi(Tr) = m\log(Tr) + O(1)$, so that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} h_1(r)\cdots h_n(r)\Omega_\pi(Tr)^k\, dr = \frac{1}{2\pi}\kappa(\mathbf{h})L^k + O(L^{k-1}), \tag{3.48}$$

and the lemma follows. □

We can conclude the next lemma from the above.

LEMMA 3.7. *For $r + s > 0$,*

$$C_{r,s}(T) = \frac{1}{2\pi}\kappa(\mathbf{h})\frac{T}{L^{r+s-1}} \sum_{\substack{n_j \ll T \\ n_1\cdots n_r = n_{r+1}\cdots n_{r+s} \\ n_1\cdots n_{r+s} \ll T^2}} \prod_{j=1}^{r} \frac{c(n_j)}{\sqrt{n_j}} \prod_{j=r+1}^{r+s} \frac{\overline{c(n_j)}}{\sqrt{n_j}} \cdot\Phi\left(-\frac{\log n_1}{L},\ldots,\frac{\log n_{r+s}}{L},\ldots,0\right) + O(T). \tag{3.49}$$

Proof. From (3.43) and (3.44), we see that for $\mathbf{n}$ in the range of summation,

$$\tilde{A}_{r,s}(\mathbf{n}, T) = \frac{T}{L^{r+s-1}}\frac{1}{2\pi}\kappa(\mathbf{h})\Phi\left(-\frac{\log n_1}{L},\ldots,\frac{\log n_{r+s}}{L},\ldots,0\right) + O\left(\frac{T}{L^{r+s}}\right), \tag{3.50}$$

and so

(3.51)

$$C_{r,s}(T) = \frac{1}{2\pi}\kappa(\mathbf{h})\frac{T}{L^{r+s-1}}\left\{1 + O\left(\frac{1}{L}\right)\right\}$$

$$\sum_{\substack{n_j \ll T \\ n_1\cdots n_r = n_{r+1}\cdots n_{r+s} \\ n_1\cdots n_{r+s} \ll T^2}} \prod_{j=1}^{r}\frac{c(n_j)}{n_j}\prod_{j=r+1}^{r+s}\overline{c(n_j)}\cdot\Phi\left(-\frac{\log n_1}{L},\ldots,\frac{\log n_{r+s}}{L},\ldots,0\right).$$

We will see below (Lemma 3.9) that the sum is $O(L^{r+s})$, and so once that is established Lemma 3.7 will follow. □

LEMMA 3.8. *Assume $r + s \geqslant 3$. Then*

$$\sum_p \sum_{l_1+\cdots+l_s=k_1+\cdots+k_r}\left|\frac{c(p^{k_1})\cdot\ \cdots\ \cdot c(p^{k_r})\overline{c(p^{l_1})}\cdot\ \cdots\ \cdot\overline{c(p^{l_s})}}{p^{k_1+\cdots+k_r}}\right| < \infty \tag{3.52}$$

(the sums with $k_i, l_j \geqslant 1$).

Proof. We first omit the restriction $\sum k_i = \sum l_j$, set $t = r + s \geqslant 3$, and write $k_{r+1} = l_1, \ldots, k_t = l_s$. The sum (3.52) is bounded by the sum

$$\sum_p \sum_{k_1,\ldots,k_t\geqslant 1}\frac{|c(p^{k_1})\cdots c(p^{k_t})|}{p^{(k_1+\cdots+k_t)/2}}. \tag{3.53}$$

Recall (2.3), which asserts that $a(p^k) \ll p^{k(1/2-\beta)}$ for $\beta = 1/(m^2+1) > 0$. We use this to bound the sum (3.53) when we restrict the exponents to $k_1 + \cdots + k_t \geqslant K > 1/\beta$:

$$\sum_p \sum_{k_1+\cdots+k_t\geqslant K}\frac{|c(p^{k_1})\cdots c(p^{k_t})|}{p^{(k_1+\cdots+k_t)/2}} \ll \sum_p\sum_{k\geqslant K}\frac{p^{k(1/2-\beta)}\log^t p}{p^{k/2}}k^t \tag{3.54}$$

$$\ll \sum_p \frac{\log^t p}{p^{K\beta}} < \infty.$$

To deal with the sum $\sum k_j < K$, we need to use the Rankin-Selberg L-function. It suffices to show that for s fixed, $s < K$, the sum

$$\sum_p \sum_{k_1+\cdots+k_t=s}\frac{|c(p^{k_1})\cdots c(p^{k_t})|}{p^{s/2}} \tag{3.55}$$

is bounded.

Recall that $t \geqslant 3$, and so we may replace all but two of the coefficients $c(p^k)$ by $p^{k(1/2-\beta)} \log p$. Doing this, we find that for fixed p, the summand in (3.55) is bounded by

$$\sum_{k_1+k_2 \leqslant s-(t-2)} \frac{|c(p^{k_1})c(p^{k_2})|}{p^{(k_1+k_2)/2}} \sum_{k_3+\cdots+k_t=s-k_1-k_2} \frac{\log^{t-2} p}{p^{(k_3+\cdots+k_t)\beta}} \tag{3.56}$$

$$\ll_s \sum_{k_1+k_2 \leqslant s-(t-2)} \frac{|c(p^{k_1})c(p^{k_2})|}{p^{(k_1+k_2)/2}} \frac{\log^{t-2} p}{p^{(t-2)\beta}}$$

$$\ll \sum_{1 \leqslant k_1, k_2 \leqslant s} \frac{|c(p^{k_1})|}{p^{(k_1+\delta')/2}} \frac{|c(p^{k_2})|}{p^{(k_2+\delta')/2}}$$

for some $\delta' > 0$. Now fix k_1, $k_2 \leqslant s$ as we may, sum over p, and use Cauchy-Schwarz:

$$\sum_p \frac{|c(p^{k_1})|}{p^{(k_1+\delta')/2}} \frac{|c(p^{k_2})|}{p^{(k_2+\delta')/2}} \ll \left\{\sum_p \frac{|c(p^{k_1})|^2}{p^{k_1+\delta'}}\right\}^{1/2} \left\{\sum_p \frac{|c(p^{k_2})|^2}{p^{k_2+\delta'}}\right\}^{1/2}. \tag{3.57}$$

Each of the sums is now seen to converge by applying **RS 1**. □

We use Lemma 3.8 to deal with sums over several prime factors.

LEMMA 3.9. *If $1 \leqslant r \leqslant s$, then*

$$\sum_{\substack{p_1^{k_1}\cdots p_r^{k_r}=q_1^{l_1}\cdots q_s^{l_s} \\ p_i^{k_i}, q_j^{l_j} \leqslant x}} \frac{c(p_1^{k_1})\cdot\cdots\cdot c(p_r^{k_r})\overline{c(q_1^{l_1})}\cdot\cdots\cdot\overline{c(q_s^{l_s})}}{p_1^{k_1}\cdots p_r^{k_r}} = \begin{cases} O((\log x)^{2r}), & r = s \\ O((\log x)^{2r-2}), & r < s. \end{cases} \tag{3.58}$$

Proof. We divide the sum (3.58) into subsums according to the number of *distinct* prime factors appearing, and we collect together factors corresponding to the same prime. The sum (3.58) then becomes a sum of products of the form

$$\prod_{p_i \neq p_j} \sum_{p_j^{k_1},\ldots,p_j^{k_a} \leqslant x} \sum_{\sum k_i = \sum l_j} \frac{c(p_j^{k_1})\cdot\cdots\cdot c(p_j^{k_a})\overline{c(p_j^{l_1})}\cdot\cdots\cdot\overline{c(p_j^{l_b})}}{p_j^{k_1+\cdots+k_a}} \tag{3.59}$$

where the product is taken over distinct primes. There are at most r factors, and by Lemma 3.8, each factor in (3.59) contributes a *bounded* quantity unless $a = b = 1$, in which case it is bounded by $O(\log^2 x)$. Thus the product (3.59) is $O(\log^{2r-2} x)$ *unless* $r = s$ and $a = b = 1$ in each factor of (3.59). Thus, if $r < s$, then (3.58) is $O(\log^{2r-2} x)$, while for $r = s$ we can relax the condition that the product in (3.59) is over *distinct* primes at the cost of introducing a bounded error in each factor, which after multiplying will introduce an error of $O(\log^{2r-2} x)$. Therefore, we find on using Proposition 2.3 that for $r = s$, (3.58) equals

$$\prod_{j=1}^{r} \sum_{p_j^{k_j} \leqslant x} \frac{|c(p_j^{k_j})|^2}{p_j^{k_j}} + O(\log^{2r-2} x) = \frac{\log^{2r} x}{2^r} + O(\log^{2r-1} x). \quad \square \tag{3.60}$$

Recall from Lemma 3.7 that

$$C_{r,s}(T) = \frac{1}{2\pi} \kappa(\mathbf{h}) \frac{T}{(L)^{r+s-1}} \cdot \sum_{\substack{n_1 \cdots n_r = n_{r+1} \cdots n_{r+s} \\ n_1 \cdots n_{r+s} \ll T^2}} \frac{c(n_1) \cdots c(n_r) \overline{c(n_{r+1})} \cdots \overline{c(n_{r+s})}}{n_1 \cdots n_r} \cdot \Phi\left(-\frac{\log n_1}{L}, \ldots, \frac{\log n_{r+s}}{L}, 0, \ldots, 0 \right) + O(T). \tag{3.61}$$

LEMMA 3.10. (1) *$C_{r,s}(T) = O(T)$ unless $r = s > 0$.*
(2) *If $r = s$, then*

$$C_{r,r}(T) = \frac{TL}{2\pi} \kappa(\mathbf{h}) \cdot \sum_{\sigma \in S_r} \int_0^{1/m} \cdots \int_0^{1/m} v_1 \cdots v_r \cdot \Phi(-v_1, \ldots, -v_r, v_{\sigma(1)}, \ldots, v_{\sigma(r)}, 0, \ldots, 0)\, dv_1 \cdots dv_r + O(T)$$

where S_r is the permutation group on r letters.

Proof. The first statement follows from the upper bound in Lemma 3.9. As for the second, we use summation by parts and the asymptotics in Proposition 2.3: From Lemma 3.8, we know that only *distinct* primes (not higher powers) are going to contribute to the main term. Thus

$$C_{r,r}(T) = \frac{1}{2\pi} \kappa(\mathbf{h}) \frac{T}{(L)^{2r-1}} \cdot \sum_{p_1, \ldots, p_r \leqslant T} \prod_{j=1}^{r} \frac{|c(p_j)|^2}{p_j} \cdot \sum_{q_1 \cdot \cdots \cdot q_r = p_1 \cdot \cdots \cdot p_r} \Phi\left(-\frac{\log p_1}{L}, \ldots, -\frac{\log p_r}{L}, \frac{\log q_1}{L}, \ldots, \frac{\log q_r}{L}, 0, \ldots, 0 \right) + O(T) \tag{3.62}$$

where the inner sum is over primes q_j. Since $p_1, \ldots, p_r$ are *distinct* primes, the primes q_j are then a permutation of p_j. Thus, for some $\sigma \in S_r$, $q_j = p_{\sigma(j)}$,

$j = 1, \ldots, r$. Therefore,

(3.63)

$$C_{r,r}(T) = \frac{1}{2\pi}\kappa(\mathbf{h})\frac{T}{(L)^{2r-1}} \cdot \sum_{p_1,\ldots,p_r \leqslant T} \prod_{j=1}^{r} \frac{|c(p_j)|^2}{p_j}$$

$$\cdot \sum_{\sigma \in S_r} \Phi\left(-\frac{\log p_1}{L}, \ldots, -\frac{\log p_r}{L}, \frac{\log p_{\sigma(1)}}{L}, \ldots, \frac{\log p_{\sigma(r)}}{L}, 0, \ldots, 0\right) + O(T),$$

and we need no longer restrict to summing over distinct primes. The sum in (3.63) can be approximated by using summation by parts in the following form: By Proposition 2.3 (assuming **H**),

$$\sum_{p \leqslant X} \frac{|c(p)|^2}{p} = \frac{1}{2}\log^2 X + O(X).$$

Hence, for reasonable functions ϕ (e.g., a C^1 function),

$$\sum_{p \leqslant T} \frac{|c(p)|^2}{p}\phi\left(\frac{\log p}{L}\right) = \int_0^{1/m} v\phi(v)\, dv \cdot \log^2 T + O(\log T), \tag{3.64}$$

and therefore

$$\begin{aligned} C_{r,r}(T) &= \frac{1}{2\pi}\kappa(\mathbf{h})\frac{T}{(L)^{2r-1}} \cdot (L)^{2r} \sum_{\sigma \in S_r} \int_0^{1/m} \cdots \int_0^{1/m} v_1 \cdots v_r \\ &\quad \cdot \Phi(-v_1, \ldots, -v_r, v_{\sigma(1)}, \ldots, v_{\sigma(r)}, 0, \ldots, 0)\, dv_1 \cdots dv_r + O(T) \\ &= \kappa(\mathbf{h})\frac{T}{2\pi}L \sum_{\sigma \in S_r} \int_0^{1/m} \cdots \int_0^{1/m} v_1 \cdots v_r \\ &\quad \cdot \Phi(-v_1, \ldots, -v_r, v_{\sigma(1)}, \ldots, v_{\sigma(r)}, 0, \ldots, 0)\, dv_1 \cdots dv_r + O(T), \end{aligned} \tag{3.65}$$

as required. □

If we use Lemma 3.10 and bring into account all possible choices of $r = 0, \ldots, [n/2]$ and signs, we obtain Theorem 3.1.

Remark: The support condition $|\xi_1| + \cdots + |\xi_n| < 2/m$. Consider first $m = 1$ and the case of $\zeta(s)$. In this case (and only in this case), there is an added term in the explicit formula (2.16) arising from the pole at $s = 1$, of the form

$$\prod_{j=1}^{n}\left\{h\left(-\frac{i}{2T}\right)T^{-m\xi_j/2} + h\left(\frac{i}{2T}\right)T^{m\xi_j/2}\right\} \sim h(0)^n T^{m(|\xi_1| + \cdots + |\xi_n|)/2}. \tag{3.66}$$

Corresponding to this all the coefficients $\Lambda(n)$ in the sum over primes are nonnegative. From (3.66) it is clear that the region $\sum_j |\xi_j| < 2$, which appears in Theorem 1.2, is precisely that which renders (3.66) smaller than the main term $T \log T$. Indeed, outside this region this polar term will be significant and the central diagonal terms in the proof of Theorem 1.2 are no longer dominant.

In the case of Dirichlet L-functions $L(s, \chi)$, the same is true but no longer because of the (nonexistent) polar term—see below. For $m \geqslant 2$, the region in which we prove Theorem 1.2 is no longer "natural": In the case of the pair correlation function, the relevant sum coming from our analysis is

$$\sum_{n_1, n_2} \frac{c(n_1)}{\sqrt{n_1}} \frac{c(n_2)}{\sqrt{n_2}} g(T(2\xi_1 \log T + \log n_1)) g(T(2\xi_2 \log T - \log n_2)) \hat{f}(\xi_1, \xi_2) \tag{3.67}$$

with $\xi_1 + \xi_2 = 0$. If we assume that $\hat{f}$ is supported near $\xi_1 = -\xi_2 = \xi > 0$, then the diagonal contribution is of size $\log T$. Consider the off-diagonal contribution: These are essentially sums over primes

$$\begin{aligned} OD &:= \frac{\log^2 T}{T^{2\xi}} \sum_{\substack{|\log m/n| \ll 1/T \\ m \neq n \sim T^{2\xi}}} c(m) c(n) \\ &= \frac{\log^2 T}{T^{2\xi}} \sum_{m \sim T^{2\xi}} \sum_{1 \leqslant h \leqslant T^{2\xi - 1}} c(m) c(m+h) \\ &\ll \frac{\log^2 T}{T^{2\xi}} \sum_{1 \leqslant h \leqslant T^{2\xi - 1}} \sum_{m \sim T^{2\xi}} c(m) c(m+h). \end{aligned} \tag{3.68}$$

The inner sum in (3.68) will have substantial cancellation if $a(m)$ are the Fourier coefficients of a cusp form on $GL(2)/\mathbf{Q}$. We expect the size of the inner sum to be of order square root of the number of terms, that is, $O(T^{\xi})$. With the square root saving, we find that the off-diagonal contribution will be bounded by

$$OD \ll \frac{\log^2 T}{T^{2\xi}} \sum_{1 \leqslant h \leqslant T^{2\xi - 1}} T^{\xi} = T^{\xi - 1} \log^2 T. \tag{3.69}$$

Hence, for $\xi < 1$ (i.e., $|\xi_1| + |\xi_2| < 2$), we see that the diagonal still dominates. That is to say that the region should be the same as in the case $m = 1$.

In this respect, we observe an important difference of the above analysis when $m = 1$ and $a(p) = \chi(p)$ is a Dirichlet character, say of conductor Q. Then the sum

$$\sum_{X \leqslant p \leqslant 2X} \Lambda(p) \chi(p) \Lambda(p + 2kQ) \chi(p + 2kQ) \tag{3.70}$$

is of order X, if we assume the "twin-prime" type conjectures of Hardy and Littlewood. Thus the inner sum of (3.68) does not have the same cancellation as for $m=2$. In fact, as with $\zeta(s)$, the region $|\xi_1|+|\xi_2|<2$, valid for $L(s,\chi)$, is the largest in which the diagonal dominates.

Proof of Theorem 3.2. We first show that in Theorem 3.1 we can take $h_j = \chi_{[-1,1]}$ to be the characteristic function of the interval $[-1,1]$. In what follows we assume the Riemann hypothesis for $L(s,\pi)$. In Theorem 3.1 we established that for h_j as in (3.5), and f satisfying (3.6),

$$\frac{1}{N(T)}\sum h_1\left(\frac{\gamma_1}{T}\right)\cdots h_n\left(\frac{\gamma_n}{T}\right)f\left(\frac{L}{2\pi}\gamma_1,\ldots,\frac{L}{2\pi}\gamma_n\right)\to\kappa(\mathbf{h})\mu(f) \tag{3.71}$$

where $N(T)=\#\{0<\gamma_j\leqslant T\}\sim TL/2\pi$, and $\mu(f)$ is the measure given by (3.9). By taking linear combinations, we obtain

$$\frac{1}{N(T)}\sum h\left(\frac{\vec{\gamma}}{T}\right)f\left(\frac{L}{2\pi}\vec{\gamma}\right)\to\kappa(h)\mu(f) \tag{3.72}$$

where $h(\vec{r})$ is a finite linear combination of functions of the form $h_1(r_1)\cdots h_n(r_n)$ and

$$\kappa(h)=\int_{-\infty}^{\infty}h(r,\ldots,r)\,dr. \tag{3.73}$$

To extend the validity of (3.72) further, we note that if $H(\vec{r})$ is piecewise continuous of rapid decrease, then, given $\varepsilon>0$, there are finite linear combinations h_1, h_2 as above so that $h_1\leqslant H\leqslant h_2$ and $\int_{-\infty}^{\infty}(h_2-h_1)(r,\ldots,r)\,dr<\varepsilon$. We use these to show that (3.72) is valid for such H: Indeed, given f as above, we can find an $f_+>0$ with $|f|<f_+$ and f_+ admissible for (3.72) (at least assuming Φ in (3.6) is C^2). If we set

$$D(H,f;T):=\frac{1}{N(T)}\sum H\left(\frac{\vec{\gamma}}{T}\right)f\left(\frac{L}{2\pi}\vec{\gamma}\right)-\kappa(H)\mu(f), \tag{3.74}$$

then

$$\begin{aligned}|D(H,f;T)|&\leqslant|D(h_1,f;T)|+|D(H-h_1,f;T)|\\&\leqslant|D(h_1,f;T)|+\frac{1}{N(T)}\sum(h_2-h_1)\left(\frac{\vec{\gamma}}{T}\right)f_+\left(\frac{L}{2\pi}\vec{\gamma}\right)+\kappa(H-h_1)|\mu(f)|\\&\leqslant|D(h_1,f;T)|+|D(h_2-h_1,f_+;T)|\\&\quad+\kappa(h_2-h_1)\mu(f_+)+\kappa(H-h_1)|\mu(f)|,\end{aligned}$$

and since (3.72) is valid for the first two terms, we find

$$\limsup_T |D(H,f;T)| \leqslant \kappa(h_2-h_1)\mu(f_+) + \kappa(H-h_1)|\mu(f)| < \varepsilon(\mu(f_+)+|\mu(f)|).$$

Since $\varepsilon > 0$ is arbitrary, it follows that $D(H,f;T) \to 0$.

With this approximation argument, we can include many more admissible functions h, and in particular the characteristic function of the cube $[-1,1]^n$. Hence Theorem 3.1 is valid with $h_j(r) = \chi_{[-1,1]}(r)$ or $\chi_{[a,b]}$, for $a < b$.

Next, we need to discuss the passage from the normalization of zeros $(L/2\pi)\gamma$ appearing in the definition of the smooth sums $C_n(f, h, T)$ in (3.3), and the normalization $\tilde{\gamma} = (m/2\pi)\gamma \log|\gamma|$ initially used to define $C_n(f, T)$ in (3.2). We explain it for the pair correlation ($n = 2$). Consider, for $\psi(x) = f(x+y, y)$,

$$\text{(3.75)} \qquad \begin{aligned} \tilde{C}(\psi, T) &= \sum_{T \leqslant \gamma_1, \gamma_2 \leqslant 2T} \psi\left(\gamma_1 \frac{m \log \gamma_1}{2\pi} - \gamma_2 \frac{m \log \gamma_2}{2\pi}\right) \\ C(\psi, T) &= \sum_{T \leqslant \gamma_1, \gamma_2 \leqslant 2T} \psi\left(\frac{L}{2\pi}(\gamma_1 - \gamma_2)\right). \end{aligned}$$

Then we claim that

$$\text{(3.76)} \qquad C(\psi, T) - \tilde{C}(\psi, T) = O(T).$$

This will show that the different normalizations lead to the same main term, and thus prove Theorem 3.2.

To see this, observe that

$$\tilde{C}(\psi, T) = \sum_{T \leqslant \gamma_1, \gamma_2 \leqslant 2T} \psi\left(\frac{L}{2\pi}(\gamma_1 - \gamma_2) + \gamma_1 \frac{m}{2\pi} \log \frac{\gamma_1}{T} - \gamma_2 \frac{m}{2\pi} \log \frac{\gamma_2}{T}\right).$$

Applying the mean value theorem, (3.76) equals

$$2 \sum_{\gamma_1 < \gamma_2} \psi'(\delta_{\gamma_1,\gamma_2,T})\left(\gamma_2 \frac{m}{2\pi} \log \frac{\gamma_2}{T} - \gamma_1 \frac{m}{2\pi} \log \frac{\gamma_1}{T}\right)$$

where $\delta_{\gamma_1,\gamma_2,T}$ is in the interval with endpoints $(m/2\pi)(\gamma_1 \log \gamma_1 - \gamma_2 \log \gamma_2)$ and $(L/2\pi)(\gamma_1 - \gamma_2)$. Thus (since $0 \leqslant \log(\gamma_j/T) \leqslant \log 2$),

$$\text{(3.77)} \qquad |\tilde{C}(\psi, T) - C(\psi, T)| \ll \sum_{T \leqslant \gamma_1 < \gamma_2 \leqslant 2T} |\psi'(\delta_{\gamma_1,\gamma_2,T})| \cdot (\gamma_2 - \gamma_1).$$

For this fixed ψ, let $\psi_1 \geqslant |\psi'|$ be a rapidly decreasing function on $\mathbf{R}$ which is even and *monotone* on $[0, \infty)$. Then, by (3.77),

$$|\tilde{C}(\psi, T) - C(\psi, T)| \ll \frac{1}{L} \sum_{T \leqslant \gamma_1, \gamma_2 \leqslant 2T} \psi_1\left(\frac{L}{2\pi}(\gamma_1 - \gamma_2)\right)\left|\frac{L}{2\pi}(\gamma_1 - \gamma_2)\right|.$$

We can find a majorant ψ_+ and h which are admissible in Theorem 3.1, satisfying $\psi_+(x) \geqslant |x|\psi_1(x)$ and $h \geqslant \chi_{[T, 2T]}$. Then, with these choices, we have

$$|C(\psi, T) - \tilde{C}(\psi, T)| \ll \frac{1}{L} \sum h\left(\frac{\gamma_1}{T}\right) h\left(\frac{\gamma_2}{T}\right) \psi_+\left(\frac{L}{2\pi}\right),$$

and we can bound the sum by $O(TL)$, by Theorem 3.1. Thus we find

$$|C(\psi, T) - \tilde{C}(\psi, T)| \ll \frac{1}{L} TL \ll T,$$

as required. This establishes Theorem 3.2.

4. Combinatorial sieving. In Section 3 we showed that the unrestricted sums $C_n(f, T)$ have a limiting distribution

$$C_n(f, T) := \sum_{i_1, \dots, i_n} f(\tilde{\gamma}_{i_1}, \dots, \tilde{\gamma}_{i_n}) \sim N(T) \cdot \int_{\mathbf{R}^n} \hat{f}(u) C_{\underline{O}}(u)\, du \tag{4.1}$$

where the sum is over *all* indices $(i_1, \dots, i_n)$. However, the n-level correlation function $R_n(f, T)$ is the same sum but over *distinct* indices; it differs from $C_n(f, T)$ by omitting all sums over diagonals $i_j = i_k$, which measure lower-order correlations. In this section we recover $R_n(f, T)$ by a combinatorial sieving.

We begin with some set-theoretic combinatorics. A set partition $\underline{F}$ of $\underline{N} = \{1, \dots, n\}$ is a decomposition of $\underline{N}$ into disjoint subsets $[F_1, \dots, F_\nu]$. The collection Π_n of all set partitions of $\underline{N}$ forms a lattice with the partial ordering given by $\underline{F} \preceq \underline{G}$ if every subset G_i is a union of subsets of $\underline{F}$. The minimal element of Π_n is $\underline{O} = [\{1\}, \{2\}, \dots, \{n\}]$, and the maximal element is $\underline{N} = \{1, 2, \dots, n\}$.

The Möbius function of a poset such as Π_n is the unique function $\mu(x, y)$ so that for any functions $f, g\colon \Pi_n \to \mathbf{R}$, satisfying

$$f(x) = \sum_{x \preceq y} g(y), \tag{4.2}$$

we have

$$g(x) = \sum_{x \preceq y} \mu(x, y) f(y). \tag{4.3}$$

In the case of Π_n, the Möbius function can be computed explicitly [15, §25], in particular,

$$\mu(\underline{O}, \underline{F}) = \prod_{j=1}^{\nu(\underline{F})} (-1)^{|F_j|-1}(|F_j| - 1)!. \tag{4.4}$$

Given a set partition $\underline{F} = [F_1, \dots, F_\nu] \in \Pi_n$, we have an embedding $\iota_{\underline{F}}: \mathbf{R}^\nu \to \mathbf{R}^n$, given by $\iota_{\underline{F}}(x_1, \dots, x_\nu) = (y_1, \dots, y_n)$, where $y_l = x_j$ if $l \in F_j$. For instance, if $\underline{F} = [\{1, 3\}, \{2, 4\}] \in \Pi_4$, then $\iota_{\underline{F}}(x, y) = (x, y, x, y)$.

We can define the lower-order correlations

$$R_{\underline{F}}(f, T) := R_{\nu(\underline{F})}(\iota_{\underline{F}}^* f, T) = \sum_{\substack{i_1, \dots, i_\nu \\ \text{distinct}}} f(\iota_{\underline{F}}(\tilde{\gamma}_{i_1}, \dots, \tilde{\gamma}_{i_\nu})). \tag{4.5}$$

This is the ν-level correlation between n zeros $(\gamma_1, \dots, \gamma_n)$ where we mandate that the indices $i_l = i_k$ are equal if $l, k \in F_j$, and $i_l \neq i_k$ if k, l are in different subsets of $\underline{F}$. Thus if $\underline{F} = \underline{N}$, then $R_{\underline{N}}(f, T) = \sum f(\tilde{\gamma}, \dots, \tilde{\gamma})$ counts the number of zeros up to height T, while if $\underline{F} = \underline{O}$, then $R_{\underline{O}}$ is just the n-level correlation function (1.3) defined in the introduction (after division by N and where $B_N = B_{N(T)}$, $N(T) = \#\{j: |\gamma_j| \leqslant T\}$). We similarly define unrestricted correlation functions by setting

$$C_{\underline{F}}(f, T) := C_\nu(\iota_{\underline{F}}^* f, T) = \sum_{i_1, \dots, i_\nu} f(\iota_{\underline{F}}(\tilde{\gamma}_{i_1}, \dots, \tilde{\gamma}_{i_\nu})) \tag{4.6}$$

so that $C_n(f, T) = C_{\underline{O}}(f, T)$.

Observe that we have an identity

$$C_{\underline{O}}(f, T) = \sum_{\underline{F}} R_{\underline{F}}(f, T) \tag{4.7}$$

or, more generally, for any $\underline{G} \in \Pi_n$,

$$C_{\underline{G}}(f, T) = \sum_{\underline{G} \preceq \underline{F}} R_{\underline{F}}(f, T). \tag{4.8}$$

This is merely partitioning the unrestricted sum for $C_{\underline{O}}$ as a sum over the various possibilities for coincidences between the indices. Thus we can use Möbius inversion to express the n-level correlation function $R_{\underline{O}}(f, T)$ in terms of the unrestricted sums $C_{\underline{F}}(f, T)$:

$$R_{\underline{O}}(f, T) = \sum_{\underline{F}} \mu(\underline{O}, \underline{F}) C_{\underline{F}}(f, T). \tag{4.9}$$

This allows us to find the limiting behaviour of $R_n(f, T)$ by using Theorem 3.2, which yields the limit of $C_{\underline{F}}(f, T)$. To describe the answer, we define δ-functions as follows: For a subset $S \subset \{1, \dots, n\}$, we put

$$\delta_S(u) = \delta\left(\sum_{l \in S} u_l\right), \tag{4.10}$$

where $\delta(x)$ = Dirac delta mass at the origin, and for a set partition $\underline{F} = [F_1, \dots, F_\nu]$ we define

$$\delta_{\underline{F}}(u) = \prod_{j=1}^{\nu(\underline{F})} \delta_{F_j}(u). \tag{4.11}$$

$\delta_{\underline{F}}$ is a delta function supported on the linear subspace

$$U_{\underline{F}} = \left\{u \in \mathbf{R}^n \colon \sum_{l \in F_j} u_l = 0, j = 1, \dots, \nu(\underline{F})\right\}. \tag{4.12}$$

LEMMA 4.1. *With the assumptions as in Theorem 3.2,*

$$C_{\underline{F}}(f, T) = N(T) \int \Phi(u) C_{\underline{F}}(u)\, du + O(T)$$

where $C_{\underline{F}}(u)$ is given by

$$C_{\underline{F}}(u) = \delta_{\underline{F}}(u) + \sum_{r=1}^{[\nu(\underline{F})/2]} \sum \prod_{t=1}^{r} \delta_{F_{i(t)} \cup F_{j(t)}}(u) \left| \sum_{l \in F_{i(t)}} u_l \right| \cdot \prod_{k \neq i(a), j(b)} \delta_{F_k}(u), \tag{4.13}$$

the sum being over all possible choices of r pairs of subsets $(F_{i(t)}, F_{j(t)})$ of $\underline{F}$, no repetitions allowed.

Proof. We first express $\imath_{\underline{F}}^* f = \hat{\Phi}_{\underline{F}}$ as a Fourier transform, in terms of the original Φ such that $f = \hat{\Phi}$:

$$\Phi_{\underline{F}}(v_1, \dots, v_\nu) = \delta(v_1 + \dots + v_\nu) \int_{\mathbf{R}^n} \prod_{j=1}^{\nu} \delta\left(v_j - \sum_{l \in F_j} u_l\right) \Phi(u)\, du. \tag{4.14}$$

Indeed,

$$\begin{aligned}
\imath_{\underline{F}}^* f(x_1, \dots, x_\nu) &= \int_{\mathbf{R}^n} \Phi(\xi) \delta(\xi_1 + \dots + \xi_n) e(-\xi \cdot \imath_{\underline{F}}(x))\, d\xi \\
&= \int_{\xi_1 + \dots + \xi_n = 0} \Phi(\xi) e\left(-\sum_{j=1}^{\nu} x_j \left(\sum_{l \in F_j} \xi_l\right)\right) d\xi \\
&= \int_{v_1 + \dots + v_\nu = 0} \left\{\int \cdots \int_{\sum_{l \in F_j} \xi_l = v_j} \Phi(\xi)\, d\xi\right\} e\left(-\sum_{j=1}^{\nu} x_j v_j\right) dv,
\end{aligned}$$

which proves (4.14). Note that (4.14) implies that if Supp $\Phi \subset \{\sum_{j=1}^{n} |\xi_j| < r\}$, then Supp $\Phi_{\underline{F}} \subset \{\sum_{j=1}^{\nu} |\eta_j| < r\}$.

Now we use Theorem 3.2 to get the asymptotics of $C_{\underline{F}}(f, T)$ (note that the conditions **TF 2, 3** descend to $\iota_{\underline{F}}^* f$):

$$C_{\underline{F}}(f, T) = C_\nu(\iota_{\underline{F}}^* f, T) = N(T) \cdot \int_{\mathbf{R}^\nu} \Phi_{\underline{F}}(v) C_\nu(v)\, dv + O(T).$$

On using (4.14), we find

$$\begin{aligned}\int_{\mathbf{R}^\nu} \Phi_{\underline{F}}(v) C_\nu(v)\, dv &= \int_{\mathbf{R}^\nu} \int_{\mathbf{R}^n} \Phi(u) \prod_{j=1}^{\nu} \delta\left(v_j - \sum_{l \in F_j} u_l\right) C_\nu(v_1, \ldots, v_\nu)\, du\, dv \\ &= \int_{\mathbf{R}^n} \Phi(u) C_\nu\left(\sum_{l \in F_1} u_l, \ldots, \sum_{l \in F_\nu} u_l\right) du.\end{aligned}$$

Now substituting the expression (3.9) for $C_\nu(u)$ in Theorem 3.2 yields (4.13). □

We have seen that

$$R_{\underline{O}}(f, T) = N(T) \int_{\mathbf{R}^n} \Phi(u) R_{\underline{O}}(u)\, du + O(T) \tag{4.15}$$

where

$$R_{\underline{O}}(u) = \sum_{\underline{F}} \mu(\underline{O}, \underline{F}) C_{\underline{F}}(u). \tag{4.16}$$

In view of (4.15), (4.16), and the definition of the GUE determinant $W_n(u)$ (1.5), Theorem 1.2 follows from Theorem 4.1.

THEOREM 4.1. *Let* $W_n(x) = \det(K(x_i - x_j))$, $K(x) = (\sin \pi x / \pi x)$. *Then, for* $\sum_j |u_j| < 2$, *the Fourier transform* $\hat{W}_n(u)$ *is equal to* $R_{\underline{O}}(u)$ *in (4.16).*

Similarly, Theorem 1.1 follows from Theorem 3.1 (with $h_1 = \cdots = h_n = h$) and the above combinatorics.

We will divide the proof of Theorem 4.1 into three propositions.

PROPOSITION 4.1. *We have*

$$\sum_{\underline{F}} \mu(\underline{O}, \underline{F}) C_{\underline{F}}(u) = \sum_{\underline{F}} \delta_{\underline{F}}(u) \prod_{j=1}^{\nu(\underline{F})} X_{F_j}(u) \tag{4.17}$$

where, for any subset $S \subset \underline{N}$,

(4.18)
$$X_S(u) = (-1)^{|S|-1}(|S| - 1)! + (-1)^{|S|-2} \sum_{S = S^+ \cup S^-} (|S^+| - 1)!(|S^-| - 1)! \left| \sum_{l \in S^+} u_l \right|.$$

PROPOSITION 4.2. *The Fourier transform of the GUE determinant*

$$\hat{W}_n(u_1, \ldots, u_n) = \int_{\mathbf{R}^n} \det(K(x_i - x_j)) e\left(\sum_{j=1}^{n} u_j x_j\right) dx$$

is given by

$$\hat{W}_n(u) = \sum_{\underline{F}} \delta_{\underline{F}}(u) \prod_{j=1}^{\nu(\underline{F})} (-1)^{|F_j|-1} Y_{F_j}(u) \tag{4.19}$$

where, for any subset $S \subset \underline{N}$,

$$Y_S(u) = \sum_{(i_1, \ldots, i_m)} \int_{-\infty}^{\infty} f_2(v) f_2(v + u_{i_1}) \cdot \cdots \cdot f_2(v + u_{i_1} + \cdots + u_{i_m})\, dv, \tag{4.20}$$

the sum over all cyclic permutations of S, and

$$f_2(v) = \begin{cases} 1, & |v| \leqslant 1/2 \\ 0, & |v| > 1/2 \end{cases} \tag{4.21}$$

(so that $K(x) = \hat{f}_2(x)$ *is the Fourier transform of* f_2*).*

PROPOSITION 4.3. *For a subset* $S \subset \underline{N}$ *with* $\sum_{l \in S} u_l = 0$, $\sum_{l \in S} |u_l| < 2$, *we have an identity*

$$X_S(u) = (-1)^{|S|-1} Y_S(u). \tag{4.22}$$

By comparing the coefficients of $\delta_{\underline{F}}(u)$ in Propositions 4.1 and 4.2, we see that Proposition 4.3 is exactly what is needed to establish Theorem 4.1. The rest of Section 4 is concerned with proving these propositions.

Proof of Proposition 4.1. It is expedient to introduce some structure on set partitions that corresponds to the combinatorics of pairs appearing in the calculations. To this end, we define a *marking* ϕ of a set partition $\underline{G} = [G_1, \ldots, G_\nu]$ to be a choice of $r \geqslant 0$ pairs of subsets $(G_{i_1}^+, G_{i_1}^-), \ldots, (G_{i_r}^+, G_{i_r}^-)$.

We will denote such a marking by

$$(\underline{G}, \phi) = [G_{i_1}^+, G_{i_1}^-; G_{i_2}^+, G_{i_2}^-; \ldots; G_{i_r}^+, G_{i_r}^-; G_{i_{2r+1}}^o; \ldots, G_{i_\nu}^o].$$

We will allow the *trivial marking* $(\underline{G}, \phi) = [G_1^o; \ldots, G_\nu^o]$.

A marking $(\underline{G}, \phi)$ *reduces* to an unmarked set partition $\underline{F}$ if, for all $j = 1, \ldots, r$, there is a subset F_j of $\underline{F}$ so that $F_j = G_{i_j}^+ \cup G_{i_j}^-$, and likewise for all k there is a subset of $\underline{F}$ so that $G_k^o = F_{i(k)}$. We will denote this symbolically as $\underline{F} = \mathrm{red}(\underline{G}, \phi)$ or $(\underline{G}, \phi) \to \underline{F}$.

Note. A marked set partition reduces to a unique unmarked one; however, there are usually several different marked partitions reducing to the same unmarked one.

With the notion of *marking* of a set partition, we can describe the formula (4.13) for $C_{\underline{G}}(u)$ as

$$C_{\underline{G}}(u) = \sum_{\phi} \delta_{\mathrm{red}(\underline{G},\phi)}(u) \prod_{j=1}^{r} \left| \sum_{l \in G_j^+} u_l \right| \tag{4.23}$$

where the sum is over all possible marking of $\underline{G}$, and for the trivial marking the empty product is interpreted as equal to one. Therefore we find

$$\begin{aligned} \sum_{\underline{G}} \mu(\underline{O}, \underline{G}) C_{\underline{G}}(u) &= \sum_{(\underline{G},\phi)} \mu(\underline{O}, \underline{G}) \delta_{\mathrm{red}(\underline{G},\phi)}(u) \prod_{j=1}^{r} \left| \sum_{l \in G_j^+} u_l \right| \\ &= \sum_{\underline{F}} \delta_{\underline{F}}(u) \sum_{(G,\phi)\to \underline{F}} \mu(\underline{O}, \underline{G}) \prod_{j=1}^{r} \left| \sum_{l \in G_j^+} u_l \right| . \end{aligned} \tag{4.24}$$

Now fix a set partition $\underline{F}$, and consider the coefficient of $\delta_{\underline{F}}$ above. Then we have a factorization identity: If $u \in U_{\underline{F}}$, then

$$\begin{aligned} &\sum_{(G,\phi)\to \underline{F}} \mu(\underline{O}, \underline{G}) \prod_{j=1}^{r} \left| \sum_{l \in G_j^+} \mu_l \right| \\ &\quad = \prod_{j=1}^{\nu(\underline{F})} \Bigg\{ (-1)^{|F_j|-1}(|F_j|-1)! \\ &\qquad + (-1)^{|F_j|-2} \sum_{F_j = G_j^+ \cup G_j^-} (|G_j^+|-1)!(|G_j^-|-1)! \left| \sum_{l \in G_j^+} u_l \right| \Bigg\} . \end{aligned} \tag{4.25}$$

To prove this, just multiply out the right-hand side of (4.25) and compare with (4.4). The factors in the product are exactly what we called X_{F_j} in Proposition 4.1, and so we find, as desired, that

$$\sum_{\underline{G}} \mu(\underline{O}, \underline{G}) C_{\underline{G}}(u) = \sum_{\underline{F}} \delta_{\underline{F}}(u) \prod_{j=1}^{\nu(\underline{F})} X_{F_j}(u). \tag{4.26}$$

This proves Proposition 4.1. □

Proof of Proposition 4.2. We expand the determinant W_n as a sum over all permutations of $\underline{N} = \{1, \ldots, n\}$:

$$W_n(x_1, \ldots, x_n) = \sum_{\sigma \in S_n} (-1)^{\sigma} \prod_{j=1}^{n} K(x_j - x_{\sigma(j)}). \tag{4.27}$$

When taking the Fourier transform of W_n, decompose each permutation as a product over disjoint cycles $\sigma = \tau_1 \cdot \cdots \cdot \tau_r$, with $\tau_j = (i_1, \ldots, i_m)$ a cycle of length $m = m(j)$. Note that $(-1)^\sigma = \prod (-1)^{m(j)-1}$. We then notice that the Fourier transform of each summand in (4.27) breaks up into a product of Fourier transforms of similar expressions over the cycles τ_j:

$$\int \prod_{j=1}^{n} K(x_j - x_{\sigma(j)}) e\left(\sum u_j x_j\right) dx$$

$$= \prod_{j=1}^{r} \int K(x_{i_1} - x_{i_2}) K(x_{i_2} - x_{i_3}) \cdots K(x_{i_m} - x_{i_1}) e(u_{i_1} x_{i_1} + \cdots + u_{i_m} x_{i_m})\, dx.$$

It is therefore sufficient to compute the Fourier transform of each factor separately when we have a *cyclic* permutation.

LEMMA 4.2. *We have*

$$\int K(x_1 - x_2) K(x_2 - x_3) \cdots K(x_m - x_1) e(u_1 x_1 + \cdots + u_m x_m)\, dx$$

$$= \delta(u_1 + \cdots + u_m) \int_{-\infty}^{\infty} f_2(v) f_2(v + u_1) \cdots f_2(v + u_1 + \cdots + u_{m-1})\, dv.$$

Proof. Set $t_j = x_j - x_{j+1}$, $j = 1, \ldots, m-1$. Then, changing variables, we find

$$\int K(x_1 - x_2) K(x_2 - x_3) \cdots K(x_m - x_1) e(u_1 x_1 + \cdots + u_m x_m)\, dx$$

$$= \int K(t_1) \cdots K(t_{m-1}) K(-t_1 \cdots - t_{m-1})$$

$$\cdot e(t_1 u_1 + \cdots + t_{m-1}(u_1 + \cdots + u_{m-1}) + x_m(u_1 + \cdots + u_m))\, dt_1 \cdots dt_{m-1}\, dx_m$$

$$= \delta(u_1 + \cdots + u_m) \int K(t_1) \cdots K(t_{m-1}) K(-t_1 \cdots - t_{m-1})$$

$$\cdot e(t_1 s_1 + \cdots + t_{m-1} s_{m-1})\, dt$$

where we have set $s_k = u_1 + \cdots + u_k$.

Since $K = \hat{f}_2$, we can use Parseval to see

$$\int_{-\infty}^{\infty} K(t_1) K(-\tau - t_1) e(t_1 s_1)\, dt_1 = \int_{-\infty}^{\infty} f_2(v) f_2(v + s_1) e(v\tau)\, dv. \tag{4.28}$$

Applying this with $\tau = t_2 + \cdots + t_{m-1}$, we find

$$\int K(t_1)\cdots K(t_{m-1})K(-t_1\cdots - t_{m-1})e(t_1s_1 + \cdots + t_{m-1}s_{m-1})\,dt$$

$$= \int f_2(v)f_2(v+s_1)e(v(t_2+\cdots+t_{m-1}))K(t_2)\cdots K(t_{m-1})$$

$$\cdot e(t_2s_2 + \cdots + t_{m-1}s_{m-1})\,dt\,dv$$

$$= \int f_2(v)f_2(v+s_1)\prod_{j=2}^{m-1}\int K(t_j)e(t_j(v+s_j))\,dt_j\,dv$$

$$= \int f_2(v)f_2(v+s_1)f_2(v+s_2)\cdots f_2(v+s_{m-1})\,dv. \quad \square$$

To finish the proof of Proposition 4.2, we decompose the permutations into products of disjoint cyclic permutations indexed by set partitions $\underline{F}$ of $\underline{N} = \{1, \ldots, n\}$:

$$S_n = \coprod_{\underline{F}} S^*(F_1) \times \cdots \times S^*(F_\nu)$$

where, for a subset $F_j \subset \underline{N}$, $S^*(F_j)$ denotes the set of all cyclic permutations of the indices in F_j. The sign of any cyclic permutation in $S^*(F_j)$ is $(-1)^{|F_j|-1}$. Proposition 4.2 now follows. $\square$

Proof of Proposition 4.3. It clearly suffices to prove Proposition 4.3 for the case $S = \underline{N}$, which we assume from now on. We will need some preparation: For $u \in \mathbf{R}^n$, with $\sum_j u_j = 0$, and $\sum_j |u_j| < 2$, and an *ordering* $\theta = (\theta(1), \ldots, \theta(n))$ of $\underline{N}$, we define the consecutive partial sums

$$s_k(\theta) = \sum_{j=1}^{k} u_{\theta(j)} = u_{\theta(1)} + \cdots + u_{\theta(k)} \tag{4.29}$$

and let $M(\theta)$, $m(\theta)$ be the maximum (respectively, the minimum) of these partial sums:

$$\begin{aligned} M_u(\theta) &= \max\{s_k(\theta), k = 1, \ldots, n\}, \\ m_u(\theta) &= \min\{s_k(\theta), k = 1, \ldots, n\}. \end{aligned} \tag{4.30}$$

Further, set

$$V(\theta) = M(\theta) - m(\theta). \tag{4.31}$$

When no ordering θ is explicitly given, we just write V for the corresponding quantity.

It is useful to think of $u_{\theta(j)}$ as the increments of a "random walk" on the real line, starting and ending at the origin, and the partial sums $s_k(\theta)$ are the positions after k steps. M and m are the farthest positions to the right (respectively, to the left), and the difference $V(\theta) = M(\theta) - m(\theta)$ is then the maximal deviation of the walk. From this description, it is apparent that $V(\theta) = V(\theta')$ if θ' is a *rotation* of θ, e.g., $\theta' = (\theta(2), \dots, \theta(n), \theta(1))$.

The connection of this to our previous discussion is the following.

LEMMA 4.3. *If* $u_1 + \cdots + u_n = 0$, $\sum_j |u_j| < 2$, *then*

$$\int_{-\infty}^{\infty} f_2(v) f_2(v + u_1) \cdots f_2(v + u_1 + \cdots + u_n)\, dv = 1 - V. \tag{4.32}$$

Proof. Recall that $f_2(v)$ is the characteristic function of the interval $I = [-1/2, 1/2]$. Thus, the integral (4.32) is the length of the intersection of the intervals I, $-s_1 + I$, $\dots$, $-s_{n-1} + I$. This intersection is nonempty if $\sum |u_j| < 2$, in which case it equals the interval $[-m - 1/2, -M + 1/2]$, whose length is $-M + 1/2 - (-m - 1/2) = 1 - V$. $\square$

Set

$$T(u) = \sum_{\{\theta\}}{}' V(\theta), \tag{4.33}$$

the sum being taken over all orderings modulo rotations, i.e., over cyclic permutations of $\{1, \dots, n\}$, of which there are $(n-1)!$. Using Lemma 4.3, we can rewrite $Y_{\underline{N}}(u)$ in (4.20) as

$$Y_{\underline{N}}(u) = (n-1)! - T(u). \tag{4.34}$$

Taking into account the definition of $X_S(u)$ (4.18), in order to prove Proposition 4.3, it suffices to prove the following identity between piecewise linear functions:

$$T(u) = \sum_{[F, F^c]} (|F| - 1)!(n - |F| - 1)! \left| \sum_{l \in F} u_l \right|. \tag{4.35}$$

This we accomplish below by adapting Spitzer's combinatorial method [34].[3]

Both sides of (4.35) are continuous in u, so it suffices to prove (4.35) for u generic, i.e., the components u_i are linearly independent over the rationals. Since

[3] Apparently, this identity is quite old—it can already be found in Kac [12].

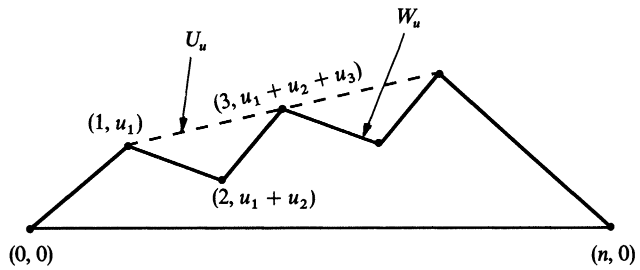

FIGURE 2

$V(\theta)$ is invariant under rotations, we may write $T(u)$ in (4.33) as

$$T(u) = \frac{1}{n}\sum_\theta V(\theta) = \frac{1}{n}\sum_\theta M(\theta) - \frac{1}{n}\sum_\theta m(\theta). \tag{4.36}$$

The sums $\sum_\theta M(\theta)$ and $\sum_\theta m(\theta)$ are easily seen to be negatives of each other, and so

$$T(u) = \frac{2}{n}\sum_\theta M(\theta). \tag{4.37}$$

For $u = (u_1, \ldots, u_r) \in \mathbf{R}^r$, define the (polygonal) walk W_u to be the walk $(0, 0) \to (1, u_1) \to \cdots \to (r, u_1 + \cdots + u_r)$. The *chord* is the segment connecting $(0, 0)$ to $(r, u_1 + \cdots + u_r)$. Define the *upper convex envelope* U_u of the walk W_u to be the lowest convex curve lying above W_u (Figure 2).

LEMMA 4.4 (Spitzer [34, Theorem 2.1]). *Given $u = (u_1, \ldots, u_r)$ in generic position, there is a unique rotation of u so that the walk lies below its chord.*

Given u, W_u, and U_u, let $0 < k_1 < k_2 < \cdots < k_v = r$ be the first coordinates of the vertices of the upper convex walk U_u. Note that the walk W_u restricted to the sets $\{1, \ldots, k_1\}$, $\{k_1 + 1, \ldots, k_2\}, \ldots$ has the following properties.

PROPERTY **SP 1.** *The walks lie below their chords.*

PROPERTY **SP 2.** *Their respective slopes are decreasing, i.e.,*

$$\frac{u_1 + \cdots + u_{k_1}}{k_1} > \frac{u_{k_1+1} + \cdots + u_{k_2}}{k_2 - k_1} > \cdots.$$

Returning to the ordering θ of $\underline{N} = \{1, \ldots, n\}$, and $u \in \mathbf{R}^n$, $\sum_j u_j = 0$, which we assume is generic, we get a convex polygon $U_u(\theta)$ above the walk $W_u(\theta)$. This convex polygon starts at (0, 0) and ends at $(n, 0)$ (Figure 2). It determines intervals as above, $G_1(\theta) = [1, \ldots, k_1]$, $G_2(\theta) = [k_1 + 1, \ldots, k_2]$, ... Each of these comes with an ordering $\psi_j(\theta)$. So, for each ordering θ, we get data $D(\theta) = (G_1(\theta), \psi_1(\theta))$, $(G_2(\theta), \psi_2(\theta))$, ... satisfying **SP 1** and **SP 2**. Moreover,

$$M_u(\theta) = \frac{1}{2} \sum_j \left| \sum_{l \in G_j(\theta)} u_l \right| . \tag{4.38}$$

Conversely, given u generic and a set partition $\underline{G} = [G_1, \ldots, G_\nu]$ of $\underline{N}$ and orderings ψ_j of G_j, we can uniquely arrange them to satisfy **SP 1** and **SP 2** by first arranging their slopes (i.e., **SP 2**) and then using Spitzer's Lemma 4.4 to adjust each ordering ψ_j by a (unique) rotation so as to satisfy **SP 1**. That is, using this bijection and (4.38), we have

(4.39)

$$\sum_\theta M_u(\theta) = \frac{1}{2} \sum_{\underline{G}=[G_1,\ldots,G_\nu]} (|G_1| - 1)! \cdot \cdots \cdot (|G_\nu| - 1)! \left(\left| \sum_{l \in G_1} u_l \right| + \cdots + \left| \sum_{l \in G_\nu} u_l \right| \right)$$

$$= \frac{1}{2} \sum_{F_1 \subset \underline{N}} (|F_1| - 1)! \left| \sum_{l \in F_1} u_l \right| \cdot \left(\sum_{\underline{F}=[F_1, F_2, \ldots, F_\nu]} (|F_2| - 1)! \cdots (|F_\nu| - 1)! \right).$$

For a fixed subset $F_1 \subset \underline{N}$, the innermost sum is clearly equal to the sum over all set partitions of the complement $F_1^c = \underline{N} - F_1$ of

$$\sum_{[F_2,\ldots,F_\nu]} (|F_2| - 1)! \cdots (|F_\nu| - 1)! . \tag{4.40}$$

This counts all the permutations of F_1^c when writing a permutation as a product of disjoint cycles. So the sum (4.40) is simply $|F_1^c|! = (n - |F_1|)!$. Hence

$$\sum_\theta M_u(\theta) = \frac{1}{2} \sum_{F_1 \subset \underline{N}} (|F_1| - 1)!(n - |F_1|)! \left| \sum_{l \in F_1} u_l \right| . \tag{4.41}$$

Grouping together the terms corresponding to F_1 and its complement F_1^c, and using $\sum_j u_j = 0$ (so that $|\sum_{l \in F_1} u_l| = |\sum_{l \in F_1^c} u_l|$), we have

$$\sum_\theta M_u(\theta) = \frac{1}{2} \sum_{[F, F^c]} ((|F| - 1)!(n - |F|)! + (n - |F| - 1)!|F|!) \left| \sum_{l \in F} u_l \right| \tag{4.42}$$

$$= \frac{n}{2} \sum_{[F, F^c]} (|F| - 1)!(n - |F| - 1)! \left| \sum_{l \in F} u_l \right| .$$

Hence, from (4.37),

$$T(u) = \sum_{[F,F^c]} (|F| - 1)!(n - |F| - 1)! \left| \sum_{l \in F} u_l \right|. \tag{4.43}$$

This concludes the proof of (4.35) and so of Proposition 4.3. □

Appendix

A.1. Our goal in this appendix is twofold. Firstly, we describe the local L-factors at ramified places, and secondly we derive the estimates (2.3) and (2.5) for the $\alpha_\pi(j, p)$ and $\mu_\pi(j)$.

Let π_p be an irreducible admissible (generic) representation of $G = GL_m$, over $\mathbf{R}$ or $\mathbf{Q}_p$, with unitary central character. From the description below it follows that if π_p is *tempered*, then $L(s, \pi_p)$ is holomorphic in $\operatorname{Re} s > 0$. The general non-tempered representation π_p can be described as a Langlands quotient: There is a standard parabolic subgroup P of type $(m_1, \dots, m_r)$ (so that the Levi complement $M_P = P/U_P \simeq GL(m_1) \times \cdots \times GL(m_r)$), tempered representations σ_j of $GL(m_j)$, and real numbers $t_1 > \cdots > t_r$ (called the Langlands parameters of π_p) so that

$$\pi_p = J(G, P; \sigma_1[t_1], \dots, \sigma_r[t_r]) \tag{A.1}$$

is the unique irreducible quotient of the induced representation $\operatorname{Ind}(G, P; \sigma_1[t_1], \dots, \sigma_r[t_r])$. Here the twist operation for $t \in \mathbf{C}$ is defined as $\sigma[t] := \sigma \otimes |\det(\cdot)|^t$. If π_p is unitary, then $\{\sigma_j[t_j]\} = \{\tilde{\sigma}_k[-t_k]\}$. In terms of this data, the L-factor is given by

$$L(s, \pi_p) = \prod_{j=1}^{r} L(s + t_j, \sigma_j) \tag{A.2}$$

where the definition of the tempered factors still needs to be given. Indeed, in this case, for σ tempered, there is a standard parabolic P of type $(m_1, \dots, m_r)$ and *square-integrable* representations τ_j of $GL(m_j)$ so that σ is isomorphic to the full induced representation $I(G, P; \tau_1, \dots, \tau_r)$. Then

$$L(s, \sigma) = \prod_{j=1}^{r} L(s, \tau_j). \tag{A.3}$$

This reduces the problem to the case of square-integrable representations. We describe these separately in the p-adic and real setting.

A.2. p-adic factors. For the p-adic case, the square-integrable representations are built out of *supercuspidal* representations as follows: If π is a unitary square-integrable representation of $GL_m(\mathbf{Q}_p)$, then there is a divisor $d|m$, a standard parabolic subgroup P of type $(d, \ldots, d)$, and a unitary supercuspidal representation ρ of GL_d so that π is the unique square-integrable constituent of $\operatorname{Ind}(G, P; \rho_1, \ldots, \rho_n)$, where $n = m/d$, and $\rho_j = \rho \otimes |\det|^{j-(n+1)/2}$, $j = 1, \ldots, n$. We will write $\pi = \Delta(n, \rho)$. We note that the contragredient of such a representation is $\tilde{\Delta}(n, \rho) = \Delta(n, \tilde{\rho})$. Then the principal L-factor is given by $L(s, \pi) = L(s + (n-1)/2, \rho)$. Next, for supercuspidal representations ρ, we have $L(s, \rho) = 1$ unless $m = 1$ (in which case the supercuspidal condition is empty) and $\rho = |\cdot|^t$ is unramified, in which case $L(s, \rho) = (1 - p^{-(s+t)})^{-1}$. This completes the description of the p-adic factors for the principal L-function.

Next, we describe the p-adic factors for the Rankin-Selberg L-function [8]. Let π, π' be irreducible, unitary generic representations of $GL_m(\mathbf{Q}_p)$ and $GL_{m'}(\mathbf{Q}_p)$, respectively. We say two representations π_1, π_2 are in the same twist class if $\deg \pi_1 = \deg \pi_2$ and $\pi_1 = \pi_2 \otimes |\det|^t$. First, if π is supercuspidal, then for $\deg \pi \geqslant \deg \pi'$, the Rankin-Selberg local factor $L(s, \pi \times \pi')$ equals one unless π' is twist-equivalent to $\tilde{\pi}$: $\pi' = \tilde{\pi}[t]$ (in particular, π' is itself supercuspidal). In that case,

$$L(s, \pi \times \tilde{\pi}[t]) = L(s + t, \pi \times \tilde{\pi}) = \frac{1}{1 - p^{-r(t+s)}} \tag{A.4}$$

where $r|\deg \pi$ is the order of the cyclic group of unramified characters $\chi = |\det|^u$ for which $\pi \otimes \chi \simeq \pi$.

Now if $\pi = \Delta(n, \rho)$, $\pi' = \Delta(n', \rho')$ are square-integrable representations of GL_m and $GL_{m'}$ respectively, then

$$L(s, \Delta(n, \rho) \times \Delta(n', \rho')) = \prod_{j=1}^{\min(n,n')} L\left(s + \frac{n+n'}{2} - j, \rho \times \rho'\right). \tag{A.5}$$

In particular, we see that $L(s, \pi \times \tilde{\pi}') = 1$ unless ρ, ρ' are in the same twist class.

For the general π_p given as a Langlands quotient $\pi = J(G, P; \sigma_1[t_1], \ldots, \sigma_r[t_r])$ as in (A.1), we can, by using induction by stages, assume that $\sigma_j = \Delta(n_j, \rho_j)$ and $t_1 \geqslant \cdots \geqslant t_r$. Then we have

$$\begin{aligned} L(s, \pi \times \tilde{\pi}) &= \prod_{j,k=1}^{r} L(s + t_j - t_k, \sigma_j \times \tilde{\sigma}_k) \\ &= \prod_{j,k=1}^{r} \prod_{v=1}^{\min(n_j,n_k)} L\left(s + t_j - t_k + \frac{n_j + n_k}{2} - v, \rho_j \times \tilde{\rho}_k\right). \end{aligned} \tag{A.6}$$

We know that the only factors which contribute are those for which ρ_j, ρ_k are in the same twist class. We can partition the components $\rho_1, \ldots, \rho_r$ into twist classes; there is a set partition $\underline{F} = [F_1, \ldots, F_h]$ of $\{1, \ldots, r\}$ and unitary supercuspidals

ρ^l, $l = 1, \ldots, h$, which are twist-inequivalent, and reals u_j so that for $j \in F_l$, $\rho_j = \rho^l[iu_j]$ is a unitary twist of ρ^l. We then find

(A.7)
$$L(s, \pi \times \tilde{\pi}) = \prod_l \prod_{j,k \in F_l} L(s + t_j - t_k, \Delta(n_j, \rho^l[iu_j]) \times \tilde{\Delta}(n_k, \rho^l[iu_k])) =: \prod_l L_{F_l}(s).$$

We have

$$\text{(A.8)} \qquad L_{F_l}(s) = \prod_{j,k \in F_l} L(s + t_j - t_k, \Delta(n_j, \rho^l[iu_j]) \times \tilde{\Delta}(n_k, \rho^l[iu_k]))$$
$$= \prod_{j,k \in F_l} \prod_{\nu=1}^{\min n_j, n_k} L\left(s + s_j - s_k + \frac{n_j + n_k}{2} - \nu, \rho^l \times \tilde{\rho}^l\right)$$

where $s_j = t_j + iu_j$. We know that $L(s, \rho^l \times \tilde{\rho}^l) = (1 - p^{-r_l s})^{-1}$ for suitable $r_l | \deg \rho^l$ (A.4), and so we find

$$\text{(A.9)} \qquad L_{F_l}(s) = \prod_{j,k \in F_l} \prod_{\nu=1}^{\min(n_j, n_k)} (1 - (p^{\nu - (n_j + n_k)/2 - s_j + s_k} p^{-s})^{r_l})^{-1}.$$

To transform this further, we note that by unitarity, the set $\{s_j\colon j \in F_l\}$ is stable under $s \mapsto -\bar{s}$. Thus

$$\text{(A.10)} \qquad L_{F_l}(s) = \prod_{j,k \in F_l} \prod_{\nu=1}^{\min(n_j, n_k)} (1 - (p^{\nu - (n_j + n_k)/2 - s_j - \bar{s}_k} p^{-s})^{r_l})^{-1}$$
$$= \prod_{j,k \in F_l} \prod_{\nu=1}^{\min(n_j, n_k)} (1 - (p^\nu z_j \bar{z}_k)^{r_l} p^{-r_l s})^{-1}$$

where

$$\text{(A.11)} \qquad z_j = p^{-(n_j/2 + s_j)}, \qquad j \in F_l.$$

To conclude, we see that, with the above notation,

$$\text{(A.12)} \qquad L(s, \pi \times \tilde{\pi}) = \prod_l \prod_{j,k \in F_l} \prod_{\nu=1}^{\min(n_j, n_k)} (1 - (p^\nu z_j \bar{z}_k p^{-s})^{r_l})^{-1}.$$

Therefore, if $t_1 > \cdots > t_r$ are the Langlands parameters of π (A.1), then the local factor $L(s, \pi_p \times \tilde{\pi}_p)$ is holomorphic for $\operatorname{Re} s > 2 \max\{|t_j|\}$ with a pole at $s = 2 \max\{|t_j|\}$. Since it is known that the local factor is holomorphic in $\operatorname{Re} s \geqslant 1$ [8], being the greatest common divisor of local integrals with this property, we see that for π_p generic and unitary, the Langlands parameters satisfy $\max\{|t_j|\} < 1/2$. As a consequence, if we write the principal local factor as $L(s, \pi_p) =$

$\prod_{j=1}^{m}(1-\alpha(p,j)p^{-s})^{-1}$, then we see that $|\alpha(p,j)| \leqslant p^{\max\{|t_j|\}} < p^{1/2}$. This proves (2.2) for all $p < \infty$.

We can expand $L(s, \pi_p \times \tilde{\pi}_p)$ and $L_{F_l}(s)$ in a Dirichlet series

$$L(s, \pi_p \times \tilde{\pi}_p) = \sum_{e=0}^{\infty} \frac{b(p^e)}{p^{es}}$$

(A.13)

$$L_{F_l}(s) = \sum_{e=0}^{\infty} \frac{b_l(p^{er_l})}{p^{er_l s}}.$$

LEMMA A.1. *If $r_l | e$, then*

(A.14)
$$b_l(p^e) \geqslant \frac{1}{e}\left|\sum_{j \in F_l} (p^{(1-n_j)/2-s_j})^e\right|^2.$$

Proof. Take logarithms in (A.10) to find

$$\log L_{F_l}(s) = \sum \frac{p^{-er_l s}}{e} \lambda_l(e)$$

with

$$\lambda_l(e) = \sum_{j,k \in F_l} \sum_{v=1}^{\min(n_j,n_k)} (p^v z_j \bar{z}_k)^{er_l}$$

$$= \sum_{1 \leqslant v \leqslant \max\{n_j : j \in F_l\}} (p^v)^{er_l} \left|\sum_{j \in F_l : n_j \geqslant v} z_j^{er_l}\right|^2.$$

On substituting $z_j = p^{-(n_j/2-s_j)}$, we get

$$\lambda_l(e) = \sum_{1 \leqslant v \leqslant \max\{n_j : j \in F_l\}} \left|\sum_{j \in F_l : n_j \geqslant v} (p^{(v-n_j)/2-s_j})^{er_l}\right|^2$$

$$\geqslant \left|\sum_{j \in F_l} (p^{(1-n_j)/2-s_j})^{er_l}\right|^2.$$

Exponentiating, we find that

$$b_l(p^{er_l}) \geqslant \frac{\lambda_l(e)}{e} \geqslant \frac{1}{e}\left|\sum_{j \in F_l} (p^{(1-n_j)/2-s_j})^{er_l}\right|^2. \quad \square$$

As a consequence of (2.19) and (A.14), we see that the $L(s, \pi \times \tilde{\pi})$ has nonnegative coefficients $b(n) \geqslant 0$.

A.3. Archimedean factors. Any irreducible unitary representation of $G = GL_m(\mathbf{R})$ is given as a Langlands quotient

$$\pi_\infty = J(G, P; \sigma_1[s_1], \ldots, \sigma_r[s_r]) \tag{A.15}$$

where σ_j are square-integrable representations and $\operatorname{Re} s_1 \geqslant \cdots \geqslant \operatorname{Re} s_j$. For the real case, $GL_m(\mathbf{R})$ does not have square-integrable representations if $n \geqslant 3$. To describe the principal L-factors in the remaining cases $n = 1, 2$, we define Gamma factors by

$$\begin{aligned} \Gamma_{\mathbf{R}}(s) &= \pi^{-s/2}\Gamma\left(\frac{s}{2}\right) \\ \Gamma_{\mathbf{C}}(s) &= 2(2\pi)^{-s}\Gamma(s). \end{aligned} \tag{A.16}$$

Note that the duplication formula reads $\Gamma_{\mathbf{C}}(s) = \Gamma_{\mathbf{R}}(s)\Gamma_{\mathbf{R}}(s + 1)$. In the case $n = 1$ (when the condition to be square-integrable is vacuous), the unitary representations are of the form $\pi(x) = |x|^t$ or $\pi(x) = \operatorname{sign}(x)|x|^t = |x|^{t+1}x^{-1}$, with $t \in i\mathbf{R}$. We then set $L(s, \pi_\infty) = \Gamma_{\mathbf{R}}(s + t)$ in the former case, and $L(s, \pi_\infty) = \Gamma_{\mathbf{R}}(s + t + 1)$ in the latter. In the case $n = 2$, the unitary square-integrable representations are unitary twists of the kth discrete series D_k, $k \geqslant 2$ (these correspond to holomorphic forms of weight k), for which the L-factor is given by $L(s, D_k) = \Gamma_{\mathbf{C}}(s + (k - 1)/2)$.

To summarize, for $p = \infty$ the local factor has the following form: We can write $m = r_1 + 2r_2$, and there are complex numbers $s_j = t_j + iu_j$, $j = 1, \ldots, r_1 + r_2$ satisfying $\sum_{j=1}^{r_1} t_j + 2\sum_{j=1}^{r_2} t_j = 0$ and integers $k_j \geqslant 2$, $j = 1, \ldots, r_2$ so that

$$L(s, \pi_\infty) = \prod_{j=1}^{r_1} \Gamma_{\mathbf{R}}(s + s_j) \prod_{j=1}^{r_2} \Gamma_{\mathbf{C}}\left(s + s_{r_1+j} + \frac{k_j - 1}{2}\right). \tag{A.17}$$

We can rewrite this using the duplication formula as a product of m real Gamma factors:

$$L(s, \pi_\infty) = \prod_{j=1}^{m} \Gamma_{\mathbf{R}}(s + \mu_\pi(j)). \tag{A.18}$$

The Rankin-Selberg local factors are defined in terms of L-factors attached to representations of the Weil group $W_{\mathbf{R}}$ by means of the Langlands correspondence, and $L(s, \pi_\infty \times \tilde{\pi}_\infty)$ can be computed from knowledge of a few basic cases. We will list them and use this knowledge to derive (2.5). Suppose that π_∞ is a unitary irreducible generic representation of $GL_m(\mathbf{R})$. If we exhibit π_∞ as a Langlands quotient $J(G, P; \sigma_1[s_1], \ldots, \sigma_r[s_r])$ as in (A.15), then

$$L(s, \pi_\infty \times \tilde{\pi}_\infty) = \prod_{j,k} L(s + s_j - s_k, \sigma_j \times \tilde{\sigma}_k). \tag{A.19}$$

It therefore suffices to check the following cases:

(1) $L(s, \sigma \times 1) = L(s, \sigma)$;

(2) $L(s, \text{sign} \times \text{sign}) = L(s, 1) = \Gamma_{\mathbf{R}}(s)$;

(3) $D_k \otimes \text{sign} \simeq D_k$ and so $L(s, D_k \times \text{sign}, s) = L(s, D_k) = \Gamma_{\mathbf{C}}(s + (k-1)/2)$;

(4) if $k_1 \geqslant k_2$, then $L(s, D_{k_1} \times D_{k_2}) = \Gamma_{\mathbf{C}}(s + (k_1 - k_2)/2)\Gamma_{\mathbf{C}}(s + (k_1 + k_2)/2 - 1)$.

In particular, $L(s, \pi_\infty \times \tilde{\pi}_\infty)$ has its first pole at $s = 2\max|t_j|$, $t_j = \operatorname{Re} s_j$. On the other hand, if π_∞ is the local component of a cuspidal automorphic representation π, then according to **RS 3** of Section 2, $L(s, \pi_\infty \times \tilde{\pi}_\infty)L(s, \pi \times \tilde{\pi})$ is holomorphic for $\operatorname{Re} s > 1$ with a simple pole at $s = 1$ coming from the Euler product $L(s, \pi \times \tilde{\pi})$. Thus $L(s, \pi_\infty \times \tilde{\pi}_\infty)$ is holomorphic for real $s \geqslant 1$, and hence $2\max|t_j| < 1$. That is, in the notations of (A.18), $\operatorname{Re} \mu_\pi(j) > -1/2$, proving (2.5).

Remark. We have given a global proof of (2.5) in order to avoid the complications involved in carrying out the analogue of the p-adic proof via the archimedean Rankin-Selberg integrals [11]. A purely local proof of (2.5) is given in [1] by going through Vogan's classification of the unitary dual of $GL_m(\mathbf{R})$ and checking which representations are generic. On the other hand, the improvements of the bounds (2.2) and (2.5), described below, do require global considerations.

A.4. Global bounds. The first improvement is for the finite places and follows a well-known argument [31].

PROPOSITION A.1. *Let π be an irreducible cuspidal automorphic representation of $GL_m/\mathbf{Q}$, and let π_p be a local constituent of π. Then*

$$|\alpha_\pi(p, j)| \leqslant p^{1/2 - 1/(m^2+1)}. \tag{A.20}$$

Proof. This is merely an application of Landau's theorem [13]. Indeed, we have shown that the series $L(s, \pi \times \tilde{\pi}) = \sum b(n)n^{-s}$ has nonnegative coefficients, and by **RS 3** a meromorphic continuation with Gamma factors and poles exactly at $s = 0, 1$. In Landau's notation, we have $\beta = 1$ and $\eta = m^2/2$, and his result gives

$$\sum_{n \leqslant x} b(n) = Ax + O_\varepsilon(x^{(2\eta-1)/(2\eta+1)+\varepsilon})$$

for all $\varepsilon > 0$. Moreover, since $b(n) \geqslant 0$, it follows that

$$b(n) \ll_\varepsilon n^{(2\eta-1)/(2\eta+1)+\varepsilon}.$$

For π_p unramified, in view of (2.19), this means that for $e \geqslant 1$,

$$\left| \sum_{j=1}^{m} \alpha_\pi(p, j)^e \right|^2 = |a_\pi(p^e)|^2 \leqslant eb(p^e) \ll_\varepsilon ep^{e((2\eta-1)/(2\eta+1)+\varepsilon)}.$$

This clearly implies that

$$|\alpha_\pi(p, j)| \leqslant p^{(1/2)((2\eta-1)/(2\eta+1))} = p^{1/2-1/(m^2+1)}.$$

For ramified primes we find from (A.14) that for all l, if $r_l|e$, then

$$\left|\sum_{j\in F_l}(p^{(1-n_j)/2-s_j})^e\right|^2 \leqslant eb(p^e) \ll_\varepsilon ep^{2e(1/2-1/(m^2+1)+\varepsilon)},$$

and therefore we get a bound on the Langlands parameters

$$\max_j |t_j| \leqslant \frac{1}{2} - \frac{1}{m^2+1}.$$

In view of the formula (A.2) for $L(s, \pi_p)$, we find that $|\alpha_\pi(p, j)| \leqslant p^{1/2-1/(m^2+1)}$ in the ramified case as well. □

As for the analogous improvement at infinity of (2.5), it is shown in [16] by a quite different method that for π_∞ spherical, we have exactly as in the finite places

$$|\mathrm{Re}\, \mu_\pi(j)| \leqslant \frac{1}{2} - \frac{1}{m^2+1}. \tag{A.21}$$

Acknowledgement. We would like to thank S. Miller for a careful reading of the manuscript and F. Shahidi for information concerning Rankin-Selberg L-functions.

References

[1] L. Barthel and D. Ramakrishnan, *A nonvanishing result for twists of L-functions of GL(n)*, Duke Math. J. **74** (1994), 681–700.
[2] H. Davenport, *Multiplicative Number Theory*, 2nd ed., Graduate Texts in Math. **74**, Springer-Verlag, New York, 1980.
[3] F. J. Dyson, *Statistical theory of the energy levels of complex systems, III*, J. Math. Phys. **3** (1962), 166–175.
[4] I. M. Gelfand and D. Kazhdan, "Representation of the group $GL(n, K)$ where K is a local field" in *Lie Groups and their Representations*, ed. by I. M. Gelfand, Wiley, New York, 1974, 95–118.
[5] R. Godement and H. Jacquet, *Zeta Functions of Simple Algebras*, Lecture Notes in Math. **260**, Springer-Verlag, Berlin, 1972.
[6] D. A. Hejhal, *On the triple correlation of zeros of the zeta function*, Internat. Math. Res. Notices **1994**, 293–302.
[7] H. Jacquet, "Principal L-Functions of the Linear Group" in *Automorphic Forms, Representations, and L-Functions*, part 2, Proc. Sympos. Pure Math. **33**, American Math. Soc., Providence, 1979, 63–86.
[8] H. Jacquet, I. I. Piatetski-Shapiro, and J. A. Shalika, *Rankin-Selberg convolutions*, Amer. J. Math. **105** (1983), 367–464.
[9] ———, *Conducteur des représentations du groupe linéaire*, Math. Ann. **256** (1981), 199–214.

[10] H. JACQUET AND J. A. SHALIKA, *On Euler products and the classification of automorphic representations I*, Amer. J. Math. **103** (1981), 499–558.

[11] ———, "Rankin-Selberg convolutions: Archimedean theory" in *Festschrift in Honor of I. I. Piatetski-Shapiro on the Occasion of his Sixtieth Birthday*, Israel Math. Conf. Proc. **2**, Weizmann, Jerusalem, 1990, 125–208.

[12] M. KAC, *Toeplitz matrices, translation kernels, and a related problem in probability theory*, Duke Math. J. **21** (1954), 501–509.

[13] E. LANDAU, *Über die Anzahl der Gitterpunkte in gewisser Bereichen*, (Zweite Abhandlung) Gött. Nach. (1915), 209–243.

[14] R. P. LANGLANDS, "Problems in the theory of automorphic forms" in *Lectures in Modern Analysis and Applications*, Lecture Notes in Math. **170**, Springer-Verlag, New York, 1970, 18–86.

[15] J. H. VAN LINT AND R. M. WILSON, *A Course in Combinatorics*, Cambridge University Press, Cambridge, 1992.

[16] W. LUO, Z. RUDNICK, AND P. SARNAK, *On Selberg's eigenvalue conjecture*, Geom. Funct. Anal. **5** (1995), 387–401.

[17] ———, *On the "Ramanujan conjectures" for GL(m)*, in preparation.

[18] M. L. MEHTA, *Random Matrices*, 2nd ed., Academic Press, Boston, 1991.

[19] C. MŒGLIN AND J.-L. WALDSPURGER, *Le Spectre Résiduel de GL(n)*, Ann. Sci. École Norm. Sup. (4) **22** (1989), 605–674.

[20] H. L. MONTGOMERY, "The pair correlation of zeros of the zeta function" in *Analytic Number Theory*, Proc. Sympos. Pure Math. **24**, Amer. Math. Soc., Providence, 1973, 181–193.

[21] A. M. ODLYZKO, *On the distribution of spacings between zeros of the zeta function*, Math. Comp. **48** (1987), 273–308.

[22] ———, *The 10^{20} zero of the Riemann zeta function and 70 million of its neighbors*, preprint, A.T.&T., 1989.

[23] I. I. PIATETSKI-SHAPIRO, "Euler subgroups" in *Lie Groups and their Representations*, ed. by I. M. Gelfand, Halsted, New York, 1975, 597–620.

[24] ———, *Arithmetic Dirichlet series: conjectures*, proceedings of conference in honor of G. Freiman, CIRM (Merseille, 1993), to appear.

[25] B. RIEMANN, *Über die Anzahl der Primzahlen unter einer gegebenen Größe*, Montasb. der Berliner Akad. (1858/60) 671–680; in *Gessamelte Mathematische Werke*, 2nd ed., Teubner, Leipzig, 1982, # VII.

[26] Z. RUDNICK AND P. SARNAK, *The n-level correlations of zeros of the zeta function*, C. R. Acad. Sci. Paris Sér. I Math. **319** (1994), 1027–1032.

[27] R. RUMELY, *Numerical computations concerning the ERH*, Math. Comp. **61** (1993), 415–440.

[28] P. SARNAK, Course notes, Princeton University, 1995.

[29] A. SELBERG, "Old and new conjectures and results about a class of Dirichlet series" in *Collected Papers, Vol. 2*, Springer-Verlag, Berlin, 1991, 47–65.

[30] J.-P. SERRE, *Abelian ℓ-adic Representations and Elliptic Curves*, W. A. Benjamin, New York, 1968.

[31] ———, Letter to J.-M. Deshouillers, 1981.

[32] F. SHAHIDI, *On certain L-functions*, Amer. J. Math. **103** (1981), 297–355.

[33] J. SHALIKA, *The multiplicity one theorem for GL(n)*, Ann. of Math. (2) **100** (1974), 171–193.

[34] F. SPITZER, *A combinatorial lemma and its applications to probability theory*, Trans. Amer. Math. Soc. **82** (1956), 323–339.

RUDNICK: RAYMOND AND BEVERLEY SACKLER SCHOOL OF MATHEMATICAL SCIENCES, TEL AVIV UNIVERSITY, 69978 TEL AVIV, ISRAEL; AND DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08544, USA

SARNAK: DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08544, USA