Topics in Probabilistic Number Theory



Valeriya Kovaleva
Queen's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2022

In this work we explore three distinct problems on the interface of number theory and random matrix theory. First, we compute the average of two shifted squares of the Riemann zeta on the critical line with shifts up to size $T^{6/5-\varepsilon}$. We give an explicit expression for such averages as well as the $(2,2)$-moment of moment of the Riemann zeta. Our result confirms a conjecture of Chandee as well as a special case of Motohashi's integral. Our approach relies on correlations of the divisor function. Next, we consider random symmetric matrices with independent entries distributed according to the Haar measure on $\mathbb{Z}_p$ for odd primes p. We derive the distribution of the canonical form for such matrices with respect to $\mathbb{Q}_p$ and $\mathbb{Z}_p$-equivalence using a simple combinatorial argument inspired by random matrix theory. Finally, jointly with Ofir Gorodetsky we prove that if X is a random matrix distributed according to the Haar measure on the finite general linear group $\mathrm{GL}_n(\mathbb{F}_q)$, then any non-trivial linear combination of $\mathrm{Tr}(X^1), \mathrm{Tr}(X^2), \dots, \mathrm{Tr}(X^k)$ has an asymptotically uniform distribution on $\mathbb{F}_q$ as $n \to \infty$ as long as $\log k = o(n)$. Our method is inspired by a classical result of Montgomery and Vaughan on the cancellation in character sums over the integers under the Riemann Hypothesis.

Statement of Originality

I confirm that the work in this thesis has not been submitted for any other academic award. The work presented is my own unless otherwise stated.

Acknowledgements

I am very grateful to my supervisor Jon Keating for his support, guidance, and patience. I am deeply thankful for the unending support of my partner Roman, who helped me get through the hardest of times. Finally, I would like to thank all my friends and family who kept giving contradictory advice so I could find my own way.

Contents

1 Background **1**

2 Riemann Zeta on the critical line **6**

2.1 Introduction 6

2.2 Preliminaries 15

2.2.1 Oscillatory integrals 15

2.2.2 Fourier and Mellin transforms 16

2.2.3 Basic spectral theory 18

2.3 Conjectures and predictions 20

2.3.1 General conjecture 20

2.3.2 Motohashi's integral 24

2.3.3 (2,2)-moment of moment 26

2.4 Groundwork 34

2.4.1 Approximate functional equation 36

2.4.2 Simplifying weights 38

2.4.3 Correlations of the divisor function 44

2.5 Explicit expression 47

2.5.1 Small shifts 48

2.5.2 Large shifts 56

2.5.3 Diagonal term and main theorem 64

2.6 Discussion and prospective work 66

3 Equivalence classes of quadratic forms **68**

3.1 Introduction 68

3.2 Preliminaries 74

3.2.1 General matrices 75

3.2.2 Quadratic forms 78

3.3 Distribution of equivalence classes of symmetric matrices 80

3.3.1 $\mathbb{Z}_p$-equivalence 80
3.3.2 $\mathbb{Q}_p$-equivalence 86
3.4 Applications 91
3.4.1 Determinant and rank 91
3.4.2 Local-global problems 94
3.4.3 Isotropy problem 99
3.5 Discussion and prospective work 104

4 Traces of random matrices over finite fields **106**
4.1 Introduction 106
4.1.1 Symmetric function analogue 108
4.1.2 Character sums 109
4.2 Preliminaries 111
4.2.1 Hayes characters 111
4.2.2 Basic bounds 114
4.3 From traces to character sums 115
4.3.1 Proof of Theorem 4.1.2 116
4.3.2 Proof of Theorem 4.1.1 117
4.3.3 Proof of Theorem 4.1.3 118
4.4 Discussion and prospective work 121

Bibliography **122**

Notation

$f \ll g$ or $f = O(g)$	$\lvert f(x)\rvert \leq C\lvert g(x)\rvert$ for some $C > 0$; $f \gg g$ is the same as $g \ll f$
$f \ll_\alpha g$ or $f = O_\alpha(g)$ and similar	Implied constant depends on parameter α
$f \asymp g$ or $f = \Theta(g)$	$f \ll g$ and $g \ll f$
$f = o(g)$	$f(x)/g(x) \to 0$ as $x \to \infty$
$f \sim g$	$f(x)/g(x) = 1 + o(1)$
$\mathrm{GL}_n(R)$	General linear group of $n \times n$ matrices over ring R
I_n	$n \times n$ identity matrix
$\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$	Real and complex normal random variables with expectation μ and variance σ^2
$\mathbb{N}$	Set of positive integers
$\mathbb{F}_q$	Finite field of q elements
$\mu(n)$	Möbius function
$d(n)$	Divisor function $d(n) = \sum_{k\mid n} 1$ for $n \in \mathbb{N}$
$\sigma_y(n)$	Sum-of-divisors function $\sigma_y(n) = \sum_{k\mid n} k^y$ for $n \in \mathbb{N}$ and $y \in \mathbb{C}$
$\tau_{(\alpha_1,\ldots,\alpha_k)}(n)$	Generalised divisor function $\tau_{(\alpha_1,\ldots,\alpha_k)}(n) = \sum_{n_1\cdots n_k = n} n_1^{\alpha_1}\cdots n_k^{\alpha_k}$ for $n \in \mathbb{N}$ and $\alpha_j \in \mathbb{C}$
$F[X]$	Set of univariate polynomials in X with coefficients in F
$[(z-a)^n]\, f(z)$	n-th coefficient in the power series of a function $f(z)$ at $z = a$
$\mathcal{M}_{n,q}$ and $\mathcal{M}_q$	Sets of monic polynomials with coefficients in $\mathbb{F}_q$ of degree n, and of all degrees respectively
$\mathcal{P}_{n,q}$ and $\mathcal{P}_q$	Sets of irreducible polynomials with coefficients in $\mathbb{F}_q$ of degree n, and of all degrees respectively

Chapter 1

Background

Number Theory is a vast subject. Classically it is divided into three major subareas: analytic, algebraic and Diophantine geometry. There are also subareas that emerged very recently, and one example is probabilistic number theory. But what does probabilistic mean in this context? An obvious almost tautological answer would be applying tools from probability theory to problems in number theory. One could also argue that as important as tools are connections, ideas, intuition.

In this work we focus on drawing connections between number theory and probability theory in a very general sense to understand problems in arithmetic statistics, character sums, and the behaviour of the Riemann zeta function $\zeta(s)$. For this purpose random matrix theory provides a very rich toolbox through its connections to physics, combinatorics, representation theory, group theory, and Gaussian multiplicative chaos (GMC). These tools have already led to a number of breakthroughs over the last few years, and will certainly continue to help make progress on long-standing problems as well as extend our understanding of existing results.

One of the most important topics in analytic number theory is the value distribution of $\zeta(s)$ on the critical line $\operatorname{Re} s = 1/2$. For example, a deep open problem in the area is the Lindelöf hypothesis stating that

$$\zeta(1/2 + it) = O(t^{\varepsilon}) \quad \text{for any } \varepsilon > 0\,.$$

This statement is equivalent to proving the same upper bound for all moments

$$\frac{1}{T}\int_0^T |\zeta(1/2 + it)|^k dt\,.$$

Random matrix theory has played a significant role in developing our understanding of the size of $\zeta(1/2 + it)$ and the statistics of its zeros on the critical line. A classic

example of utilising random matrix theory as a guiding tool is the analogy between the determinant of random unitary matrices and ζ on the critical line.

The Hilbert–Pólya conjecture states that non-trivial zeros of ζ should behave like eigenvalues of some self-adjoint operator. This idea is further supported by Montgomery's pair correlation conjecture [101] after Dyson's observation that the pair correlation functions of the zeros of the Riemann zeta suggested by Montgomery and the eigenvalues of a random Hermitian matrix are in fact the same. In particular, let γ_n be the height of the n-th zero of the Riemann zeta, and let

$$w_n = \frac{\gamma_n}{2\pi} \log \frac{\gamma_n}{2\pi} .$$

Let $\delta(x)$ be the Dirac delta. Montgomery proved that for a bump function $f(x)$ such that $\hat{f}$ is supported on $|x| < 1$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n,m\leq N} f(w_n - w_m) = \int_{-\infty}^{+\infty} f(x) \left(\delta(x) + 1 - \left(\frac{\sin(\pi x)}{\pi x} \right)^2 \right) dx ,$$

and conjectured that for fixed $\alpha, \beta > 0$

$$\lim_{T\to\infty} \frac{1}{T} \#\{w_n, w_m \in [0;T] : \alpha \leq w_n - w_m \leq \beta\} = \int_{\alpha}^{\beta} \left(\delta(x) + 1 - \left(\frac{\sin(\pi x)}{\pi x} \right)^2 \right) dx ,$$

whereas the corresponding statement for random matrices does indeed hold due to Dyson [50].

Another matrix model exhibiting the same eigenvalue pair correlation function is the Circular Unitary Ensemble, which stands for random matrices drawn uniformly from the unitary group. However, the Riemann zeta and the unitary ensemble might share statistics other than just the pair correlation function. Keating and Snaith [92] conjectured that one such statistic could be moments, thus predicting the exact asymptotic of the moments of the Riemann zeta and other families of L-functions. Despite much work, this conjecture has only been verified in very few cases [66, 81], which predate the work of Keating and Snaith. Independent heuristic number theoretic arguments have later produced the same prediction.

Let τ be the uniform random variable on the interval $[T, 2T]$. A classical result of Selberg [124, 125] known as Selberg's Central Limit Theorem (Selberg's CLT) states that

$$\frac{\log \zeta(1/2 + i\tau)}{\sqrt{\frac{1}{2} \log\log T}} \to \mathcal{N}_{\mathbb{C}}(0, 1)$$

in distribution as $T \to \infty$. This was originally proved using the method of moments, and for a more recent proof see [119].

Let $P_N(A, \theta)$ be the characteristic polynomial of a random unitary $N \times N$ matrix at $e^{i\theta}$. Keating and Snaith [92] proved that

$$\frac{\log P_N(A, \theta)}{\sqrt{\frac{1}{2}\log N}} \to \mathcal{N}_{\mathbb{C}}(0, 1)$$

in distribution as $N \to \infty$. This further shows that $\zeta(1/2 + i\tau)$ and $P_N(A, \theta)$ indeed behave quite similarly.

Another recent development that proved useful here is the framework of log-correlated fields. A log-correlated field is a family of random variables $\{X(v)\}_{v \in S}$ on a metric space S such that the correlation between two variables $X(v)$ and $X(v')$ is essentially guided by $\log|v - v'|$. This is not an exact statement, but rather a qualitative one, and it is meant to describe the general behaviour of a field "in the bulk" as well as the type of the singularity when $v = v'$ ("logarithmic singularity"). This framework allows us to predict, for example, the maximum size of random variables in this family. Both the logarithm of the characteristic polynomial of a random unitary matrix, and $\log \zeta(1/2 + i\tau)$ have log-correlated structure, as well as a classic random field in statistical mechanics called the Gaussian free field. Thus we expect $P_N(A, \theta)$, $\log \zeta(1/2 + i\tau)$ and the exponential of the Gaussian free field known as the Gaussian Multiplicative Chaos (GMC) to behave similarly. For a detailed exposition on this topic see Arguin [1] or Bailey and Keating [7].

Using this framework, Fyodorov, Hiary, and Keating [60] conjectured that if τ is a uniform random variable on $[T, 2T]$ and c is fixed, then

$$\max_{h \in [-c,c]} \log|\zeta(1/2 + i\tau + ih)| = \log\log T - \frac{3}{4}\log\log\log T + E(T)\,,$$

where $E(T)$ is bounded in probability, and even more precisely is expected to behave like a sum of two independent Gumbel random variables. Note that while the first order term is very much consistent with Selberg's central limit theorem, the coefficient of the second order would have been different if not for log-correlations.

The work on this conjecture and its analogue for $P_N(A, \theta)$ has been going on somewhat in parallel. The first order for $\log|\zeta|$ was established independently by Arguin, Belius, Bourgade, Radziwiłł and Soundararajan [3] and Najnudel [108]. Harper [70] obtained the upper bound up to second order term without the error term, and Arguin, Bourgade and Radziwiłł [4] established the upper bound on $\max \log|\zeta|$ in strong form and gave an explicit upper bound for the error term (in probability).

In Chapter 2 we present a problem closely related to these ideas. Following the work of Bailey and Keating, let $c = c(T)$ and define the $(k, 2\gamma)$-moment of moment of the Riemann zeta function as the average of short averages of length $2c$ on the interval $[0, T]$

$$\mathcal{M}_{k,2\gamma}(c, T) := \frac{1}{T}\int_0^T \left(\frac{1}{2c}\int_{-c}^{c} |\zeta(1/2 + it + ih)|^{2\gamma} dh\right)^k dt\,.$$

This average is closely connected to the distribution of $\max_{h\in[-c,c]} \log|\zeta(1/2+i\tau+ih)|$. One may study this distribution using the partition function of $\log|\zeta(1/2 + i\tau + ih)|$ defined as

$$Z(\gamma) := \frac{1}{2c}\int_{-c}^{c} |\zeta(1/2 + it + ih)|^{2\gamma} dh$$

so $\mathcal{M}_{k,2\gamma}(c, T) = \mathbb{E}Z(\gamma)^k$ is the moment generating function of $Z(\gamma)$. Following [60, 61], Bailey and Keating [6] conjecture that with proper smoothing $\mathcal{M}_{k,2\gamma}(c, T)$ is polynomial in $\log T$ with a given leading order for $k, \gamma \in \mathbb{N}$. This conjecture does not extend to all $k, \gamma > 0$, and this is also predicted by the connection with the critical multiplicative chaos.

In this work we establish the full asymptotic for $\mathcal{M}_{k,2\gamma}(c, T)$ when $\gamma = 1$, $k = 2$ and $c(T) \ll T^{6/5-\varepsilon}$. Curiously, though the leading order agrees with the Bailey-Keating conjecture, the full asymptotic is not in fact polynomial in $\log T$ when $c(T)$ is a constant, and likely one cannot obtain a power saving unless there is some additional smoothing.

More generally, consider the average of four zetas with different shifts

$$\int_0^T \zeta(\frac{1}{2} + it + i\alpha_1)\zeta(\frac{1}{2} + it + i\alpha_2)\zeta(\frac{1}{2} - it - i\beta_1)\zeta(\frac{1}{2} - it - i\beta_2)dt\,.$$

Analysing the Dirichlet series, one can obtain a prediction for its asymptotic. Motohashi [105] proved that this prediction holds for shifts up to size $1/\log T$. This prediction also generalises to an arbitrary number of shifts using the principles from random matrix theory (the "recipe" [39]). Whether the prediction holds for larger shifts is still an open problem, which largely boils down to obtaining a good error term for correlations of generalised divisor functions. The only well-studied case is the case of two shifts [9]. We consider a particular case of this average which can be understood as the correlation of squares of the Riemann zeta,

$$\int_0^T |\zeta(\frac{1}{2} + it + i\alpha)|^2|\zeta(\frac{1}{2} + it + i\beta)|^2 dt\,,$$

and find the asymptotic for the widest range of shifts given the current error term for the correlations of the divisor function with a power saving error.

As we allude above, there are other ways to apply ideas from random matrix theory to number theory. Many classical arithmetic problems ranging from the elementary ones such as the density of square-free numbers to more difficult such as the density of primes, can be extended to integer matrices, see for example [100, 98, 23]. Arithmetic problems over higher dimensions are typically much more difficult. Indeed, the Bateman-Horn conjecture predicting the density of numbers giving prime values of multivariate polynomials is very much open. However, one may be able to utilise the symmetry of certain polynomials such as the determinant of a matrix. Some special cases of polynomials are in fact more approachable, see Bhargava [11]. Further, there are problems over integer matrices such as isotropy for symmetric matrices [15] where the local-global principle applies, and in that case one may ask if there is a unified approach to random matrices over the reals and the p-adic numbers.

An alternative motivation to studying random integer matrices comes from algebraic number theory. In particular, there is a series of works [56, 99, 116, 16, 115, 113, 133] studying the connection between the distribution of cokernels of random matrices and the Cohen-Lenstra heuristics and ranks of elliptic curves.

Finally, in Chapter 4 we present yet another angle of connecting random matrix theory and number theory. A classical result of Diaconis and Shahshahani states that traces of a random unitary matrix are asymptotically independent and normally distributed. One may ask if a similar statement holds for random invertible matrices over finite fields, and what are the limitations in this case. The first rather recent result in this direction is due to Gorodetsky and Rodgers [64] who proved that traces $\mathrm{Tr}(A^k)$ of uniform random matrices $A \in \mathrm{GL}_n(\mathbb{F}_q)$ indeed converge to the uniform distribution for $k \leq C(q)n$. In this chapter we discuss our joint work with Ofir Gorodetsky devoted to studying error terms for character sums parametrised by random matrices from $\mathrm{GL}_n(\mathbb{F}_q)$. In a distinct manner from the first two problems, we take a problem from random matrix theory and tackle it with a purely number theory argument. In particular, we reduce our initial question to a question about cancellation in character sums over function fields. We obtain cancellation for twisted sums of Hayes characters in quite a wide range thus extending the result of Gorodetsky and Rodgers to $\log k = o(n)$. Our method is inspired by a classical result of Montgomery and Vaughan [102], and the work of Granville and Soundararajan [65].

Chapter 2

Riemann Zeta on the critical line

In this chapter we compute the average of two shifted squares of the Riemann zeta on the critical line with shifts up to size $T^{6/5-\varepsilon}$. We give an explicit expression for such averages as well as the $(2,2)$-moment of moment of the Riemann zeta. Our result confirms a conjecture of Chandee as well as a special case of Motohashi's integral.

2.1 Introduction

The behaviour of the Riemann zeta function on the critical line $\{1/2+it : t \in \mathbb{R}\}$ is one of the major topics of interest in analytic number theory. Behaviour in this context allows for numerous interpretations varying from the Lindelöf hypothesis stating that

$$\zeta(1/2+it) \ll_\varepsilon t^\varepsilon \quad \forall \varepsilon > 0$$

to more probabilistic problems such as studying the distribution of

$$\max_{|h|\leq c} |\zeta(1/2+i\tau+ih)|\,,$$

where τ is a random variable on the interval $[T, 2T]$. These interpretations are also quite different in the level of difficulty. The Lindelöf hypothesis currently does not seem accessible via modern techniques with the best known upper bound due to Bourgain [22] with

$$\zeta(1/2+it) \ll t^{13/84+\varepsilon}\,.$$

One can also state the Lindelöf hypothesis through the moments, or averages of the Riemann zeta. Let $k \geq 0$, then the $2k$-moment of the Riemann zeta function on the critical line is the average of $|\zeta(1/2+it)|^{2k}$ on the interval $[0,T]$

$$\mathcal{M}_{2k}(T) := \frac{1}{T}\int_0^T |\zeta(1/2+it)|^{2k} dt\,.$$

Then the Lindelöf hypothesis is equivalent to

$$\mathcal{M}_{2k}(T) \ll_{k,\varepsilon} T^{\varepsilon} \quad \forall \varepsilon > 0,\ k \in \mathbb{N}\,.$$

Moreover, it is believed that an even stronger statement might hold. Namely, that

$$\mathcal{M}_{2k}(T) = P_{k^2}(\log T/2\pi) + O(T^{-a+\varepsilon})\,, \tag{2.1}$$

where P_{k^2} is a polynomial of degree k^2 and $a > 0$. Keating and Snaith [92] conjectured that the leading coefficient of P_k equals $a_k g_k$, where

$$a_k = \prod_p \left(\left(1 - \frac{1}{p}\right)^{k^2} \sum_{j=0}^{\infty} \frac{\Gamma^2(j+k)}{j!^2\Gamma^2(k)} p^{-j} \right)$$

comes from primes, and

$$g_k = \prod_{j=0}^{k-1} \frac{j!}{(j+k)!}$$

comes from random matrix theory. Later Conrey, Farmer, Keating, Rubinstein and Snaith [40] extended this conjecture to predict lower order terms as well.

This conjecture has only been verified for $k = 1, 2$. In particular, Hardy and Littlewood [66] determined the asymptotic for $\mathcal{M}_2(T)$, Ingham [81]– for $\mathcal{M}_4(T)$, and various authors over the years worked on obtaining the full form of P_k with an appropriate error term. Conjectures for $k = 3, 4$ have been obtained independently using a number-theoretic heuristic by Conrey and Ghosh [41], and Conrey and Gonek [42]. Ng [109] and more recently Ng, Shen and Wong [110] gave a rigorous justification of this heuristic for $k = 3, 4$ conditionally on an asymptotic for divisor function correlations with a certain error size. Conrey and Keating devoted a series of papers [33, 34, 35, 36, 37] to developing the number theory heuristics for all k. Recently Baluyot and Conrey [8] gave a different expression for the heuristic suggested by Keating and Conrey using Mellin transforms.

Unconditional lower bounds for all rational $k \geq 0$ are due to Heath-Brown [74], for all real $k > 1$ are due to Radziwiłł and Soundararajan [118], and for all real $k \in (0, 1)$ due to Heap and Soundararajan [72]. Harper [68] obtained a matching upper bound conditionally on the Riemann Hypothesis improving on the work of Soundararajan [126] for all real $k \geq 0$.

Additionally, there are two types of partial results towards higher moments. The first one corresponds to twisted moments of the Riemann zeta

$$\int_0^T |\zeta(1/2 + it)|^4 |P(1/2 + it)|^2 dt\,,$$

where $P(1/2+it)$ is a Dirichlet polynomial of length T^a, and a is as large as possible. Watt's mean value theorem [131] implies that this integral is $O(T^{1+\varepsilon})$ for $a < 1/4$. Further, Hughes and Young [79] computed its asymptotic with a power saving when $a < 1/11$, and Bettin, Bui, Li and Radziwiłł [10] extended the range to $a < 1/4$ by utilising Watt's work [131]. Such integrals also come up when estimating the proportion of zeros of the Riemann zeta lying on the critical line, see Bui, Conrey and Young [27]. For related results concerning other families of L-functions see Hough [77], Zacharias [134], and Kowalski, Michel, and VanderKam [96].

The second type of partial results relates to Dirichlet L-functions rather than the Riemann zeta. Namely, for $k > 2$ one may prove an analogue of (2.1) for Dirichlet L-functions in an extended family, where in addition to averaging over characters mod q we average over the conductor and possibly over t. Some results in this direction are due to Conrey, Iwaniec and Soundararajan [43], and Chandee and Li [31].

When it comes to random variables, behaviour does not necessarily mean size. One may also study correlations, which is a generalisation of the moments, or

$$\mathcal{M}_{2k}(T;\alpha,\beta) := \frac{1}{T}\int_0^T \prod_{j=1}^{k} \zeta(1/2+it+i\alpha_j)\zeta(1/2-it-i\beta_j)dt \tag{2.2}$$

with shifts $\alpha = (\alpha_1,\dots,\alpha_k)$ and $\beta = (\beta_1,\dots,\beta_k)$. Such integrals are also called shifted averages of the Riemann zeta, and studying them gives us a better structural understanding of the Riemann zeta. The formula for $k = 1$ is due to Ingham [81] and Bettin [9], and $k = 2$ for shifts up to $1/\log T$ is due to Motohashi [105]. The general conjecture for all k is due to Conrey, Farmer, Keating, Rubinstein and Snaith [39]. In particular, the conjecture states that

$$\begin{aligned}\mathcal{M}_{2k}(f;\alpha,\beta) := \int_0^{+\infty} f(t)\prod_{j=1}^{k} \zeta(1/2+it+i\alpha_j)\zeta(1/2-it-i\beta_j)dt = \\ \int_0^{\infty} f(t)Q(t;\alpha,\beta)dt + E_f\,,\end{aligned} \tag{2.3}$$

where f is an appropriate weight function, and $Q(t;\alpha,\beta)$ is a certain multiple contour integral in variables $z_1,\dots,z_{2k}$ involving $\zeta(1+z_i-z_j)$ and the Vandermonde determinant $\prod_{i<j}(z_i-z_j)$. The error term E_f is conjectured to be of size $\ll T^{-a}$ for some $a > 0$ when f is normalised and concentrated on the interval $[0;T]$. This is expected to hold for all $\alpha_j,\beta_j \ll T^{2-\varepsilon}$. We also expect that $Q(t;0,0)$ is a polynomial of degree k^2 in $\log t/2\pi$, but in general the explicit expression for $Q(t;\alpha,\beta)$ is more involved.

Tackling such averages in full generality has proven to be very difficult as this involves correlations of generalised divisor functions $\tau_{(l_1,\dots,l_k)}(n) = \sum_{n_1\cdots n_k=n} n_1^{l_1}\cdots n_k^{l_k}$

$$\sum_{n\le X} \tau_{-i\alpha}(n)\tau_{i\beta}(n+r)\,.$$

Except for the case $\alpha=\beta=0$, we are currently unable to obtain an asymptotic with a power saving, possibly due to the lack of analytic continuation of the corresponding Estermann zeta function.

One of the particular cases more approachable for analysis is when $\alpha=(\alpha_1,\dots,\alpha_1)$ and $\beta=(\beta_1,\dots,\beta_1)$. Then with a slight abuse of notation we are now interested in

$$\mathcal{M}_{2k}(T;\alpha,\beta) := \frac{1}{T}\int_0^T |\zeta(1/2+it+i\alpha)|^k|\zeta(1/2+it+i\beta)|^k dt\,,$$

where $\alpha,\beta\in\mathbb{R}$. In this case, Chandee [30] conjectured that

$$\mathcal{M}_{2k}(T;\alpha,\beta)\begin{cases} \asymp_k \log^{k^2} T, & \text{if } \lim_{T\to\infty}|\alpha-\beta|\log T = 0\,;\\ \asymp_{k,c} \log^{k^2} T, & \text{if } \lim_{T\to\infty}|\alpha-\beta|\log T = c\neq 0\,;\\ \asymp_k \frac{1}{|\alpha-\beta|^{k^2/2}}\log^{k^2/2} T, & \text{if } \lim_{T\to\infty}|\alpha-\beta|\log T = \infty\,. \end{cases} \tag{2.4}$$

In the same work Chandee obtained the corresponding lower bound unconditionally, and recently Ng, Shen and Wong [111] proved that Chandee's conjecture holds under the Riemann Hypothesis.

Averages of this type also occur when studying the maximum. Suppose one would like to study the distribution of

$$\max_{h\in[-c,c]} \log|\zeta(1/2+i\tau+ih)|\,, \tag{2.5}$$

where τ is a random variable, and c is a constant which may depend on T. One way to go about such a problem is to study the partition function of $\log|\zeta(1/2+i\tau+ih)|$

$$Z(\beta) := \frac{1}{2c}\int_{-c}^{c} e^{2\gamma\log|\zeta(1/2+it+ih)|}dh = \frac{1}{2c}\int_{-c}^{c}|\zeta(1/2+it+ih)|^{2\gamma}dh\,.$$

Then moments of moments of the Riemann zeta

$$\mathcal{M}_{k,2\gamma}(T) := \frac{1}{T}\int_0^T\left(\frac{1}{2c}\int_{-c}^{+c}|\zeta(1/2+it+ih)|^{2\gamma}dh\right)^k dt = \mathbb{E}Z(\gamma)^k$$

give the moment generating function of $Z(\gamma)$. When $k\in\mathbb{N}$,

$$\mathcal{M}_{k,2\gamma}(T) = \frac{1}{T(2c)^k}\int_{[-c;c]^k}\int_0^T |\zeta(1/2+it+ih_1)|^{2\gamma}\cdots|\zeta(1/2+it+ih_k)|^{2\gamma}dt\,dh_1\cdots dh_k$$

is a k-fold average of length $2c$ of a shifted average.

This statistic is more sensitive to the behaviour of the Riemann zeta on short intervals. For example, it exhibits a phase transition which you cannot see when studying $\mathcal{M}_{2k}$. In particular, Arguin, Ouimet, Radziwiłł [5] showed that if τ is a uniform random variable on $[T, 2T]$, then with probability $1 - o(1)$ as $T \to \infty$

$$\int_{-1}^{1} |\zeta(1/2 + i\tau + ih)|^{2\gamma} dh = \begin{cases} (\log T)^{\gamma^2 + o(1)}, & \text{if } \gamma \leq 1\,; \\ (\log T)^{2\gamma - 1 + o(1)}, & \text{if } \gamma > 1\,. \end{cases} \tag{2.6}$$

Their result is in fact stated for averages of length up to $\log^{\theta} T$. Recently Arguin and Bailey [2] established new bounds for large deviations in Selberg's CLT and impoved some of the results in [5].

Although rigorous analysis of $Z(\gamma)$ would be quite challenging, heuristic arguments have managed to shed light on the behaviour of the maximum. In particular, using the framework of log-correlated fields Fyodorov, Hiary, and Keating [60] made a very precise conjecture on (2.5), which was later partially verified by Najnudel [108]; Arguin, Belius, Bourgade, Radziwiłł, and Soundararajan [3]; and Arguin, Bourgade, Radziwiłł [4]. Although these particular works did not utilise the moments of moments, $\mathcal{M}_{k,2\gamma}$ are still interesting in their own right.

Following the work of Fyodorov, Hiary, and Keating, and Fyodorov and Keating, Bailey and Keating [6] conjectured that for $k, \gamma \in \mathbb{N}$

$$\mathcal{M}_{k,2\gamma}(T) = a_{k,\gamma} c_{k,\gamma} \left(\log \frac{T}{2\pi} \right)^{k^2\gamma^2 - k + 1} \left(1 + O_{k,\gamma}\left((\log T)^{-1} \right) \right), \tag{2.7}$$

where $a_{k,\gamma}$ is an Euler product, and $c_{k,\gamma}$ comes from random matrix theory. When $k = 1$, the asymptotic is the same as for $\mathcal{M}_{2\gamma}(T)$. We might also expect a stronger statement similar to (2.1) with

$$\mathcal{M}_{k,2\gamma}(T) = a_{k,\gamma} c_{k,\gamma} P_{k^2\gamma^2 - k + 1}(\log T/2\pi) + O(T^{-a})\,.$$

Note that this conjecture does not extend to all $k, \gamma > 0$. This is related to the phase transition at $k\gamma^2 = 1$ as in (2.6). This phenomenon, also observed in random matrix theory and Gaussian Multiplicative Chaos, may be occurring due to the log-correlated structure of $\log \zeta(1/2 + it)$. Namely, let $h_1, h_2 \in \mathbb{R}$ with $d = |h_1 - h_2|$, then

$$\frac{1}{T} \int_{T}^{2T} \log |\zeta(1/2 + it + ih_1)| \log |\zeta(1/2 + it + ih_2)| dt =$$

$$\begin{cases} \frac{1}{2} \log \log T + O(1), & \text{if } d \ll \frac{1}{\log T}\,; \\ \frac{1}{2} \log \frac{1}{d} + O(1), & \text{if } \frac{1}{\log T} \ll d \ll 1\,; \\ O(1/d), & \text{if } d \gg 1 \end{cases}$$

with the most interesting regime being $1/\log T \ll d \ll 1$. For the proof of this fact see Bourgade [21] or Harper [67].

Properly defined log-correlated probabilistic objects [1] form a universality class. This means that in certain ways they are very similar to each other, and one example of such similarity is the behaviour of their maxima. Another property they have in common is the fact that $\gamma = 1$ is a special value corresponding to the critical case (the point of phase transition). Using this connection and the ideas from Critical Multiplicative Chaos, Harper [69] showed that for $k \in [0, 1]$

$$\frac{1}{T}\int_T^{2T}\left(\int_{-1/2}^{1/2}|\zeta(1/2+it+ih)|^2dh\right)^k \ll \left(\frac{\log T}{1+(1-k)\sqrt{\log\log T}}\right)^k$$

as opposed to expected $(\log T)^k$. Even though this result was very unexpected on the number theory side, it is entirely consistent with the framework of log-correlated fields.

In this chapter we consider

$$\mathcal{M}_4(T,\alpha,\beta) = \frac{1}{T}\int_0^T |\zeta(1/2+it+i\alpha)|^2|\zeta(1/2+it+i\beta)|^2dt$$

for $\alpha, \beta \in \mathbb{R}$. This problem is related to both general moment calculations, and trying to understand the structural properties of the Riemann zeta. On the one hand we have the conjectures of Motohashi and Chandee for such an average. On the other hand, this average is related to the critical case of $\gamma = 1$, and is not technically included in the conjecture of Bailey and Keating.

Theorem 2.1.1 (Kovaleva)**.** *Let $T \geq 10$ be large, let α and β be positive real numbers. Let $\delta = \beta - \alpha$ and $\tilde{T} = \sqrt{(T+\alpha)(T+\beta)}$. Let $\gamma_0 := [z^0]\zeta(1+z)$ be the Euler-Mascheroni constant, and set*

$$h(z,s) := \frac{1}{\zeta(2+2s)}\left(\zeta(1+z+s)\zeta(1-z+s) - \frac{2s\zeta(1+2s)}{s^2-z^2}\right).$$

Then

$$\frac{1}{T}\int_0^T |\zeta(1/2+it+i\alpha)|^2|\zeta(1/2+it+i\beta)|^2dt = \frac{1}{T}\int_0^T Q(t;\alpha,\beta)dt + \frac{E_{\alpha,\beta}(T)}{T},$$

where $Q(t;\alpha,\beta) = D(t;\alpha,\beta) + OD(t;\alpha,\beta)$ with

$$D(t;\alpha,\beta) = 2\,\mathrm{Re}\left(\operatorname*{Res}_{s\in\{0,i\delta\}}\frac{\zeta^4(1+s)}{\zeta(2+2s)}\frac{(T/2\pi)^s}{s-i\delta}\right);$$

$$OD(t;\alpha,\beta) = \frac{d^2}{ds^2}\left(h(i\delta,s)\frac{(t+\alpha)^{s/2}(t+\beta)^{s/2}}{(2\pi)^s}e^{2\gamma_0 s}\right)\Bigg|_{s=0} - \frac{1}{4}h(i\delta,0)\log^2\frac{t+\alpha}{t+\beta};$$

and

$$E_{\alpha,\beta}(T) \ll T^{\varepsilon} \begin{cases} \tilde{T}^{4/5} + \delta^{11/4}, & \text{if } \delta \ll \tilde{T}^{1/3}\,; \\ \tilde{T}\delta^{-1/4} + \delta^{1/3}\tilde{T}^{1/2} + \delta^{5/6}, & \text{otherwise}\,. \end{cases}$$

Corollary 2.1.2. *The formula for Motohashi's integral* (2.3)*, and Chandee's conjecture* (2.4) *for $k = 2$ hold in the range allowed by Theorem 2.1.1.*

Integrating the expression in Theorem 2.1.1, one can also obtain the expression for the $(2,2)$-moment of the Riemann zeta

$$\mathcal{M}_{2,2}(T; g) = \frac{1}{T}\int_0^T \left(\int_{-\infty}^{+\infty} g(h)|\zeta(1/2 + it + ih)|^2 dh\right)^2 dt$$

for an appropriate weight function g. Normally we choose $g(x) = c^{-1}g_0(x/c)$, where g_0 is piece-wise smooth and rapidly decaying outside of the interval $[-1, 1]$. In the classic setup g_0 is either the indicator function of an interval or $g_0(x) = e^{-|x|}$, but sometimes it is convenient to choose a Schwarz function.

Corollary 2.1.3 ($(2,2)$-moment of moment)**.** *Let $T \geq 10$, and let $c > 0$ be either constant, or growing with T such that $c = c(T) \ll T^{1-\varepsilon}$. Let $N = \log\frac{T}{2\pi}$, and let $g(x) = c^{-1}g_0(x/c)$, where $g_0(x)$ is a fixed normalised even weight function rapidly decaying outside of $[-1, 1]$. Let*

$$P_3(x) := \operatorname*{Res}_{s=0} \frac{\zeta^4(1+s)e^{sx}}{\zeta(2+2s)(1+s)} = b_0x^3 + b_1x^2 + b_2x + b_3$$

be a polynomial of degree 3 in x, and

$$a_j(g) := \int_{-\infty}^{+\infty} \frac{d^j}{ds^j}\left(\frac{h(it, s)}{1+s}\right)\Big|_{s=0} (g * g)(t)dt \ll 1\,.$$

Then

$$\mathcal{M}_{2,2}(T; g) = D(T; g) + OD(T; g) + E_c + O(\max_{w \gg cN} |\hat{g}(w)|^2),,$$

where

$$\begin{aligned} D(T; g) &= 2\int_0^N \hat{g}^2(y/2\pi)P_3(N - y)dy\,; \\ OD(T; g) &= a_0(g)(N + 2\gamma_0)^2 + 2a_1(g)(N + 2\gamma_0) + a_2(g)\,; \\ E_c &\ll T^{-1+\varepsilon}\begin{cases} T^{11/12}, & \text{if } c \ll T^{1/3}\,; \\ c + Tc^{-1/4} + c^{1/3}T^{1/2}, & \text{otherwise}\,. \end{cases} \end{aligned}$$

Behaviour of $\mathcal{M}_{2,2}(T; g)$ depends on the rate of decay of $\hat{g}$, which in turn depends on both the rate of growth of $c(T)$ and the rate of decay of $\hat{g}_0$. Let us first discuss the classic case when g is the indicator function on the interval.

Corollary 2.1.4. *In the notation of Corollary 2.1.3 let* $g(x) = \mathbf{I}\{|x| \le c\}/2c$, *then*

$$D(T;g) = \frac{\pi}{c}P_3(N) - \frac{1}{c^2}P_3'(N)(\log(2cN) + \gamma_0) + \frac{1}{2c^2}(3b_0N^2 - 2b_2) - \frac{b_0}{4c^4} - \frac{b_3}{c^2N}\,.$$

If c *is fixed, then*

$$OD(T;g) = \tilde{a}_0(c)(N + 2\gamma_0)^2 + 2\tilde{a}_1(c)(N + 2\gamma_0) + \tilde{a}_2(c)\,,$$

where $\tilde{a}_j(c) = a_j(g) \neq 0$ *are independent of* T. *If* $c(T) \to \infty$ *as* $T \to \infty$, *then*

$$OD(T;g) = (N + 2\gamma_0 - 1)^2(1 + O(c^{-1+\varepsilon})) \sim \mathcal{M}_2(T)^2\,.$$

When c is fixed two notable things occur. First, this is not strictly speaking a polynomial in $\log T/2\pi$ because of $\log\log T$-factors despite what one may conjecture using random matrix theory. The leading term however still agrees with the conjecture of Bailey and Keating (2.7) exactly. Second, $b_3 \neq 0$, so the error term

$$-\frac{b_3}{c^2\log\frac{T}{2\pi}} + O\left(\frac{1}{c^2\log^2\frac{T}{2\pi}}\right)$$

is certainly not a power saving unless $c \gg T^a$. Another way to see this is as follows. Using Perron's formula the diagonal contribution can essentially be written as

$$\frac{1}{T}\sum_{n\le T/2\pi} d^2(n)\int_1^{\frac{T}{2\pi n}} \hat{g}^2\left(\frac{1}{2\pi}\log w\right)dw = \frac{1}{2\pi i}\int_{(1)}\frac{\zeta^4(1+s)}{(s+1)\zeta(2+2s)}\left(\frac{T}{2\pi}\right)^s \mathcal{L}\hat{g}^2(s)ds\,,$$

where $\mathcal{L}f(s) = \int_0^\infty f(x)e^{-sx}dx$ is the Laplace transform of f. With our choice of $g(x) = \mathbf{I}\{|x| \le c\}/2c$ say with $c = \pi$

$$\mathcal{L}\hat{g}^2(s) = -\frac{1}{4\pi^2}\left((s+2\pi i)\log(s+2\pi i) + (s-2\pi i)\log(s-2\pi i) - 2s\log s\right)$$

has branch cuts, and the branch cut coming from $\log s$ is the source of both of the issues above. For a general function g_0, at least naively, if $\mathcal{L}\hat{g_0}^2(s)$ cannot be extended analytically beyond $\operatorname{Re} s \ge 0$, we cannot use the standard trick for obtaining a power saving by moving the line of integration, and integrating along branch cuts generally does not yield the same result.

Corollary 2.1.5. *In the notation of Corollary 2.1.3 let* $g(x) = c^{-1}g_0(x/c)$, *where* g_0 *is a normalised smooth even function such that* $\hat{g}_0(y) \ll e^{-Ay}$ *for some* $A > 0$. *Then*

$$D(T;g) = \sum_{j=0}^{3}\frac{(-1)^j}{c^{j+1}j!}P_3^{(j)}(N)\int_0^\infty y^j\hat{g}_0^2(y/2\pi)dy + O(N^3e^{-4\pi cAN})\,.$$

If c is fixed, then

$$OD(T;g) = \tilde{a}_0(c)(N+2\gamma_0)^2 + 2\tilde{a}_1(c)(N+2\gamma_0) + \tilde{a}_2(c)\,,$$

where $\tilde{a}_j(c) = \tilde{a}_j(c,g_0) = a_j(g) \neq 0$ are independent of T. If $c(T) \to \infty$ as $T \to \infty$, then

$$OD(T;g) = (N+2\gamma_0-1)^2(1+O(c^{-1+\varepsilon})) \sim \mathcal{M}_2(T)^2\,.$$

In particular, when c is fixed, $\mathcal{M}_{2,2}(T;g)$ is indeed a polynomial of degree 3 with a power saving.

Finally, we would like to make the following observation. When working with the Riemann zeta on the critical line, one may sometimes like to think that correlations between $\zeta(1/2+iT+ih_1)$ and $\zeta(1/2+iT+ih_2)$ only matter when $|h_1-h_2| \ll 1/\log T$. This is of course untrue both factually and morally as the log-correlated structure beyond this regime manifests majorly in short interval problems. Let $\theta > 0$, and set $c(T) = \pi \log^{\theta}(T)$. Then as $T \to \infty$

$$\mathcal{M}_{2,2}(T;\mathbb{I}) \sim \frac{1}{\pi^2}\log^{3-\theta} T + \log^2 T$$

with coefficients independent of θ. The off-diagonal term becomes the main term once $\theta > 1$, and we observe that

$$\mathcal{M}_{2,2}(T;g) \sim \mathcal{M}_2(T)^2\,.$$

Thus, decorrelation between $\zeta(1/2+iT+ih_1)$ and $\zeta(1/2+iT+ih_2)$ *on average* occurs when $|h_1-h_2| \gg \log T$.

Now let us outline the structure of the chapter. We start with presenting some basic tools and theory we will need throughout the chapter in Section 2.2. We also cover some basic spectral theory which is necessary to understand Motohashi's result on the correlations of the divisor function and consequently the formula for shifted moments. In Section 2.3 we formulate general conjectures for shifted moments, and discuss the case of four zetas in detail. We prove the main result in Sections 2.4 and 2.5. In Section 2.4 we lay the groundwork for the computation by transforming the initial problem into a tangible form. This is done in a standard manner for these types of problems, namely, we develop an approximate functional equation for

$$|\zeta(1/2+it+i\alpha)|^2|\zeta(1/2+it+i\beta)|^2$$

and use it to represent this product as a weighted double Dirichlet series. Further, in Section 2.5 we use Motohashi's result on correlations of the divisor functions to

evaluate the off-diagonal term in two distinct regimes depending on the size of $\delta = \beta - \alpha$. When δ is small, we use this result as a black box, and when δ is large, we use it in full strength and work with the explicit expansion of the error term. Finally, in Section 2.6 we discuss how our results build into the general narrative, and present possible directions for further investigation.

2.2 Preliminaries

Let us start with several standard results for bounding oscillatory integrals.

2.2.1 Oscillatory integrals

The following two lemmas can be found in Titchmarsh [128, Lemma 4.3, Lemma 4.5].

Lemma 2.2.1 (Oscillation lemma)**.** *Let $f(x)$ be a real differentiable function, and $g(x)$ be a real continuous function. If $f'(x)/g(x)$ is monotone on the interval $[a, b]$ with $|f'(x)/g(x)| \geq Y > 0$, then*

$$\int_a^b g(x)e^{if(x)}dx \ll \frac{1}{Y}.$$

Implied constant does not depend on f and g.

Lemma 2.2.2 (Van der Corput lemma)**.** *Let $f(x)$ be a real twice differentiable function, and $g(x)$ be a real continuous function such that $f'(x)/g(x)$ is monotone on the interval $[a, b]$. If $|f''(x)| \geq Y > 0$ and $|g(x)| \geq N > 0$ on $[a, b]$, then*

$$\int_a^b g(x)e^{if(x)}dx \ll \frac{N}{Y^{1/2}}$$

with constant independent of f and g.

We also have the following expression when $f'(x)$ vanishes on the interval [a,b].

Lemma 2.2.3 (Stationary phase [80])**.** *Let $f(x)$ be a differentiable function, and $g(x)$ be a continuous function on the interval $[a, b]$. Let $s \neq a, b$ be the stationary point of f i.e. $f'(s) = 0$, such that $f''(s) \neq 0$. Suppose that on $[a, b]$ we have $f^{(l)}(x) \ll AM^{-l}$ for $l \geq 2$ and $g^{(k)}(x) \leq BN^{-k}$ for $k \geq 0$. Additionally, assume $f''(x) \geq A/M^2$. Then*

$$\int_a^b g(x)e^{if(x)}dx = e^{if(s)+\frac{\pi i}{4}\operatorname{sign}(f''(s))}\frac{(2\pi)^{1/2}g(s)}{|f''(s)|^{1/2}} - \frac{g(a)e^{if(a)}}{if'(a)} + \frac{g(b)e^{if(b)}}{if'(b)} + O\left(\frac{BM^4}{A^2}\left(\frac{1}{(s-a)^3} + \frac{1}{(s-b)^3}\right)\right) + O\left(\frac{BM}{A^{3/2}}\left(1 + \frac{M}{N}\right)^2\right).$$

When $g(x)$ is a sufficiently nice function, we also have the following asymptotic expansion, see for example [127, Lemma 2.7-2.8].

Lemma 2.2.4 (Fresnel phase)**.** *Let $g(x)$ be an infinitely differentiable function of compact support, and $Y \in \mathbb{R}$. Then for any $N \in \mathbb{N}$*

$$\int g(x)e^{iYx^2}dx = \sum_{j=1}^{N} b_j g^{(2j)}(0)Y^{-j-1/2} + O_{g,N}(Y^{-N-3/2})$$

for some complex coefficients b_j. Similarly, for an infinitely differentiable function $f(x)$ with a unique stationary point s

$$\int g(x)e^{iYf(x)}dx = \sum_{j=1}^{N} b_{j,f} g^{(2j)}(s)Y^{-j-1/2} + O_{f,g,N}(Y^{-N-3/2})\,.$$

We will also need the asymptotic expansion for $\Gamma(s)$.

Lemma 2.2.5 (Stirling's formula)**.** *Let $\Gamma(s)$ be the Gamma-function. Let $T > 0$ be large, and let $s = \sigma + it$ such that $t \in [T, 2T]$, then for any $N \in \mathbb{N}$*

$$\log\Gamma(s) = \left(s - \frac{1}{2}\right)\log s - s + \frac{1}{2}\log 2\pi + \sum_{j=0}^{N} a_j s^{-(2j+1)} + O_N(T^{-(2N+1)})$$

for some real coefficients a_j.

2.2.2 Fourier and Mellin transforms

Let f be a L^1-integrable function on $\mathbb{R}$. Then one can define its Fourier transform denoted $\mathcal{F}f$ or $\hat{f}$ with the following normalisation:

$$\mathcal{F}f(\xi) = \hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ix\xi}dx\,.$$

If f is k times differentiable, and $f^{(k)}$ is L^1-integrable, then

$$\mathcal{F}f^{(k)}(\xi) = (2\pi i\xi)^k \mathcal{F}f(\xi)\,.$$

Conversely, if $g(x) = x^k f(x)$ is L^1-integrable, and $\hat{f}$ is k times differentiable, then

$$\mathcal{F}g(\xi) = (-2\pi i)^{-k}\frac{d^k}{d\xi^k}\hat{f}(\xi)$$

We can also state this result for a line integral.

Lemma 2.2.6. *Let $f(z)$ be an analytic function rapidly decaying on vertical strips, that is, $f^{(l)}(z) \ll_{l,\operatorname{Re} z} |z|^{-k}$ for any $k, l \in \mathbb{Z}^{\geq 0}$. Let $f_0(x) = f(ix)$ for $x \in \mathbb{R}$. Then for any integer $\nu \geq 0$ and $\xi \in \mathbb{R}$*

$$\int_{(1)} e^{z\xi} z^\nu f(z) dz = i \frac{d^\nu}{d\xi^\nu} \hat{f}_0(-\xi/2\pi) \,.$$

Proof. Move the line of integration to (0). Since $f(z)$ is entire and rapidly decaying,

$$\int_{(1)} e^{z\xi} z^\nu f(z) dz = \int_{(0)} e^{z\xi} z^\nu f(z) dz = i \int_{-\infty}^{\infty} e^{iy\xi} y^\nu f(iy) dy \,.$$

The statement of the lemma then follows. □

Let us also state the following fundamental results.

Theorem 2.2.7 (Fourier convolution)**.** *Let f, g and fg be L^1-integrable, then*

$$\mathcal{F}(fg)(\xi) = (\mathcal{F}f * \mathcal{F}g)(\xi) = \int_{-\infty}^{\infty} \mathcal{F}f(x) \mathcal{F}g(\xi - x) dx \,.$$

Theorem 2.2.8 (Plancherel theorem)**.** *Let f be L^1 and L^2-integrable, then*

$$\int_{-\infty}^{\infty} \hat{f}^2(x) dx = \int_{-\infty}^{\infty} f^2(x) dx \,.$$

Theorem 2.2.9 (Poisson summation)**.** *Let f be a Schwartz function, then*

$$\sum_{n \in \mathbb{Z}} f(n) = \sum_{n \in \mathbb{Z}} \hat{f}(n) \,.$$

Let $\mathcal{M}f$ denote the Mellin transform of f

$$\mathcal{M}f(z) = \int_0^{\infty} f(x) x^{z-1} dx \,.$$

Many properties of the Fourier transform including the convolution theorem translate into the corresponding properties of the Mellin transform.

Finally, let us introduce some basics of spectral theory relating to modular forms and Maass forms. We will only need notation and some existing results in the area, so this does not include any general theory of automorphic forms or Hecke operators.

2.2.3 Basic spectral theory

Let $\{\lambda_j = 1/4 + \kappa_j^2\}_{j\geq 1,\kappa_j>0}$ be the discrete spectrum of the non-Euclidean Laplacian acting on $\mathrm{SL}_2(\mathbb{Z})$-automorphic forms on the upper half plane. The space of all cusp forms admits an orthogonal basis $\{\phi_j\}_{j\geq 1}$ such that ϕ_j is an eigenfunction attached to the eigenvalue λ_j, and also simultaneously an eigenfunction of all Hecke operators $\{T(n)\}_{n\geq 1,n=-1}$. This basis is unique up to normalisation, and functions ϕ_j are then referred to as Maass cusp forms attached to respective eigenvalues λ_j.

For $n = -1$ we have $T(-1)\phi_j(z) = \varepsilon_j\phi_j(z)$ with $\varepsilon_j \in \{\pm 1\}$ defining the parity of ϕ_j, and for $n \geq 1$

$$(T(n)\phi_j)(z) = \frac{1}{\sqrt{n}} \sum_{ad=n} \sum_{b=1}^{d} \phi_j\left(\frac{az+b}{d}\right) = t_j(n)\phi_j(z)$$

produces Hecke eigenvalues $t_j(n)$. Hecke eigenvalues also come up as coefficients in the expansion

$$\phi_j(x+iy) = \rho_j(1)\sqrt{y}\sum_{n\neq 0} t_j(n)K_{i\kappa_j}(2\pi|n|y)e^{2\pi inx}\,,$$

where K_l is the lth Bessel function, and $\rho_j(1)$ is the normalisation constant.

Each $t_j(n)$ is a real multiplicative function. Trivially we have

$$|t_j(n)| \leq \sigma_1(n) = \sum_{k|n} k\,,$$

but coefficients $t_j(n)$ are conjectured to satisfy $t_j(n) \ll n^{\varepsilon}$. This statement is known as the Ramanujan-Petersson conjecture, which has been settled in many other cases, but not for Maass cusp forms. The classical bound due to Serre is $t_j(n) \ll d(n)n^{1/5}$ (see [103, 107]), and the current best bound is due to Kim and Sarnak [93] with

$$t_j(n) \ll n^{7/64}\,. \tag{2.8}$$

The Dirichlet series

$$H_j(s) = \sum_{n=1}^{\infty} \frac{t_j(n)}{n^s}\,,$$

is the Hecke L-function of ϕ_j. Using its functional equation one can then see, similarly to $\zeta(s)$, that $H_j(s)$ grows at mostly polynomially in $|\kappa_j| + |s|$. Size of $H_j(1/2+it)$ depends on both t and κ_j, and extending the Lindelöf hypothesis to this case we expect $H_j(1/2+it) \ll (t\kappa_j)^{\varepsilon}$. In the regime when t and κ_j have different sizes the best bound known is due to Jutila and Motohashi [88] with

$$H_j(1/2+it) \ll (|t|+\kappa_j)^{1/3+\varepsilon}\,.$$

Similarly, consider the space of holomorphic cusp forms of weight $2k$, where $k \geq 3$. Let $\vartheta(k)$ be its dimension and $\{\phi_{j,k}\}_{1\leq j\leq \vartheta(k)}$ be its orthogonal basis. In particular, those are eigenfunctions of the Hecke operators $T_j(n)$ with eigenvalues $t_{j,k}(n)$ as

$$(T_k(n)\phi_{j,k})(z) = \frac{1}{\sqrt{n}}\sum_{ad=n}\left(\frac{a}{d}\right)^k \sum_{b=1}^{d}\phi_{j,k}\left(\frac{az+b}{d}\right) = t_{j,k}(n)\phi_{j,k}(z)\,.$$

Similarly to $t_j(n)$, coefficients $t_{j,k}(n)$ are real and multiplicative, and come up in the Fourier expansion

$$\phi_{j,k}(z) = \rho_{j,k}(1)\sum_{n=1}^{\infty} t_{j,k}(n)e^{2\pi inz}\,,$$

where $\rho_{j,k}(1)$ is the normalisation constant. In contrast to $t_j(n)$, for which the Ramanujan-Petersson conjecture is open, we do have $|t_{j,k}(n)| \leq d(n) \ll n^{\varepsilon}$, see work of Deligne [45, 46]. The Dirichlet series

$$H_{j,k}(s) = \sum_{n=1}^{\infty}\frac{t_{j,k}(n)}{n^s}\,,$$

is the Hecke L-function of $\phi_{j,k}$. Again, we have $H_{j,k}(s) \ll (k+|s|)^b$, and expect $H_{j,k}(1/2+it) \ll (tk)^{\varepsilon}$. The best result when t and k are of different size is also due to Jutila and Motohashi [88] with

$$H_{j,k}(1/2+it) \ll (|t|+k)^{1/3+\varepsilon}\,.$$

Set

$$\alpha_j = \frac{|\rho_j(1)|^2}{\cosh \pi\kappa_j}\,,$$

and

$$\alpha_{j,k} = \frac{16\Gamma(2k)|\rho_{j,k}(1)|^2}{(4\pi)^{2k+1}}\,.$$

Then $\sum_{\kappa_j\leq K}\alpha_j \ll K^2$ and $\sum_{j=1}^{\vartheta(k)}\alpha_{j,k} \ll k$. At the central point for $m = 2, 4$ we have

$$\begin{aligned}\sum_{\kappa_j\leq X}\alpha_j H_j^m(1/2) &\ll X^{2+\varepsilon}\,;\\ \sum_{k\leq X}\sum_{j=1}^{\vartheta(k)}\alpha_{j,k}H_{j,k}^m(1/2) &\ll X^{2+\varepsilon}\,,\end{aligned} \tag{2.9}$$

see [104] and [47, Theorem 2]. In both cases $H_j(1/2)$ and $H_{j,k}(1/2)$ are in fact real. When $t \neq 0$ we also have

$$\begin{aligned}\sum_{\kappa_j\leq K}\alpha_j|H_j(1/2+it)|^2 &\ll K^{2+\varepsilon} + K^{\varepsilon}t^{2/3+\varepsilon}\,;\\ \sum_{k\leq K}\sum_{j=1}^{\vartheta(k)}\alpha_{j,k}|H_{j,k}(1/2+it)|^2 &\ll K^{2+\varepsilon} + K^{\varepsilon}t^{2/3+\varepsilon}\end{aligned} \tag{2.10}$$

due to Jutila [87]. In fact, Jutila's proof relies on corresponding bounds for linear approximations of H_j and $H_{j,k}$ respectively, and this version will be more relevant in our case.

Lemma 2.2.10 (Jutila [87]). *Let $t \in \mathbb{R}$ and $N \in \mathbb{N}$ such that $N \ll t$. Let*

$$S_j(1/2+it) = \sum_{N \le n \le 2N} \frac{t_j(n)}{n^{1/2+it}}\,;$$
$$S_{j,k}(1/2+it) = \sum_{N \le n \le 2N} \frac{t_{j,k}(n)}{n^{1/2+it}}\,.$$

Then

$$\sum_{K \le \kappa_j \le 2K} \alpha_j |S_j(1/2+it)|^2 \ll K^{2+\varepsilon} + K^{\varepsilon} t^{2/3+\varepsilon}\,,$$
$$\sum_{K \le k \le 2K} \sum_{j=1}^{\vartheta(k)} \alpha_{j,k} |S_{j,k}(1/2+it)|^2 \ll K^{2+\varepsilon} + K^{\varepsilon} t^{2/3+\varepsilon}\,.$$

For a detailed account of spectral theory of automorphic forms see Iwaniec [84].

2.3 Conjectures and predictions

First, we would like to discuss the conjectures for shifted moments, and specifically the case of $k = 2$ in more detail.

2.3.1 General conjecture

Let $\overline{\alpha} = (\alpha_1, \dots, \alpha_k)$, $\overline{\beta} = (\beta_1, \dots, \beta_k) \in \mathbb{R}^k$, and $\overline{\gamma} = (\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k) \in \mathbb{R}^{2k}$ be their concatenation. Sometimes shifts are chosen to be in $\mathbb{C}$ with small real part, but here we restrict to imaginary shifts along the critical line. We will omit the bars and simply write α, β, γ throughout this particular section.

Let $\sigma > 1$, and consider the integral

$$\int_0^T \zeta(\sigma+it+i\alpha_1)\cdots\zeta(\sigma+it+i\alpha_k)\zeta(\sigma-it-i\beta_1)\cdots\zeta(\sigma-it-i\beta_k)dt\,.$$

Let $l_1, \dots, l_k$ be arbitrary complex numbers, and let $\tau_{(l_1,\dots,l_k)}(n)$ be the generalised divisor function

$$\tau_{(l_1,\dots,l_k)}(n) = \sum_{n_1\cdots n_k = n} n_1^{l_1}\cdots n_k^{l_k}\,.$$

In particular, if $l_j = 0$ for all j,

$$\tau_{(0,\dots,0)}(n) = \sum_{n_1\cdots n_k = n} 1$$

is the k-fold divisor function, sometimes denoted $d_k(n)$. It is also a non-trivial but well-known fact that $d_k(n) \ll_{k,\varepsilon} n^{\varepsilon}$ for any $\varepsilon > 0$.

Expanding every zeta into its Dirichlet series gives

$$\prod_{j=1}^{k} \zeta(\sigma + it + i\alpha_j)\zeta(\sigma - it - i\beta_j) = \sum_{n,m=1}^{\infty} \frac{\tau_{-i\alpha}(n)\tau_{i\beta}(m)}{(nm)^{\sigma}} \left(\frac{m}{n}\right)^{it} . \tag{2.11}$$

Integrating over t,

$$\int_0^T \prod_{j=1}^{k} \zeta(\sigma + it + i\alpha_j)\zeta(\sigma - it - i\beta_j)dt = T\sum_{n=1}^{\infty} \frac{\tau_{-i\alpha}(n)\tau_{i\beta}(n)}{n^{2\sigma}} + \sum_{n\neq m} \frac{\tau_{-i\alpha}(n)\tau_{i\beta}(m)}{(nm)^{\sigma} i \log m/n} \left(\left(\frac{m}{n}\right)^{iT} - 1\right) .$$

The first term is a convergent series. For $\sigma > 1$ and sufficiently small ε

$$\sum_{n=1}^{\infty} \frac{\tau_{-i\alpha}(n)\tau_{i\beta}(n)}{n^{2\sigma}} \ll \sum_{n=1}^{\infty} \frac{d_k(n)^2}{n^{2\sigma}} \ll_{k,\varepsilon} \sum_{n=1}^{\infty} \frac{n^{\varepsilon}}{n^{2\sigma}} ,$$

thus the diagonal contribution, that is when $m = n$, is

$$T\sum_{n=1}^{\infty} \frac{\tau_{-i\alpha}(n)\tau_{i\beta}(n)}{n^{2\sigma}} = C_{\alpha,\beta,\sigma} T .$$

The off-diagonal contribution, when $n \neq m$, turns out to be negligible due to oscillation. Assume $m > n$ and split the sum into two ranges, $n < m < 2n$ and $m \geq 2n$. Then for sufficiently small $\varepsilon > 0$

$$\sum_{\substack{n\geq 1 \\ m\geq 2n}} \frac{\tau_{-i\alpha}(n)\tau_{i\beta}(m)}{(nm)^{\sigma} \log m/n} \left(\left(\frac{m}{n}\right)^{iT} - 1\right) \ll_{k,\varepsilon} \sum_{\substack{n\geq 1 \\ m>2n}} \frac{(nm)^{\varepsilon}}{(nm)^{\sigma} \log 2} \ll_{k,\sigma} 1 .$$

For the complementary range notice that

$$\frac{1}{\log m/n} \left(\left(\frac{m}{n}\right)^{iT} - 1\right) \ll \frac{\left|\sin\left(\frac{1}{2}T\log m/n\right)\right|}{\log m/n} \ll \min\{T, \log^{-1} m/n\} .$$

Then setting $r = m - n$ and using $\log m/n = \log(1 + r/n) \asymp r/n$ we have

$$\sum_{\substack{n\geq 1 \\ n<m<2n}} \frac{\tau_{-i\alpha}(n)\tau_{i\beta}(m)}{(nm)^{\sigma} \log m/n} \left(\left(\frac{m}{n}\right)^{iT} - 1\right) \ll_{k,\varepsilon} \sum_{\substack{n\geq 1 \\ 1\leq r<n}} \frac{n^{\varepsilon}}{n^{2\sigma}} \min\{T, n/r\} \ll_{k,\sigma} 1 .$$

When $\sigma \leq 1$ we still expect the main contribution to come from the diagonal term, though not in a straightforward manner. This should be interpreted in the sense that

we can reconstruct the value of the integral from the diagonal contribution in (2.11), but not that the off-diagonal contributions are negligible.

Let $\mathbf{p}^{-i\alpha} = (p^{-i\alpha_1}, \dots, p^{-i\alpha_k})$ and $\mathbf{p}^{i\beta} = (p^{i\beta_1}, \dots, p^{i\beta_k})$, then setting $n = m$ in (2.11),

$$\zeta_\gamma(s) := \sum_{n=1}^{\infty} \frac{\tau_{-i\alpha}(n)\tau_{i\beta}(m)}{n^{2s}} = \prod_p \sum_{m=0}^{\infty} \frac{h_m(\mathbf{p}^{-i\alpha})h_m(\mathbf{p}^{i\beta})}{p^{2ms}}\,,$$

where $h_m(\mathbf{x})$ is the m-th complete symmetric polynomial at $\mathbf{x} = (x_1, \dots, x_k)$. Looking at this Euler product, we see that

$$\zeta_\gamma(s) = F(s;\alpha,\beta) \prod_{j,l=1}^{k} \zeta(2s + i(\alpha_j - \beta_l))\,,$$

where $F(s;\alpha,\beta)$ has an absolutely convergent product for $\operatorname{Re} s > 1/4$. The product of zetas has an analytic continuation to the whole complex plane with single poles at

$$s = \frac{1}{2} - \frac{i}{2}(\alpha_j - \beta_l)\,.$$

Thus if we wanted to make a prediction for the average with $s = 1/2$ we immediately run into singularities of $\alpha_j = \beta_l$ for some j and l, and even computing residues here is not straightforward. In order to fix that, let us symmetrise the conjecture with respect to permutations of the coordinates of γ that do not fix α and β when $s = 1/2$.

Recall the functional equation of the Riemann zeta

$$\pi^{-s/2}\Gamma(s/2)\zeta(s/2) = \pi^{-(1-s)/2}\Gamma((1-s)/2)\zeta(1-s)\,, \tag{2.12}$$

or $\zeta(s) = \chi(s)\zeta(1-s)$, where

$$\chi(s) = \pi^{s-1/2}\frac{\Gamma((1-s)/2)}{\Gamma(s/2)} = 2^s\pi^{s-1}\sin(\pi s/2)\Gamma(1-s)\,. \tag{2.13}$$

Let $\pi \in S_{2k}$ be a permutation on the coordinates of γ with $\pi(\gamma) = (\tilde{\alpha}, \tilde{\beta})$, and define

$$G_\gamma(\pi) := \prod_{\substack{1\le j\le k \\ \tilde{\alpha}_j \notin \alpha, \tilde{\beta}_j \notin \beta}} \chi(1/2 - it - i\tilde{\alpha}_j)\chi(1/2 + it + i\tilde{\beta}_j)\,,$$

where the product is taken only over shifts in γ that moved to the other half of the vector. This factor allows us to symmetrise the initial product of zetas so that

$$\prod_{j=1} \zeta(1/2 + it + i\alpha_j)\zeta(1/2 - it - i\beta_j) = G_\gamma(\pi)\prod_{j=1}\zeta(1/2 + it + i\tilde{\alpha}_j)\zeta(1/2 - it - i\tilde{\beta}_j)\,.$$

Additionally, sum over all permutations in $S_{2k}/(S_k \times S_k)$. Then the conjecture takes form

$$Q(t;\alpha,\beta) := \sum_{\pi \in S_{2k}/(S_k \times S_k)} G_\gamma(\pi)\zeta_{\pi(\gamma)}(1/2)\,.$$

One can check that Q has removable singularities at $\alpha_j = \beta_l$, so we can think of Q as entire. The sum above contains $\binom{2k}{k}$ terms and is generally difficult to analyse for larger k. In the simplest case $k = 1$ we have

$$Q(t,\alpha,\beta) = \zeta(1+i\alpha-i\beta) + \chi(1/2+it+i\alpha)\chi(1/2-it-i\beta)\zeta(1+i\beta-i\alpha)\,.$$

Conjecture 2.3.1 (Conrey, Farmer, Keating, Rubinstein, and Snaith [39])**.** *Let $f(t)$ be an appropriate weight function, and let $Q(t;\alpha,\beta)$ be as above, then*

$$\mathcal{M}_{2k}(f;\alpha,\beta) = \int_0^\infty f(t)\prod_{j=1}^{k}\zeta(1/2+it+i\alpha_j)\zeta(1/2-it-i\beta_j)dt = \int_0^\infty f(t)Q(t;\alpha,\beta)dt + E_f\,,$$

where E_f is negligible, and

$$\mathcal{M}_{2k}(T;\alpha,\beta) = \frac{1}{T}\int_0^T \prod_{j=1}^{k}\zeta(1/2+it+i\alpha_j)\zeta(1/2-it-i\beta_j)dt = \frac{1}{T}\int_0^T Q(t;\alpha,\beta)dt + O(T^{-a+\varepsilon})$$

for some $a > 0$ (sometimes $a = 1/2$).

The only cases where this is known to hold are $k = 1, 2$. In particular, Ingham [81] showed that for $\alpha, \beta \ll 1$

$$\mathcal{M}_2(T,\alpha,\beta) = \frac{1}{T}\int_0^T Q(t;\alpha,\beta)dt + O(T^{-1/2+\varepsilon})\,,$$

and Bettin [9] extended this to $\alpha, \beta \ll T^{2-\varepsilon}$ with error $O(T^{-1}\max\{T,|\alpha|,|\beta|\}^{1/2+\varepsilon})$.

The case of $k = 2$ was first studied by Motohashi [105] and is known as Motohashi's integral. This result predated any of the generalised conjectures and in fact inspired them. Motohashi showed that the above mentioned conjecture holds for $\alpha_j, \beta_j \ll 1/\log T$, and gave an explicit expansion for the error term using spectral theory of automorphic forms. This is still the best range known for distinct shifts when $k = 2$.

Let us now discuss the case $k = 2$ in more detail.

2.3.2 Motohashi's integral

Let $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$, and let f be an appropriate weight function. Following Motohashi's work [105], consider the integral

$$\int_{-\infty}^{+\infty} f(t)\zeta(1/2+it+i\alpha_1)\zeta(1/2+it+i\alpha_2)\zeta(1/2-it-i\beta_1)\zeta(1/2-it-i\beta_2)dt\,.$$

In this case for α_j and β_j distinct

$$\zeta_\gamma(1/2) = \frac{\zeta(1+i\alpha_1-i\beta_1)\zeta(1+i\alpha_1-i\beta_2)\zeta(1+i\alpha_2-i\beta_1)\zeta(1+i\alpha_2-i\beta_2)}{\zeta(2+i\alpha_1-i\beta_1+i\alpha_2-i\beta_2)}\,.$$

Set $Z(\alpha_1,\alpha_2,\beta_1,\beta_2) := \zeta_\gamma(1/2)$, and define

$$\kappa_{x,y}(t) := \chi(1/2+it+ix)\chi(1/2-it-iy)\,.$$

Adapting the general case to $k=4$, we expect that

$$\mathcal{M}_4(f;\alpha,\beta) = \int_{-\infty}^{+\infty} Q(t;\alpha,\beta)f(t)dt + E_f(\alpha,\beta)\,, \tag{2.14}$$

where $E_f(\alpha,\beta)$ is negligible, and

$$\begin{aligned} Q(t;\alpha,\beta) = {} & Z(\alpha_1,\alpha_2,\beta_1,\beta_2) + \kappa_{\alpha_1,\beta_1}(t)\kappa_{\alpha_2,\beta_2}(t)Z(\beta_1,\beta_2,\alpha_1,\alpha_2)+ \\ & \kappa_{\alpha_1,\beta_1}(t)Z(\beta_1,\alpha_2,\alpha_1,\beta_2) + \kappa_{\alpha_2,\beta_2}(t)Z(\alpha_1,\beta_2,\beta_1,\alpha_2)+ \\ & \kappa_{\alpha_2,\beta_1}(t)Z(\alpha_1,\beta_1,\alpha_2,\beta_2) + \kappa_{\alpha_1,\beta_2}(t)Z(\beta_2,\alpha_2,\beta_1,\alpha_1) \end{aligned} \tag{2.15}$$

is an entire function.

For any $v \in \mathbb{R}$ we have $\chi(1/2+iv) = e^{if(v)}$, where

$$f(v) = -v\log\frac{|v|}{2\pi e} + \frac{\pi}{4}\operatorname{sign} v + O(1/v)\,,$$

and for $v > 1$ the error term is in fact a convergent series coming from the asymptotic expansion of $\Gamma(s)$. For $x, y \in \mathbb{R}$ set

$$\kappa_{x,y}(t) = e^{i\phi_{x,y}(t)}\,.$$

When $|x-y| < |t+x|$, this function reduces to a more convenient form

$$\phi_{x,y}(t) = (y-x)\log\frac{|t+x|}{2\pi e} + O(|y-x|^2/|t+x|)\,. \tag{2.16}$$

Additionally, for $m \in \mathbb{N}$

$$\begin{aligned} (-1)^m\frac{d^m}{ds^m}\kappa_{0,is}(y)\Big|_{s=0} = {} & \chi(1/2+iy)\chi^{(m)}(1/2-iy) = \\ & \log^m\frac{|y|}{2\pi} + O(1/y)\,. \end{aligned} \tag{2.17}$$

Let $\tau = t/2\pi$, and $\mu = \max\{|\alpha_j|, |\beta_j|\}$. If $\mu \ll t^{1/2-\varepsilon}$, then using (2.16) we can rewrite (2.15) in a more convenient way as

$$\begin{aligned} Q(t;\alpha,\beta) = \big(&Z(\alpha_1,\alpha_2,\beta_1,\beta_2) + \tau^{i(\beta_1+\beta_2-\alpha_1-\alpha_2)} Z(\beta_1,\beta_2,\alpha_1,\alpha_2) + \\ &\tau^{i(\beta_1-\alpha_1)} Z(\beta_1,\alpha_2,\alpha_1,\beta_2) + \tau^{i(\beta_2-\alpha_2)} Z(\alpha_1,\beta_2,\beta_1,\alpha_2) + \\ &\tau^{i\beta_2-i\alpha_1} Z(\beta_2,\alpha_2,\beta_1,\alpha_1) + \tau^{i\beta_1-i\alpha_2} Z(\alpha_1,\beta_1,\alpha_2,\beta_2)\big)\left(1+O(\mu^2/t)\right). \end{aligned}$$

Motohashi proved that the prediction (2.14) holds for shifts up to size $1/\log T$, but it remains an open question for larger shifts.

Let us now consider a particular case when $\alpha_j = \beta_j$ for $j = 1, 2$. Set

$$D(z,t) := L(z)\kappa_{0,iz}(t) + L(-z)\kappa_{-iz,0}(t)\,,$$

where

$$L(z) := \frac{\zeta^4(1+z)}{\zeta(2+2z)}\,.$$

Further, let $\gamma_0 = [z^0]\zeta(1+z)$, and define

$$\begin{aligned} H(z,t) := \frac{1}{2}\frac{d^2}{ds^2}\left(h_0(z,s)e^{2\gamma_0 s}(1+\kappa_{0,-is-iz}(t)\kappa_{0,-is-iz}(t))\right)\Big|_{s=0} - \\ \frac{1}{4}h_0(z,0)\frac{d^2}{ds^2}\left(\kappa_{0,-is-iz}(t)\kappa_{0,is-iz}(t)\right)\Big|_{s=0}, \end{aligned} \tag{2.18}$$

where

$$h_0(z,s) := \frac{\zeta(1+z+s)\zeta(1-z+s)}{\zeta(2+2s)}\,. \tag{2.19}$$

Take the limit of (2.15) as $\alpha_1-\beta_1$ and $\alpha_2-\beta_2$ tend to zero via a residue computation or otherwise. Let $\alpha := \alpha_1$, $\beta := \alpha_2$ and $\delta = \beta - \alpha$, then the first four terms in (2.15) give $H(i\delta, t+\alpha)$, and the last two terms give $D(i\delta, t+\alpha)$.

Further, assuming $|t+\alpha| \le |t+\beta|$ are large and using (2.17) we have

$$H(i\delta,t+\alpha) = \left(\frac{d^2}{ds^2}\,h_0(i\delta,s)e^{s(\xi+2\gamma_0)}\Big|_{s=0} - \frac{1}{4}h_0(i\delta,0)\log^2\left|\frac{t+\beta}{t+\alpha}\right|\right)(1+O(1/|t+\alpha|))\,, \tag{2.20}$$

where

$$\xi = \frac{1}{2}\log\frac{|(t+\alpha)(t+\beta)|}{4\pi^2}\,.$$

When $|\alpha| < |\beta| = o(t)$ this further reduces to

$$\frac{d^2}{ds^2}\left(h_0(i\delta,s)e^{s(\log t/2\pi+2\gamma_0)}\right)\Big|_{s=0}(1+O(|\beta|/t)) = \sum_{j=0}^{2} H_j(i\delta)\left(\log\frac{t}{2\pi}+2\gamma_0\right)^j.$$

The sum of the two terms above $F(z,t) := D(z,t) + H(z,t)$ has a removable singularity at $z = 0$ with

$$\lim_{z\to 0} F(z,t) = \operatorname*{Res}_{z=0} \frac{F(z,t)}{z} = P_4\left(\log \frac{t}{2\pi}\right), \tag{2.21}$$

where $P_4(x)$ is a polynomial of degree 4 giving the fourth moment of the Riemann zeta. Comparing to Heath-Brown [73] and Conrey [38] we also observe that the diagonal contribution (of the truncated Dirichlet series) in the fourth moment computation, that is for $\mathcal{M}_4(T;0,0)$, is given by

$$2T \operatorname*{Res}_{z=0} \frac{\zeta^4(1+z)}{z(z+1)\zeta(2+2z)} \left(\frac{T}{2\pi}\right)^z = \int_0^T \operatorname*{Res}_{z=0} \left(z^{-1} D(z,t)\right) dt + o(T).$$

The off-diagonal contribution (of the truncated Dirichlet series) is given by

$$T\frac{d^2}{ds^2} \left(\frac{\zeta^2(1+s) - \frac{2}{s}\zeta(1+2s)}{(s+1)\zeta(2+2s)} \left(\frac{Te^{2\gamma_0}}{2\pi}\right)^s \right) \Big|_{s=0} = \int_0^T \operatorname*{Res}_{z=0} \left(z^{-1} H(z,t)\right) dt + o(T).$$

Thus we will refer to $D(i\delta, t+\alpha)$ and $H(i\delta, t+\alpha)$ as the diagonal and the off-diagonal terms respectively. The correspondence of these terms to specific terms in (2.15) is irrelevant since it is affected by the choice of lines in $\mathbb{C}^4$ along which we compute the limits of separate groups of terms. Integrating $F(i\delta, t)$ over t gives an expression identical to one in Theorem 2.1.1.

Now let us analyse how $F(z,t)$ behaves on average as $z = i\delta$ ranges over an interval of some length by computing the corresponding moment of moment.

2.3.3 (2,2)-moment of moment

Let $c = c(T)$ be either fixed or growing with T. Let $g_0(x)$ be a normalised even rapidly decaying weight function concentrating on the interval $[-1,1]$ of total mass 1. Set $g(x) = c^{-1} g_0(x/c)$, and consider

$$\mathcal{M}_{2,2}(T;g) := \frac{1}{T}\int_0^T \left(\int_{-\infty}^{+\infty} g(h)|\zeta(1/2+it+ih)|^2 dh\right)^2 dt.$$

Let

$$\overline{Q}(t,g) := \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(\alpha)g(\beta)F(i(\alpha-\beta),t)d\alpha d\beta,$$

then we expect that up to a small error $\mathcal{M}_{2,2}(T;g) \approx \frac{1}{T}\int_0^T \overline{Q}(t,g)dt$.

Recall that

$$\frac{1}{X}\int_0^X |\zeta(1/2+it)|^2 dt = M(X) + E(X),$$

where $M(X) = \log X + 2\gamma_0 - 1$, and $E(X) \ll X^{-1/2+\varepsilon}$; see [128, Theorem 7.4]. This is not the sharpest error term, but it suffices for our purposes. Using this expression for the internal integral in $\mathcal{M}_{2,2}(T;g)$ when $g(x) = \frac{1}{2c}\mathbf{I}\{|x| \le c\}$ yields

$$\mathcal{M}_{2,2}(T;g) - \frac{1}{4T}\int_0^T (M(X+c) - M(X-c))^2 dt \ll \frac{(T+c)^{1/2+\varepsilon}}{c} + \frac{(T+c)^{1+\varepsilon}}{c^2}\,.$$

Thus when $c(T) \gg T^{1/2+\varepsilon}$ computing $\mathcal{M}_{2,2}(T;g)$ with a power saving does not in fact require any new results. With this in mind, let us move on to the computation.

Let $\tilde{f}$ denote the Taylor series (holomorphic) part of the expansion of a meromorphic function f around $z=0$, that is if $f(z) = \sum_{n\ge -N} a_j z^j$, then $\tilde{f}(z) = \sum_{n\ge 0} a_j z^j$.

To compute $\overline{Q}(t,g)$ we again would like to split it into the diagonal and the off-diagonal contributions. Function $F(z,t)$ is entire in z, but $D(z,t)$ and $H(z,t)$ are not. However, $F(z,t) = \tilde{F}(z,t)$, thus integrating $F = D + H$ is the same as integrating $\tilde{F} = \tilde{D} + \tilde{H}$, where we truncated the series of F, D, H as functions in z. Let

$$\overline{Q}(t;g) = \iint g(\alpha)g(\beta)\tilde{D}(i(\alpha-\beta),t)d\alpha d\beta + \iint g(\alpha)g(\beta)\tilde{H}(i(\alpha-\beta),t)d\alpha d\beta\,,$$

where both integrals real-valued. Denote these two integrals by $I_1(t;g)$ and $I_2(t;g)$ respectively.

To compute $\int_0^T I_1(t;g)dt$ we will first need the following technical lemma.

Lemma 2.3.2. *Let $T \ge 10$ and let $x, y \in \mathbb{R}$. Then*

$$\int_0^T \kappa_{x,y}(t)dt = \int_0^T \chi(1/2+it+ix)\chi(1/2-it-iy) = \frac{(t+x)\kappa_{x,y}(t)}{1+i(y-x)}\bigg|_0^T + O\left(1 + \min\{a^{1/2}, \max_{t\in[0,T]}|\phi'(t)|^{-1}\}\right).$$

Proof. If $x = y$, the statement of the lemma is trivial. Let $a = y - x \neq 0$. Recall that

$$\kappa_{x,y}(t) = \chi(1/2+it+ix)\chi(1/2-it-iy) = e^{i\phi_{x,y}(t)}\,,$$

and, in particular, when $|t+x| > |a|$,

$$\phi_{x,y}(t) = a\log\frac{|t+x|}{2\pi e} + O(a^2/|t+x|)\,.$$

Using this approximation, integrate by parts:

$$\int_0^T \kappa_{x,y}(t)dt = \int_0^T |t+x|^{ia} e^{i\phi_{x,y}(t) - ia\log|t+x|}dt = \frac{(t+x)\kappa_{x,y}(t)}{1+ia}\bigg|_0^T - i\int_0^T \frac{(t+x)\kappa_{x,y}(t)}{1+ia}\left(\phi'_{x,y}(t) - \frac{a}{t+x}\right)dt\,.$$

If $t+x \asymp a$, then by the oscillation lemma for either the first or the second derivative

$$\int_{t+x\asymp a} \frac{(t+x)\kappa_{x,y}(t)}{1+ia}\left(\phi'_{x,y}(t)-\frac{a}{t+x}\right)dt \ll \min\{a^{1/2}, \max_{t\in[0,T]}|\phi'(t)|^{-1}\}\,.$$

Otherwise if $t+x=o(a)$ or $a=o(t+x)$, then

$$\phi'_{x,y}(t)-\frac{a}{t+x}=\log\left|1+\frac{a}{t+x}\right|-\frac{a}{t+x}+O\left(\frac{1}{|t+x|^2}+\frac{1}{|t+y|^2}\right)$$

together with the oscillation lemma for the first derivative gives

$$\int_{t+x\not\asymp a} \frac{(t+x)\kappa_{x,y}(t)}{1+ia}\left(\phi'_{x,y}(t)-\frac{a}{t+x}\right)dt \ll \frac{|a|}{|1+ia|}\ll 1\,.$$

This then implies the statement of the lemma. □

Lemma 2.3.3. *Let $T\geq 10$, and let $c=c(T)$, either fixed or growing with T. Let $g(x)=\frac{1}{c}g_0(x/c)$, where $g_0(x)$ is a weight function rapidly decaying outside of $[-1,1]$. Then*

$$\frac{1}{T}\int_0^T I_1(t;g)dt = 2\int_0^{\log\frac{T}{2\pi}}\hat{g}^2(\frac{w}{2\pi})P_3(\log\frac{T}{2\pi}-w)dw+O(\max_{y\gg\log T}|\hat{g}(y)|^2+T^{-1/2+\varepsilon})$$

where $d>0$, and $P_3(\log y)=Res_{z=0}\frac{L(z)}{1+z}y^z$.

Proof. First, let us show that this term is only meaningful when $c=c(T)\ll T^{\varepsilon}$. Suppose $c=c(T)\to\infty$ as $T\to\infty$, then

$$\int_0^T I_1(t,g)dt=\left(\iint_{|x-y|\ll 1}+\iint_{|x-y|\gg 1}\right)\int_0^T \tilde{D}(i(x-y),t)dt\,g(x)g(y)dxdy=$$
$$2\,\mathrm{Re}\iint_{|x-y|\gg 1}L(i(x-y))\int_0^T\kappa_{x,y}(t)dt\,g(x)g(y)dxdy+O(T^{1+\varepsilon}/c)\,.$$

When $|x-y|\gg 1$, by Lemma 2.3.2

$$\int_0^T\kappa_{x,y}(t)dt\ll\min\{T,(T+|x|)/|x-y|+|x-y|^{1/2}\wedge\max_{t\in[0,T]}|\phi'(t)|^{-1}\}\,.$$

Using the above together with basic bounds $\zeta(1+it)\ll t^{\varepsilon}$ when $t\gg 1$, and integrating over x,y

$$\int_0^T I_1(t,g)dt\ll(T+c)^{\varepsilon}(1+T/c+\min\{c^{1/2},T/c^{1/2}\})\,.$$

In particular, this is $O(T^{1/2+\varepsilon})$ uniformly for $c\gg T^{1/2}$.

Thus we only need to consider the case when $c(T)$ is small, say $c(T) \ll T^{1/2}$. In this range using (2.16)

$$\frac{1}{T}\int_0^T D(i(x-y),t)dt = 2\,\mathrm{Re}\,\frac{L(i(x-y))}{1+i(x-y)}\left(\frac{T}{2\pi}\right)^{i(x-y)}(1+O(T^{\varepsilon-1}(|x|+|y|)))\,.$$

Recall the classic result $\sum_{n\le x} d^2(n) = xP_3(\log x) + E(x)$, where $P_3(\log y) = \mathrm{Res}_{z=0}\frac{L(z)}{1+z}y^z$ and $E(x) \ll x^{1/2+\varepsilon}$. Let $z \in \mathbb{C}$ with $\mathrm{Re}\,z > 1$, then by Perron's formula

$$L(z) = (1+z)\int_1^\infty P_3(\log x)x^{-z-1}dx + (1+z)\int_1^\infty E(x)x^{-z-2}dx\,.$$

We can directly check that the Taylor series part of $\frac{1}{T}\int_0^T L(z)\tau^z dt$ is

$$\left(\frac{T}{2\pi}\right)^z\left(\int_1^{T/2\pi} P_3(\log x)\frac{dx}{x^{z+1}} + \int_1^\infty E(x)\frac{dx}{x^{z+2}}\right)$$

and this expression is valid for all $\mathrm{Re}\,z > -1/2+\varepsilon$. Set $z = i(\alpha-\beta)$ and integrate the above with respect to α and β, then together with Lemma 2.3.2 this gives

$$\begin{aligned}\frac{1}{T}\int_0^T I_1(t;g)dt =& 2\int_1^{T/2\pi} P_3(\log x)\hat{g}^2\left(\frac{1}{2\pi}\log\frac{T}{2\pi x}\right)\frac{dx}{x} + \\ & 2\int_1^\infty \frac{E(x)}{x^2}\hat{g}^2\left(\frac{1}{2\pi}\log\frac{T}{2\pi x}\right)dx + O(T^{-1+\varepsilon}c)\,.\end{aligned}$$

Splitting into $x \ll T^{1-a_2}$ and $x \gg T^{1-a_2}$ for any $a_2 \in (0,1)$, and using $E(x) \ll x^{1/2+\varepsilon}$

$$\int_1^\infty \frac{E(x)}{x^2}\hat{g}^2\left(\frac{1}{2\pi}\log\frac{T}{2\pi x}\right)dx \ll T^{\frac{a_2-1}{2}+\varepsilon} + \max_{w\gg T^{a_2}}\left|\hat{g}(\frac{\log w}{2\pi})\right|^2.$$

Gathering everything together,

$$\frac{1}{T}\int_0^T I_1(t;g)dt = 2\int_0^{\log\frac{T}{2\pi}}\hat{g}^2(\frac{w}{2\pi})P_3(\log\frac{T}{2\pi}-w)dw + O(T^{-1/2+\varepsilon} + \max_{y\gg\log T}\left|\hat{g}(y)\right|^2)$$

for any $c(T) \gg 1$. □

Moving on to the off-diagonal part, first check directly (say by computing the Laurent series) that

$$h(z,s) = \tilde{h}_0(z,s) = \frac{1}{\zeta(2+2s)}\left(\zeta(1+z+s)\zeta(1-z+s) - \frac{2s\zeta(1+2s)}{s^2-z^2}\right), \quad (2.22)$$

so $\tilde{H}(z,t)$ is obtained by replacing h_0 by h in (2.18).

Lemma 2.3.4. *Let $T \geq 10$, and let $c = c(T)$, either fixed or growing with T. Let $g(x) = \frac{1}{c} g_0(x/c)$, where $g_0(x)$ is a weight function rapidly decaying outside of $[-1, 1]$. If $c = o(T)$, then*

$$\frac{1}{T}\int_0^T I_2(t; g)dt = a_2(g) + 2a_1(g)(\log T/2\pi + 2\gamma_0) + a_0(g)(\log T/2\pi + 2\gamma_0)^2 + O(T^{-1+\varepsilon}c)\,,$$

where

$$a_j(g) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{d^j}{ds^j}\left(\frac{h(i(x-y), s)}{1+s}\right)\Big|_{s=0} g(x)g(y)dxdy \ll 1\,.$$

If $c = c(T) \to \infty$, then

$$\frac{1}{T}\int_0^T I_2(t; g)dt = \frac{1}{T}\int_0^T \left(\int \log\frac{|t+x|}{2\pi} g(x)dx + 2\gamma_0\right)^2 dt\left(1 + O((T+c)^{-1+\varepsilon} + c^{-1+\varepsilon})\right)\,.$$

Proof. Let $\alpha, \beta \in \mathbb{R}$ such that $\delta = \beta - \alpha$. Let $t_1 = t + \alpha$, and

$$\xi = \frac{1}{2}\log\frac{|(t+\alpha)(t+\beta)|}{4\pi^2} = \frac{1}{2}\log\frac{|t_1(t_1+\delta)|}{4\pi^2}\,,$$

then

$$\tilde{H}(i\delta, t_1) = \left(\frac{d^2}{ds^2} h(i\delta, s)e^{s(\xi+2\gamma_0)}\Big|_{s=0} - \frac{1}{4}h(i\delta, 0)\log^2\frac{|t_1+\delta|}{|t_1|}\right)\left(1 + O\left(\frac{1}{|t_1| \wedge |t_1+\delta|}\right)\right)\,. \tag{2.23}$$

Computing $I_2(t; g)$ then boils down to computing integrals of the form

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^{(j)}(i(x-y), 0)\log^k\frac{|t+x|}{2\pi}\log^l\frac{|t+y|}{2\pi} g(x)g(y)dxdy\,.$$

When $c \ll 1$, this is trivial, so assume $c(T) \to \infty$ as $T \to \infty$. Recall the approximate functional equation [128, Theorem 4.11] for $\zeta(1+it)$ when $|t| \ll x$

$$\zeta(1+it) = \sum_{n \leq x}\frac{1}{n^{1+it}} - \frac{x^{-it}}{it} + O(1/x)\,,$$

and expand $h(i\delta, 0)$ into the corresponding truncated Dirichlet series. Then exactly as we have shown for (2.11) only the diagonal term matters, while the contribution of the off-diagonal term is $O(c^{\varepsilon-1})$. For example, the leading term in t is

$$h(i\delta, 0)\log\frac{|t+\alpha|}{2\pi}\log\frac{|t+\beta|}{2\pi}$$

and integrating over α and β together with the approximate functional equation gives

$$\iint h(i\delta,0)\log\frac{|t+\alpha|}{2\pi}\log\frac{|t+\beta|}{2\pi}g(\alpha)g(\beta)d\alpha d\beta =$$
$$\frac{1}{\zeta(2)}\iint\sum_{n,m\ll c}\frac{1}{nm}\left(\frac{m}{n}\right)^{i\delta}\log\frac{|t+\alpha|}{2\pi}\log\frac{|t+\beta|}{2\pi}g(\alpha)g(\beta)d\alpha d\beta+O(c^{\varepsilon-1}) =$$
$$\frac{1}{\zeta(2)}\sum_{n,m\ll c}\frac{1}{nm}\iint\left(\frac{m}{n}\right)^{i\delta}\log\frac{|t+\alpha|}{2\pi}\log\frac{|t+\beta|}{2\pi}g(\alpha)g(\beta)d\alpha d\beta+O(c^{\varepsilon-1})\,.$$

When $m=n$,

$$\sum_{n\ll c}\frac{1}{n^2}=\zeta(2)+O(1/c)\,,$$

and when $m\neq n$ using the oscillation lemma for any appropriate $g(x)$ satisfying the assumption in the statement of the lemma

$$\sum_{m\neq n}\frac{1}{mn}\iint\left(\frac{m}{n}\right)^{i\delta}\log\frac{|t+\alpha|}{2\pi}\log\frac{|t+\beta|}{2\pi}g(\alpha)g(\beta)d\alpha d\beta\ll$$
$$\frac{\log^2|t+c|}{c^2}\sum_{m\neq n}\frac{1}{mn\log^2 m/n}\ll c^{-1+\varepsilon}\log^2|t+c|\,.$$

Thus

$$\iint h(i\delta,0)\log\frac{|t+\alpha|}{2\pi}\log\frac{|t+\beta|}{2\pi}g(\alpha)g(\beta)d\alpha d\beta =$$
$$\left(\iint\log\frac{|t+\alpha|}{2\pi}\log\frac{|t+\beta|}{2\pi}g(\alpha)g(\beta)d\alpha d\beta\right)(1+O(c^{\varepsilon-1}))=$$
$$\left(\int\log\frac{|t+x|}{2\pi}g(x)dx\right)^2\left(1+O(c^{\varepsilon-1})\right)\,.$$

Computing other terms in a similar way then gives the statement of the lemma. □

Finally, we would like to give an integral representation of the off-diagonal contribution though a Parseval type identity.

Lemma 2.3.5. *Let $T\geq 10$, and let $c=c(T)$ such that $c=o(T)$. Let $g(x)=\frac{1}{c}g_0(x/c)$, where $g_0(x)$ is a weight function rapidly decaying outside of $[-1,1]$. Let $s\in\mathbb{C}$ be such that $\Re s\geq 0$, then*

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}h(i\delta,s)g(x)g(y)dxdy=\sum_{n=1}^{\infty}\frac{1}{n^{s+1}}\int_n^{\infty}((x))\left(\frac{\hat{g}^2(\frac{1}{2\pi}\log x/n)}{x^{s+1}}\right)'dx\,,$$

where $((x))=x-\lfloor x\rfloor-1/2$.

Proof. Let $\delta = x - y \in \mathbb{R}$ and $s \in \mathbb{C}$ such that $\Re s > 0$, then

$$h(i\delta, s) = \sum_{n,m=1}^{\infty} \frac{(n/m)^{i\delta}}{(nm)^{1+s}} - \sum_{n=1}^{\infty} \frac{1}{n^{1+2s}} \left(\frac{1}{s+i\delta} + \frac{1}{s-i\delta} \right).$$

Integrating this expression with respect to x, y, we have

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(i\delta, s) g(x) g(y) dx dy = \sum_{n,m=1}^{\infty} \frac{\hat{g}^2(\frac{1}{2\pi} \log n/m)}{(nm)^{1+s}} - 2 \sum_{n=1}^{\infty} \frac{1}{n^{1+s}} \int_n^{\infty} \frac{\hat{g}^2(\frac{1}{2\pi} \log x/n)}{x^{s+1}} dx \,.$$

This in turn equals

$$2 \sum_{n=1}^{\infty} \frac{1}{n^{1+s}} \sum_{m>n} \frac{\hat{g}^2(\frac{1}{2\pi} \log n/m)}{m^{1+s}} - 2 \sum_{n=1}^{\infty} \frac{1}{n^{1+s}} \int_n^{\infty} \frac{\hat{g}^2(\frac{1}{2\pi} \log x/n)}{x^{s+1}} dx + \zeta(2+2s) =$$
$$2 \sum_{n=1}^{\infty} \frac{1}{n^{1+s}} \int_n^{\infty} \{x\} \left(\frac{\hat{g}^2(\frac{1}{2\pi} \log x/n)}{x^{s+1}} \right)' dx + \zeta(2+2s) =$$
$$2 \sum_{n=1}^{\infty} \frac{1}{n^{1+s}} \int_n^{\infty} ((x)) \left(\frac{\hat{g}^2(\frac{1}{2\pi} \log x/n)}{x^{s+1}} \right)' dx \,,$$

where $((x)) = x - [x] - 1/2$. □

Before gathering all the terms together, let us again note that for $c(T) \gg T^{1/2+\varepsilon}$ is already within existing theory, though our computation above, of course, yields the same result. So for the purposes of simplicity, let us assume that $c(T) \ll T^{1/2+\varepsilon}$. Then, assuming that the error term in Motohashi's integral (2.14) gives a power saving $O(T^{-d})$ for some $d > 0$,

$$\begin{aligned} \mathcal{M}_{2,2}(T; g) = \frac{1}{T} \int_0^T \overline{Q}(t, g) dt + \frac{1}{T} \iint E_{\mathbb{I}}(\alpha, \beta) g(\alpha) g(\beta) d\alpha d\beta dt = \\ 2 \int_0^N \hat{g}^2(y/2\pi) P_3(N-y) dy + \sum_{j=0}^{2} a_{2-j}(g) (N + 2\gamma_0)^j + E_{g,T} \,, \end{aligned} \tag{2.24}$$

where $N = \log T/2\pi$ and $E_{g,T} \ll T^{-1/2+\varepsilon} + T^{-d} + \max_{w \gg \log T} |\hat{g}(w)|^2$.

In particular, if $g(x) = \frac{1}{2c} \mathbf{I}\{|x| \le c)\}$, then

$$\hat{g}(x/2\pi) = \frac{\sin(cx)}{cx} \,,$$

and

$$\int_0^N \hat{g}^2(y/2\pi)P_3(N-y)dy = \frac{1}{c}\left(\frac{\pi}{2}-\frac{1}{2cN}\right)P_3(N) - \frac{1}{2c^2}P_3'(N)(\log(2cN)+\gamma_0)+ \frac{1}{4c^2}NP_3''(N) - \frac{1}{24c^2}N^2P_3'''(N) - \frac{1}{48c^4}P_3'''(N) + O(1/c^6N^2) = \frac{\pi}{2c}P_3(N) - \frac{1}{2c^2}P_3'(N)\log(2cN) + \frac{1}{c^2}Q_2(N) + \frac{B_1}{c^4} + \frac{B_2}{c^2N} + O(1/c^6N^2)\,,$$

where $Q_2(x)$ is some polynomial of degree 2 independent of T and c, and $B_1, B_2 \neq 0$ are absolute constants. This implies that

$$\mathcal{M}_{2,2}(T;g) = \frac{\pi}{c}P_3(N) - \frac{1}{2c^2}P_3'(N)\log(2cN) + \frac{1}{c^2}Q_2(N) + \frac{B_1}{c^4} + \frac{B_2}{cN} + \sum_{j=0}^{2} a_{2-j}(g)(N+2\gamma_0)^j + O\left(\frac{1}{N^2}\right) \sim \frac{1}{\pi c}\log^3 T + a_0(g)\log^2 T\,.$$

Note that the $\log\log T$-factor comes from

$$\int_0^N y\hat{g}^2(y/2\pi)dy = \frac{1}{c^2}\int_0^N \frac{\sin^2(cy)dy}{y}\,,$$

and the error term

$$\frac{B_2}{c^2N} + O\left(\frac{1}{c^2N^6}\right) + O(\max_{w\gg\log T}|\hat{g}(w)|^2) = \frac{B_2}{c^2N} + O\left(\frac{1}{c^2N^2}\right)$$

is not polynomial in T. Both of these issues vanish if $g_0(x)$ is an appropriate weight function analytic on a vertical strip. In particular, by the Paley-Wiener theorem this implies $\hat{g}_0(y) \ll e^{-Ay}$ for some $A > 0$, thus for $g(x) = c^{-1}g_0(x/c)$

$$\int_0^N \hat{g}^2(y/2\pi)P_3(N-y)dy = \sum_{j=0}^{3}\frac{(-1)^j}{c^{j+1}j!}P_3^{(j)}(N)\int_0^\infty y^j\hat{g}^2(y/2\pi)dy + O(N^3e^{-4\pi cAN})\,.$$

In particular, by the Plancherel theorem the leading coefficient is $\frac{\pi}{c}\|\hat{g}_0\|_2^2 = \frac{\pi}{c}\|g_0\|_2^2$.

As $c(T)$ grows, if $c(T) \gg \log T$, the off-diagonal term becomes the main term. The leading term of $\int_0^T I_2(t,g)dt$ is

$$a_0(g)\log^2\frac{T}{2\pi} \sim \log^2\frac{T}{2\pi}\,.$$

This is of course consistent with the probabilistic interpretation

$$\mathbb{E}[XY] = \frac{1}{T}\int_0^T |\zeta(1/2+it+ih_1)|^2|\zeta(1/2+it+ih_2)|^2dt \approx \left(\frac{1}{T}\int_0^T |\zeta(1/2+it+ih)|^2dt\right)^2 = \mathbb{E}[X]\mathbb{E}[Y]$$

once random variables X and Y are uncorrelated.

Now let us prove (2.14) indeed applies when $\alpha_j = \beta_j$.

2.4 Groundwork

Let us formulate the problem on the interval $[T; 2T]$. Summing over dyadic intervals $[T_1; T_2] = [T/2^j; T/2^{j+1}]$ then translates the result over $[T; 2T]$ to $[0; T]$.

Let T be large. Let $t \in [T, 2T]$, and let $\alpha, \beta \in \mathbb{R}$ such that $0 \le \alpha \le \beta \ll T^{2-\varepsilon}$ with $\delta := \beta - \alpha$. We can also allow $\alpha, \beta < 0$ as long as $\alpha, \beta = o(T)$. Let $W(t)$ be an appropriate weight function. In the classic formulation $W(t)$ is the indicator function of the interval $[T, 2T]$

$$W(t) = \mathbf{I}_{[T,2T]}(t)\,,$$

but this choice of function is more difficult for analysis. Another possible choice is

$$W(t) = G_\Delta * \mathbf{I}_{[T,2T]}(t)\,, \tag{2.25}$$

where $G_\Delta(t) = \Delta^{-1} G(t/\Delta)$ is a normalised smooth (infinitely differentiable) function. Here Δ is the parameter regulating the precision of approximation satisfying

$$T^{1/2+\varepsilon} \ll \Delta \ll T^{1-\varepsilon}\,.$$

For example, one can take

$$G_\Delta(t) = (\sqrt{\pi}\Delta)^{-1} e^{-t^2/\Delta^2}\,.$$

Finally, one may choose $W(t)$ to be a smooth function compactly supported on $[T, 2T]$, again, with the same parameter Δ controlling the precision. For example, set

$$p(x) = \exp\left(-\frac{1}{x}\exp\left(-\frac{1}{1-x}\right)\right)$$

and for compact support on the interval $[T_1, T_2]$ take

$$W(t) = \begin{cases} p(|t - T_j|/\Delta), & |t - T_j| \le \Delta\,; \\ 1, & T_1 + \Delta \le t \le T_2 - \Delta\,; \\ 0, & \text{otherwise}\,. \end{cases} \tag{2.26}$$

Both choices (2.25) and (2.26) essentially give

$$W^{(k)}(t) \ll \Delta^{-k} \mathbf{I}\{|t - T_j| \le \Delta \log T\}\,,$$

and $|\operatorname{supp} \mathcal{F}W| \ll \Delta^{-1} \log T$ up to negligible error.

In most cases, we do not have to choose a specific function W since it is enough to know that there exists an infinitely differentiable function of compact support approximating the indicator function of an interval with arbitrary precision, and with

derivatives as above. Though sometimes it is convenient to have a specific function in mind to simplify computations.

Now consider the integral

$$M(W,\alpha,\beta) = \int_0^\infty W(t)|\zeta(\frac{1}{2}+it+i\alpha)|^2|\zeta(\frac{1}{2}+it+i\beta)|^2dt = \int_0^\infty W(t)\kappa(t)\zeta^2(\frac{1}{2}+it+i\alpha)\zeta^2(\frac{1}{2}-it-i\beta)dt\,,$$

where we used the functional equation of zeta $\zeta(s) = \chi(s)\zeta(1-s)$ so

$$\kappa(t) = \kappa_{\alpha,\beta}(t) := \chi\left(1/2 - i(t+\alpha)\right)\chi\left(1/2 + i(t+\beta)\right)\,. \tag{2.27}$$

Sometimes it would also be convenient to use the approximation for $|t| \geq 1$

$$\kappa(t) = e^{i\phi(t)}(1+O(t^{-1}))\,, \tag{2.28}$$

where

$$\phi(t) = \phi_{\alpha,\beta}(t) := (t+\alpha)\log\frac{t+\alpha}{2\pi e} - (t+\beta)\log\frac{t+\beta}{2\pi e}\,. \tag{2.29}$$

In particular, for $\alpha \leq \beta \ll T^{1/2-\varepsilon}$ the phases almost cancel out with

$$e^{i\phi(t)} = (t/2\pi)^{i(\alpha-\beta)}(1+O(\delta|\alpha+\beta|/T))\,.$$

Since W only deviates from $\mathbf{I}_{[T,2T]}$ in a meaningful way when $|t-T_j| \ll T^\varepsilon\Delta$ we have

$$|M(W,\alpha,\beta) - M(\mathbf{I}_{[T,2T]},\alpha,\beta)| \ll T^\varepsilon\Delta\,,$$

so instead of the integral on the interval $[T,2T]$ we will be evaluating a weighted integral.

All the results apply for both choices of smoothing (2.25) and (2.26). The distinction boils down to working with general Schwartz functions, or Schwarz functions with compact support ("bump functions"). We will write Schwartz functions with (approximately) finite support to reflect this difference that a general Schwarz function is not compactly supported, but it still "concentrated" on a compact set up to small error. To avoid abuse of notation, we will write "negligible error" to mean $O((1+|x|)^{-b})$ for any $b > 0$ and omit this error altogether.

Due to the choice of smoothing all sums and integrals throughout the chapter converge absolutely unless stated otherwise, thus we can change their order freely due to Fubini's theorem.

Finally, due to the rate of decay, we can essentially use

$$\int_0^{+\infty} W(t)f(t)dt,\ \int_{-\infty}^{+\infty} W(t)f(t)dt,\ \int_{t\in\operatorname{supp} W} W(t)f(t)dt$$

interchangeably.

2.4.1 Approximate functional equation

Using a similar idea to the heuristic calculation presented in the section on general conjectures, we would first like to show that even though the Dirichlet series of $\zeta(s)$ does not converge for $\Re s < 1$, we can still use this representation with some adjustment. This is in fact a rigorous statement known as the approximate functional equation of $\zeta(s)$ connecting $\zeta(s)$ and its truncated Dirichlet series. Thus we start with constructing an approximate functional equation for

$$\zeta^2(\frac{1}{2}+it+i\alpha)\zeta^2(\frac{1}{2}-it-i\beta)\,.$$

Recall the functional equation of zeta

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \pi^{-(1-s)/2}\Gamma((1-s)/2)\zeta(1-s)\,, \tag{2.30}$$

and let

$$\Gamma_{\alpha,\beta}(z;t) := \left(\Gamma(\frac{1}{2}(z+it+i\alpha))\Gamma(\frac{1}{2}(z-it-i\beta))\right)^2\,. \tag{2.31}$$

Then the functional equation (2.30) implies that

$$\Lambda_{\alpha,\beta}(z;t) := \pi^{-2z+i(\beta-\alpha)}\Gamma_{\alpha,\beta}(z;t)\zeta^2(z+it+i\alpha)\zeta^2(z-it-i\beta) \tag{2.32}$$

satisfies the functional equation

$$\Lambda_{\alpha,\beta}(z;t) = \Lambda_{\beta,\alpha}(1-z;t)\,. \tag{2.33}$$

and is meromorphic on $\mathbb{C}$ with double poles at

$$z \in P_{t,\alpha,\beta} := \{1-it-i\alpha, -it-i\alpha, 1+it+i\beta, it+i\beta\}\,.$$

Now we would like to construct an approximate functional equation using a smooth approximation of the cutoff function

$$\frac{1}{2\pi i}\int_{(\sigma)} y^s\frac{ds}{s} = \begin{cases} 1, & y>1\,; \\ 1/2, & y=1\,; \\ 0, & y<0\,. \end{cases}$$

Let $G(z)$ be a nice analytic function; in particular, choose $G(s)$ to be even, entire and rapidly decaying on vertical strips with $G(0)=1$. Then to avoid any convergence issues we can replace the integral above with

$$\frac{1}{2\pi i}\int_{(\sigma)} y^s G(s)\frac{ds}{s}\,,$$

where σ is any real number such that $\sigma > 1$. Usually we can think of $G(z)$ as $G(z) = e^{z^2}$.

Lemma 2.4.1. *Let $T \geq 10$, and $Q = Q(T)$. Let $G(z) = G(z, Q) = e^{z^2/Q}$, and*

$$V_{\alpha,\beta}(x,t) := \frac{1}{2\pi i}\int_{(1)} (\pi^2 x)^{-z} \frac{\Gamma_{\alpha,\beta}(1/2+z;t)}{\Gamma_{\alpha,\beta}(1/2;t)} G(z)\frac{dz}{z}\,.$$

Let $t \in [T, 2T]$ and $\alpha, \beta \in \mathbb{R}$ such that $t+\alpha,\ t+\beta \gg Q^{1/2}T^{\varepsilon}$, then for any $a > 0$

$$|\zeta(1/2+it+i\alpha)\zeta(1/2-it-i\beta)|^2 = \\ 2\,\mathrm{Re}\sum_{n,m=1}^{\infty} \frac{d(n)d(m)}{n^{1/2+it+i\alpha}m^{1/2-it-i\beta}}\kappa(t)V_{\alpha,\beta}(nm,t) + O(T^{-a})\,.$$

Proof. Consider the contour integral

$$\frac{1}{2\pi i}\int_{(1)} \Lambda_{\alpha,\beta}(1/2+z;t)G(z)\frac{dz}{z}\,.$$

Move the line of integration to $\Re z = -1$. The residue at $z = 0$ gives $\Lambda_{\alpha,\beta}(1/2;t)$, and residues at poles have small contributions. Indeed, using Stirling's formula given in Lemma 2.2.5 in the form

$$\log\Gamma(z) = (z-1/2)\log z - z + O(1)$$

and standard estimates $\zeta(\sigma+it), \zeta'(\sigma+it) \ll t^2$ when $t \gg 1$ and $|\sigma| \leq 1$ we have

$$\frac{1}{\Gamma_{\alpha,\beta}(1/2;t)}\sum_{z\in P_{t,\alpha,\beta}} \operatorname*{Res}_z \left(\Lambda_{\alpha,\beta}(1/2+z;t)\frac{G(z)}{z}\right) \ll T^B \sum_{z\in P_{t,\alpha,\beta}} |G(z)| \ll T^B e^{-T^{\varepsilon}}$$

for some fixed absolute constant B. Thus the above is $O(T^{-a})$ for any $a > 0$.

Changing variables to $w = -z$ and using functional equation (2.33) we then obtain

$$\frac{\Lambda_{\alpha,\beta}(1/2;t)}{\Gamma_{\alpha,\beta}(1/2;t)} = \frac{1}{2\pi i}\int_{(1)} \frac{\Lambda_{\alpha,\beta}(1/2+z;t) + \Lambda_{\beta,\alpha}(1/2+z;t)}{\Gamma_{\alpha,\beta}(1/2;t)} G(z)\frac{dz}{z} + O(T^{-a})\,.$$

Using

$$\frac{\Gamma_{\alpha,\beta}}{\Gamma_{\beta,\alpha}}(1/2;t) = \kappa^2(t)$$

and expanding $\zeta^2(1/2+z+it+i\alpha)\zeta^2(1/2+z-it-i\beta)$ into a double series yields

$$\sum_{n,m=1}^{\infty} \frac{d(n)d(m)}{\sqrt{nm}}\left(\frac{m^{i(t+\beta)}}{n^{i(t+\alpha)}}V_{\alpha,\beta}(nm,t) + \kappa^{-2}(t)\frac{m^{i(t+\alpha)}}{n^{i(t+\beta)}}V_{\beta,\alpha}(nm,t)\right) + O(T^{-a})\,.$$

The statement of the lemma then follows immediately. Replacing $\kappa(t)$ with $e^{i\phi(t)}$ as in (2.28) increases the error term to $O(T^{-1/2})$. □

This form of the approximate functional equation is not very convenient for analysis, but what it really does is truncate the double Dirichlet series at $mn \ll (T+\alpha)(T+\beta)$. In the next subsection we make this statement rigorous.

2.4.2 Simplifying weights

Throughout the section we write $W(t)$ for either of the choices of smoothing (2.25) and (2.26) with parameter Δ. We also take $G(z) = G(z, Q) = e^{z^2/Q}$.

Our computations will be contingent on parameters Q and Δ. We will assume that $T^\varepsilon \ll Q \ll T$ and $T^{1/2+\varepsilon} \ll \Delta \ll T^{1-\varepsilon}$. Moreover, we will assume that $\Delta Q^{1/2} \ll T$, though this is not necessary. We will choose specific Q and Δ at the very end, and the choice will depend on both T and $\delta = \beta - \alpha$.

Let us simplify the weights $V_{\alpha,\beta}(nm, t)$. In particular, we will show that we can replace $V_{\alpha,\beta}(nm, t)$ by a well-understood contour integral

$$J(x,t) = J_{\alpha,\beta}(x,t) := \frac{1}{2\pi i}\int_{(1)} \left(\frac{(t+\alpha)(t+\beta)}{4\pi^2 x}\right)^z G(z)\frac{dz}{z} \tag{2.34}$$

as we alluded to in the previous subsection. Let $\xi = \log \frac{(t+\alpha)(t+\beta)}{4\pi^2 x}$. Integral $J_{\alpha,\beta}(x,t)$ behaves very similarly to the cut-off function as for any $b > 0$

$$J_{\alpha,\beta}(x,t) = \int_{-\infty}^{\xi} \hat{G}(y/2\pi)dy = \begin{cases} G(0) + O(|\xi|^{-b}), & \text{if } \xi > 0; \\ O(|\xi|^{-b}), & \text{if } \xi \le 0\,. \end{cases}$$

Thus $J_{\alpha,\beta}(x,t) = \mathbf{I}\{\xi > 0\} + O(|\xi|^{-b})$. Moreover,

$$\frac{d}{dt}J_{\alpha,\beta}(x,t) = \frac{2t+\alpha+\beta}{(t+\alpha)(t+\beta)}\hat{G}(\xi/2\pi)\,. \tag{2.35}$$

In particular, for $G(z) = e^{z^2/Q}$

$$J_{\alpha,\beta}(x,t) = \sqrt{Q}\int_{-\infty}^{\xi} e^{-\frac{Q}{4}y^2}dy = \begin{cases} 2 + O(e^{-\frac{Q}{4}\xi^2}), & \text{if } \xi > 0; \\ O(e^{-\frac{Q}{4}\xi^2}), & \text{if } \xi \le 0\,. \end{cases}$$

We start with approximating

$$\frac{\Gamma_{\alpha,\beta}(1/2+z;t)}{\Gamma_{\alpha,\beta}(1/2;t)}$$

using Stirling's formula.

Lemma 2.4.2. *Let $t \in [T, 2T]$ for some large number $T > 0$. Let $\Gamma_{\alpha,\beta}(z;t)$ be as above, and let M be some cutoff parameter such that $M \ll T^{1/2+\varepsilon}$. Let $z \in \mathbb{C}$ with imaginary part $y = \Im z$ and fixed real part. Then*

$$\frac{\Gamma_{\alpha,\beta}(1/2+z;t)}{\Gamma_{\alpha,\beta}(1/2;t)} = \begin{cases} \exp\left(z\log\frac{(t+\alpha)(t+\beta)}{4}\right)(1 + E_{\alpha,\beta}(z;t)), & \text{if } y \le M\,; \\ e^{O(T+|y|)}, & \text{otherwise}\,. \end{cases}$$

Here $E_{\alpha,\beta}(z;t)$ is the error term represented by a convergent series and bounded in terms of α, β, M and T.

Proof. Let us start with $y = \Im z > M$. In this regime Stirling's formula gives the trivial bound

$$\frac{\Gamma_{\alpha,\beta}(1/2+z;t)}{\Gamma_{\alpha,\beta}(1/2;t)} \ll (T+|y|)^C$$

for some absolute constant $C > 0$.

For $y = \Im z < M$ Stirling's formula in the form of the asymptotic expansion gives

$$\frac{\Gamma_{\alpha,\beta}(1/2+z;t)}{\Gamma_{\alpha,\beta}(1/2;t)} = \exp\left(z\log\frac{(t+\alpha)(t+\beta)}{4}\right)(1+R_{\alpha,\beta}(z;t)+O(T^{-N}))\,, \tag{2.36}$$

where N can be chosen as large as necessary and

$$R_{\alpha,\beta}(z;t) = \sum_{n,m} b_{n,m}\left(\frac{z^n}{(t+\alpha)^m} + \frac{(-1)^m z^n}{(t+\beta)^m}\right) = \frac{iz^2(\beta-\alpha)}{(t+\alpha)(t+\beta)} + \frac{z^3}{(t+\alpha)^2} + \dots \tag{2.37}$$

for some complex coefficients $b_{n,m}$ and a finite range of m and n depending on N. Moreover, the indices satisfy $m \geq n-1$ for $n \geq 3$ and $m \geq n$ for $n \leq 2$. □

Lemma 2.4.3. *Let $0 \leq \alpha \leq \beta \ll T^{2-\varepsilon}$ with $\delta = \beta - \alpha$. Set $\tilde{T} = \sqrt{(T+\alpha)(T+\beta)}$. Let $B_{\alpha,\beta}(x,t) = V_{\alpha,\beta}(x,t) - J_{\alpha,\beta}(x,t)$, then*

$$\int_{-\infty}^{+\infty} W(t) \sum_{n,m\geq 1} \frac{d(n)d(m)}{\sqrt{nm}} \frac{m^{i(t+\beta)}}{n^{i(t+\alpha)}} \kappa(t) B_{\alpha,\beta}(nm,t)dt \ll$$
$$T^{\varepsilon}\begin{cases} \frac{Q^{1/2}}{\Delta}(\delta+Q^{1/2}), & \text{if } \delta \ll Q\,; \\ \delta^{1/2}Q + \frac{Q^{3/2}(T+\beta)}{\delta^{1/2}}, & \text{if } \delta \gg T^{\varepsilon}\,. \end{cases}$$

In particular, this is $O(T^{1/2+\varepsilon})$ when $\delta \ll T^{1/2+\varepsilon}$ and $T^{\varepsilon}\delta^{1/2}Q$ when $\delta \gg T^{1/2+\varepsilon}$.

Proof. First, Fubini's theorem allows us to change the order of summation and integration due to absolute convergence. To bound the internal integral in

$$\sum_{n,m\geq 1} \frac{d(n)d(m)}{\sqrt{nm}} \frac{m^{i\beta}}{n^{i\alpha}} \int_{-\infty}^{+\infty} W(t)\big(\frac{m}{n}\big)^{it} \kappa(t) B_{\alpha,\beta}(nm,t)dt \tag{2.38}$$

let us first bound $B_{\alpha,\beta}(nm,t)$.

When $y = \Im z > \sqrt{Q}\log T$ a trivial bound on $\Gamma_{\alpha,\beta}$ suffices as $G(z)$ is Schwartz. In particular, Lemma 2.4.2 with $M = Q^{1/2}\log T$ gives

$$\int_{y>\sqrt{Q}\log T} (\pi^2 x)^{-z} \frac{\Gamma_{\alpha,\beta}(1/2+z;t)}{\Gamma_{\alpha,\beta}(1/2;t)} G(z)\frac{dz}{z} \ll x^{-1}T^{-a}\,.$$

When $y = \Im z \le \sqrt{Q}\log T$ apply the asymptotic expansion in (2.36) and (2.37). For any integer $\nu \ge 0$, any real number $\mu \in \mathbb{R}$ and any $a > 0$ we have

$$\int_{y\le\sqrt{Q}\log T} e^{2\pi z\mu} z^\nu G(z)dz = \int_{(1)} e^{2\pi z\mu} z^\nu G(z)dz + O(T^{-a})\,.$$

We have $G(z) = F(z/\sqrt{Q})$, so by Lemma 2.2.6 and properties of the Schwartz class

$$\int_{(1)} e^{2\pi z\mu} z^\nu G(z)dz = \frac{d^\nu}{d\mu^\nu}\hat{G}(\mu) \ll Q^{(\nu+1)/2}|\mu|^{-b}$$

for any $b > 0$. In particular, for $G(z) = e^{z^2/Q}$ with $\hat{G}^{(\nu)}(\mu) \ll Q^{(\nu+1)/2}e^{-\frac{Q}{8}\mu^2}$. Taking $\mu = \log\frac{(t+\alpha)(t+\beta)}{4\pi^2 mn}$ and looking at the largest terms in (2.37) we then have

$$B_{\alpha,\beta}(x,t) \ll Ae^{-\frac{Q}{8}\xi^2} + x^{-1}T^{-a}\,, \tag{2.39}$$

where

$$A = A(\alpha,\beta,Q,T) = \frac{\delta Q}{(T+\alpha)(T+\beta)} + \frac{Q^{3/2}}{(T+\alpha)^2} = \frac{\delta Q}{\tilde{T}^2} + \frac{Q^{3/2}}{(T+\alpha)^2}\,.$$

Now let us bound the internal integral in (2.38). Applying the Fourier convolution theorem to the product $W\cdot(\kappa B_{\alpha,\beta})$ we can rewrite

$$\int_{-\infty}^{+\infty} W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)B_{\alpha,\beta}(nm,t)dt = \left(\mathcal{F}W * \mathcal{F}\kappa\, B_{\alpha,\beta}(nm,\cdot)\right)(\log m/n)\,. \tag{2.40}$$

For both choices of W (with parameter Δ) we have

$$\hat{W}(y) \ll (1+\Delta|y|)^{-k}(1/T+|y|)^{-1}\,.$$

Recall that $B_{\alpha,\beta}(x,t)$ had an expression in terms of derivatives of G. Moreover, it essentially restricts its argument t to the range where

$$\xi = \xi(t) = \log\frac{(t+\alpha)(t+\beta)}{4\pi^2 mn} = \log\frac{(t+\alpha)(t+\beta)}{4\tau^2} \ll Q^{-1/2}\log T\,.$$

Let $\tau = 2\pi\sqrt{nm}$, and let q satisfy $(q+\alpha)(q+\beta) = \tau^2$. Set $\eta_y(x) := y + \phi'(x)$ and $\eta(x) := \log m/n + \phi'(x)$. Then $\xi(t) \ll Q^{-1/2}\log T$ implies

$$\xi = \log\frac{(t+\alpha)(t+\beta)}{4\tau^2} = \frac{w}{\tau^*} + O\left(\frac{w^2}{(\tau^*)^2}\right) \ll \frac{\log T}{Q^{1/2}}\,,$$

where $w = t - q$ and $\tau^* = \tau^2/(2q+\alpha+\beta)$.

Suppose $\delta \ll Q^{1-\varepsilon}$, then

$$\int e^{iyt}\kappa(t)\hat{G}^{(\nu)}(\xi(t)/2\pi)dt = e^{iyq}\kappa(q)\int e^{iw\eta_y(q)}\hat{G}^{(\nu)}(w/2\pi\tau^*)(1+E(w))dw\,,$$

where $E(w)$ is a convergent series with largest terms

$$\frac{iw^2\delta}{\tau^2} + \frac{Q^{1/2}w^2}{(\tau^*)^2} \ll \frac{\delta}{Q} + \frac{1}{Q^{1/2}}\,.$$

Thus $\mathcal{F}f^{(k)}(u) = (2\pi iu)^k \mathcal{F}f(u)$ together with (2.37) implies

$$\mathcal{F}(\kappa\, B_{\alpha,\beta})(y) \ll A\tau^* Q^{-1/2} G(\tau^* \eta_y(q))^{1+\varepsilon}\,.$$

Then (2.40) is upper-bounded by

$$\begin{aligned}
&AT^{1+\varepsilon}Q^{-1/2}\int_{-\infty}^{+\infty} G(\tau^*\eta_y(q))\hat{W}(\frac{1}{2\pi}(\log m/n - y))dy \ll \\
&AT^{1+\varepsilon}Q^{-1/2}|\{y : |\log m/n - y| \le \Delta^{-1}\log T,\ |\eta_y(q)| \le Q^{1/2}\log T/T\}| + \\
&AT^{2+\varepsilon}Q^{-1/2}|\{y : |\log m/n - y| \le T^{-1}\log T,\ |\eta_y(q)| \le Q^{1/2}\log T/T\}| \ll \\
&AT^{1+\varepsilon}Q^{-1/2}\,,
\end{aligned}$$

which is only non-negligible when

$$\eta(q) \ll \frac{\log T}{\Delta} + \frac{Q^{1/2}\log T}{T} \ll \frac{\log T}{\Delta}\,.$$

Summing over m, n, this is bounded by

$$\frac{AT^{1+\varepsilon}}{Q^{1/2}} \sum_{nm \ll \tilde{T}^2} \frac{1}{\sqrt{nm}} \mathbf{I}\{\eta(q) \ll \frac{\log T}{\Delta}\} \ll \frac{Q^{1/2}}{\Delta}(\delta + Q^{1/2})\,.$$

When δ is large, we may also use the oscillation lemma (Lemma 2.2.2) for the second derivative to bound the integral in question. If $\delta \gg T^{\varepsilon}$, then

$$\int_{-\infty}^{+\infty} W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)B_{\alpha,\beta}(nm,t)dt \ll \frac{\max|B_{\alpha,\beta}(nm,t)|}{\min|\phi''(t)|^{1/2}} \ll \frac{A\tilde{T}}{\delta^{1/2}}\,. \tag{2.41}$$

This bound is not quite optimal, but it suffices for our purposes. Summing over m, n,

$$\frac{A\tilde{T}}{\delta^{1/2}} \sum_{nm \ll (T+\alpha)(T+\beta)} \frac{1}{\sqrt{nm}} \ll \delta^{1/2}Q + \frac{Q^{3/2}(T+\beta)}{\delta^{1/2}}\,.$$

Between the two regimes we can always choose Δ and Q to obtain a power saving for $\alpha, \beta \ll T^{2-\varepsilon}$. In particular,

$$\int_{-\infty}^{+\infty} W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)B_{\alpha,\beta}(nm,t)dt \ll \begin{cases} T^{1/2} + \frac{Q^{1/2}\delta}{\Delta}, & \delta \ll Q \text{ and } Q \ll T; \\ \delta^{1/2}Q + \frac{Q^{3/2}(T+\beta)}{\delta^{1/2}}, & \delta \gg T^{\varepsilon}\,, \end{cases} \tag{2.42}$$

where for $\delta \gg T^{1/2}$ we can set $Q = T^{\varepsilon}$. □

We can remove $J(mn,t)$ completely and turn our weighted Dirichlet series into a truncated one. This is not necessary for our computation, but let us show how to do that nonetheless to illustrate this equivalence between weighting and truncating.

Lemma 2.4.4. *Let $0 \le \alpha \le \beta \ll T^{1-\varepsilon}$ such that $\delta = \beta - \alpha \ll Q^{1-\varepsilon}$. Let $\tau = 2\pi\sqrt{nm}$, and let q satisfy $(q+\alpha)(q+\beta) = \tau^2$. Then*

$$M(W;\alpha,\beta) = 2\,\mathrm{Re}\sum_{\tau\le 2T}\frac{d(n)d(m)}{n^{1/2+i\alpha}m^{1/2-i\beta}}\int_q^\infty W(t)\kappa(t)\left(\frac{m}{n}\right)^{it}dt + T^\varepsilon O(\frac{T}{\sqrt{Q}} + \frac{Q(\delta+Q^{1/2})}{T} + \frac{T\delta}{Q}).$$

Proof. Let

$$\xi = \log\frac{(t+\alpha)(t+\beta)}{4\pi^2 nm}\,,$$

and consider

$$\int_{-\infty}^{\infty} W(t)\kappa(t)\left(\frac{m}{n}\right)^{it}(J_{\alpha,\beta}(nm,t) - \mathbf{I}\{\xi\ge 0\})dt\,.$$

If $|q - T_j| \ll T\log T/\sqrt{Q} + \Delta\log T \ll T\log T/\sqrt{Q}$, then applying the oscillation lemma gives that the above integral is bounded by

$$\frac{T}{\sqrt{Q+\delta}} \wedge \frac{\delta^{1/2}}{|\log m/n|}\,.$$

Thus the contribution of such $q = q(m,n)$ to the total sum is bounded by

$$T^\varepsilon\left(\frac{T}{\sqrt{Q}} + \Delta\right)\left(1 + \frac{\delta}{(\delta+Q)^{1/2}}\right) \ll \frac{T^{1+\varepsilon}}{\sqrt{Q}}\,.$$

The contribution of $q \notin [T, 2T]$ is negligible due to rate of decay of W. For the rest of the range we only need $|t-q| \ll T\log T/\sqrt{Q}$.

Let $\eta(t) = \log m/n + \phi'(t)$. If $\eta(q) \ll 1/T$,

$$\int_{-\infty}^{+\infty} W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)(J_{\alpha,\beta}(nm,t) - \mathbf{I}\{\xi\ge 0\})dt \ll \frac{T}{\sqrt{Q+\delta}}\,.$$

For $\eta(q) \gg 1/T$ by partial integration we have

$$\int_{-\infty}^{+\infty} W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)(J_{\alpha,\beta}(nm,t) - \mathbf{I}\{\xi\ge 0\})dt = \frac{1}{i\eta(q)}\int_{-\infty}^{+\infty}\big(\frac{m}{n}\big)^{it}\kappa(t)\left(W(t)\kappa(t)e^{-(t-q)\eta'(q)}(J_{\alpha,\beta}(nm,t) - \mathbf{I}\{\xi\ge 0\})\right)' dt\,.$$

This gives us three integrals. The first one is negligible

$$\int_{-\infty}^{+\infty}\big(\frac{m}{n}\big)^{it}\kappa(t)W'(t)(J_{\alpha,\beta}(nm,t) - \mathbf{I}\{\xi\ge 0\})dt \ll T^{-a}$$

since $|t-q| \ll T\log T/\sqrt{Q}$ and $|t-T_j| \le \Delta \log T$ do not overlap. Next,

$$\int_{-\infty}^{+\infty} \big(\frac{m}{n}\big)^{it} \kappa(t) W(t) \left(\phi'(t)-\phi'(q)\right) \left(J_{\alpha,\beta}(nm,t) - \mathbf{I}\{\xi \ge 0\}\right) dt \ll \frac{\delta}{TQ^{1/2}} \left(\frac{T}{Q^{1/2}} \wedge \frac{\delta^{1/2}}{|\log m/n|} \right)$$

since $\phi'(t)-\phi'(q) \ll \delta|t-q|/T^2$. The final integral equals

$$\int_{-\infty}^{+\infty} W(t) \big(\frac{m}{n}\big)^{it} \kappa(t) (J_{\alpha,\beta}(nm,t) - \mathbf{I}\{\xi \ge 0\})' dt = -W(q)\big(\frac{m}{n}\big)^{iq} \kappa(q) + \int_{-\infty}^{+\infty} W(t) \big(\frac{m}{n}\big)^{it} \kappa(t) \frac{2t+\alpha+\beta}{(t+\alpha)(t+\beta)} \hat{G}(\xi(t)) dt \,.$$

For $w = t-q$ and $\tau^* = \tau^2/(2q+\alpha+\beta)$

$$\xi(t) = \xi(q+w) = \log \frac{(t+\alpha)(t+\beta)}{4\pi^2 mn} = \frac{w}{\tau^*} + O\left(\frac{w^2}{\tau^2}\right)$$

thus the above integral equals

$$W(q)\big(\frac{m}{n}\big)^{iq} \kappa(q) \frac{Q^{1/2}}{\tau^*} \int_{-\infty}^{+\infty} e^{iw\eta(q)} e^{-\frac{Q}{(\tau^*)^2} w^2} (1 + E_1(w) + iE_2(w)) dw \,, \tag{2.43}$$

where $E_j(w)$ are convergent real-valued series with largest terms $\frac{w}{T} + \frac{w^3 Q}{T^3}$ and $\frac{w^2\delta}{T^2}$ respectively. This gives us

$$W(q)\big(\frac{m}{n}\big)^{iq} \kappa(q) e^{-\frac{(\tau^*)^2\eta(q)^2}{4Q}} (1 + E_1 + iE_2) \,.$$

with $E_1 \ll Q^{-1/2}$ and $E_2 \ll \delta/Q$. Computing the Fourier transform in (2.43) and gathering all the phases, we see that the contribution of this term to the sum over m, n is

$$\mathrm{Re} \sum_{n,m} \frac{d(n)d(m)}{(nm)^{1/2} i\eta(q)} e^{i\psi_{\alpha,\beta}(n,m)} W(q) (e^{-\frac{(\tau^*)^2\eta(q)^2}{4Q}} - 1)(1+E_1) + \sum_{n,m} \frac{d(n)d(m)}{(nm)^{1/2}\eta(q)} W(q) (e^{-\frac{(\tau^*)^2\eta(q)^2}{4Q}} - 1) E_2 \,,$$

where

$$\psi_{\alpha,\beta}(n,m) = (q + (\alpha+\beta)/2) \log \frac{m}{n} + \phi(q) + \frac{\delta}{2} \log mn \,.$$

Set $m - n = \delta/2\pi + r$, then

$$\psi_{\alpha,\beta}(n,m) = \frac{\sqrt{\delta^2+4\tau^2}}{2} \eta(q) = 2\pi r + \frac{4\pi^2 r^2 \delta}{(\delta + 4\pi n)^2} + O\Big(\frac{r^3}{(\delta+4\pi n)^2}\Big) \,, \tag{2.44}$$

thus the first term has contribution at most

$$T^{\varepsilon} \sum_{\eta \ll \sqrt{Q} \log T/T} \frac{W(q)}{(nm)^{1/2}\eta(q)} \frac{r^2\delta}{(\delta + 4\pi n)^2} \ll \frac{Q(\delta + Q^{1/2})}{T}\,, \tag{2.45}$$

and the second

$$T^{\varepsilon} \sum_{\eta \ll \sqrt{Q} \log T/T} \frac{W(q)E_2}{(nm)^{1/2}\eta(q)} \ll \frac{T\delta}{Q}\,, \tag{2.46}$$

The total error is then

$$\frac{T^{1+\varepsilon}}{\sqrt{Q}} + \frac{Q(\delta + Q^{1/2})}{T} + \frac{T\delta}{Q}\,.$$

When $\delta \ll Q \ll T$ there is always a power saving. □

Now we would like to evaluate the diagonal $m = n$ and the off-diagonal $m \neq n$ contributions separately. The former is quite simple, while the latter constitutes the bulk of the computation. Thus we start with the off-diagonal terms.

In our problem there are two distinct regimes. When δ is small, the integral

$$\int W(t) J_{\alpha,\beta}(nm, t) \big(\frac{m}{n}\big)^{it} \kappa(t) dt$$

behaves similarly to the Fourier transform of $WJ_{\alpha,\beta}$ at $\eta(T)$. When δ is large, the main contribution comes from the stationary point instead.

The off-diagonal term relies on bounds for the error term in the correlations on the divisor function

$$\sum_{n \leq X} d(n)d(n+r)\,.$$

Let us first discuss what is known for such sums.

2.4.3 Correlations of the divisor function

Let $d(n) = \sum_{k|n} 1$ be the divisor function, and let $r \in \mathbb{N}$, then

$$\sum_{n \leq N} d(n)d(n+r) = S(N, r) + E(N, r)\,,$$

where $S(x, r)/x$ is a polynomial of degree 2 in $\log x$ with coefficients depending on r, and $E(x, r) = o(x)$ is the error term. In fact, we have the following expansions [38] for the main term and its derivative:

$$\begin{aligned} S(x, r) &= x \sum_{k=1}^{\infty} \frac{\mu(k)}{k^2} \sum_{d|r} \frac{1}{d} \left(\left(\log \frac{xe^{2\gamma_0 - 1}}{d^2 k^2} \right)^2 + 1 \right)\,; \\ S'(x, r) &= \sum_{k=1}^{\infty} \frac{\mu(k)}{k^2} \sum_{d|r} \frac{1}{d} \left(\log \frac{xe^{2\gamma_0}}{d^2 k^2} \right)^2\,. \end{aligned} \tag{2.47}$$

Let us first state the result for correlations of the divisor function without using any spectral theory.

Theorem 2.4.5 (Motohashi [106]). *Let $N \geq 1$ be large, and let $r \in \mathbb{N}$ such that $1 \leq r \leq N^{10/7}$. Let $\sum_{n \leq N} d(n)d(n+r) = S(N,r) + E(N,r)$, then uniformly for all r in this range*

$$E(N,r) \ll (N(N+r))^{1/3+\varepsilon} + r^{9/40}(N(N+r))^{1/2} + r^{7/10}N^{\varepsilon}\,.$$

In particular, $E(N,r) \ll N^{2/3+\varepsilon}$ when $r \ll N^{20/27}$.

The error term here relies on certain spectral bounds, and cannot be improved further than $N^{1/2}$ as shown in the same work [106].

In addition, we have the following results on the mean error term due to Ivic and Motohashi [82, 83]. Let $R \ll N^{1/2-\varepsilon}$, then

$$\sum_{r=1}^{R} |E(N,r)| \ll (RN)^{2/3+\varepsilon} \tag{2.48}$$

uniformly in R. Additionally, if $t_j(n) \leq n^a$ as in (2.8), then

$$\int_{N}^{2N} E^2(x,r)dx \ll r^{2a}N^{2+\varepsilon}$$

uniformly for $r \leq N^{1/(2-2a)}$.

Now let us move on to the explicit expression. Let $U(x)$ be an infinitely differentiable function of compact support. Let $F(a,b,c;x) = {}_2F_1(a,b,c;x)$ be the hypergeometric function. Set

$$\Xi(z;U) := \frac{\Gamma^2(1/2+z)}{\Gamma(1+2z)} \int_0^{\infty} U(x)x^{-1/2-z}F(1/2+z,1/2+z,1+2z;-1/x)dx\,, \tag{2.49}$$

and

$$\Theta(y;U) = \frac{1}{2}\,\mathrm{Re}\left(\left(1+\frac{i}{\sinh \pi y}\right)\Xi(iy;U)\right)\,. \tag{2.50}$$

Finally, let $m(x,r) := \lim_{h\to\infty} m(x,r,h)$, where

$$\begin{aligned} m(x,r,h) = \sigma_{1+2h}(r)\frac{\zeta^2(1+h)}{\zeta(2+2h)}x^h(1+x)^h + r^{2h}\sigma_{1-2h}(r)\frac{\zeta^2(1-h)}{\zeta(2-2h)} + \\ r^h\sigma_1(r)\frac{\zeta(1+h)\zeta(1-h)}{\zeta(2)}(x^h+(1+x)^h)\,. \end{aligned} \tag{2.51}$$

Let $\frac{d^l}{dy^l}\sigma_y(r) = \sigma_y^{(l)}(r) = \sum_{d|r} d^y \log^l d$. Then taking the limit as $h \to 0$ similarly to how we computed $h_0(z,s)$, obtain

$$m(x,r) = \frac{\sigma_1(r)}{\zeta(2)}(\log x \log(1+x) + \log x(1+x)(2\gamma_0 - \log r) + (2\gamma_0 - \log r)^2) + \\ (\log x(1+x) + 4\gamma_0 - 2\log r)\frac{d}{dh}\left(\frac{\sigma_{1+2h}(r)}{\zeta(2+2h)}\right)\Big|_{h=0} + \frac{d^2}{dh^2}\left(\frac{\sigma_{1+2h}(r)}{\zeta(2+2h)}\right)\Big|_{h=0} = \\ \frac{d^2}{dh^2}\left((xr^{-1}e^{2\gamma_0})^h \frac{\sigma_{1+2h}(r)}{\zeta(2+2h)}\right)\Big|_{h=0}(1 + O(1/|x|)) \ .$$

Expanding

$$\frac{\sigma_{1+2h}(r)}{\zeta(2+2h)} = \sum_{k=1}^{\infty}\frac{\mu(k)}{k^2}\sum_{d|r} d^{1+2h}k^{-2h}$$

and substituting this series into $m(x,r)$ yields

$$m(x,r) = \sum_{k=1}^{\infty}\frac{\mu(k)}{k^2}\sum_{d|r} d\frac{d^2}{dh^2}\left(\frac{d^2 x e^{2\gamma_0}}{rk^2}\right)^h\Big|_{h=0}(1+O(1/|x|)) = \\ \sum_{k=1}^{\infty}\frac{\mu(k)}{k^2}\sum_{d|r} d\log^2\frac{d^2 x e^{2\gamma_0}}{rk^2}(1+O(1/|x|)) \ ,$$

which is clearly similar to $S'(x,r)$ as in (2.47).

Recall that $H_j(s)$ is the Hecke-Maass L-function corresponding to the eigenvalue $1/4+\kappa_j^2$ of the hyperbolic Laplacian. It is also the Dirichlet series of the Hecke eigenvalue $t_j(n)$. Similarly, $H_{j,k}(s)$ is the Hecke L-function corresponding to j-th function from the orthogonal basis of the space of weight $2k$ holomorphic cusp forms. It is also the Dirichlet series of the corresponding Hecke eigenvalue $t_{j,k}(n)$. Additionally, we defined α_j and $\alpha_{j,k}$ to be appropriately normalised coefficients in the Fourier expansions of the corresponding basis functions.

Now let us state the theorem for the asymptotic of divisor correlations in full strength using notation from spectral theory.

Theorem 2.4.6 (Motohashi [106]). *Let U be a bump function, then for any $r \geq 1$*

$$\sum_{n\geq 1} d(n)d(n+r)U(n/r) = \int_0^{\infty} m(x,r)U(x)dx + E(r;U)\,,$$

where $E(r;U) = e_1(r;U) + e_2(r;U) + e_3(r;U)$ has the following spectral expansion in

terms of the continuous, discrete and residual spectrum:

$$e_1(r;U) := \frac{r^{1/2}}{\pi}\int_{-\infty}^{\infty}\frac{r^{-iy}\sigma_{2iy}(r)|\zeta(1/2+iy)|^4}{|\zeta(1+2iy)|^2}\Theta(y;U)dy\,;$$

$$e_2(r;U) := r^{1/2}\sum_{j=1}^{\infty}\alpha_j t_j(r)H_j^2(1/2)\Theta(\kappa_j;U)\,;$$

$$e_3(r;U) := \frac{1}{4}r^{1/2}\sum_{k=6}^{\infty}\sum_{j=1}^{\vartheta(k)}(-1)^k\alpha_{j,k}t_{j,k}(r)H_{j,k}^2(1/2)\Xi(k-1/2;U)\,,$$

where $\Xi(z;U)$ *and* $\Theta(z;U)$ *are as defined in* (2.49) *and* (2.50) *respectively. In particular, this implies Theorem 2.4.5.*

Let $s\in\mathbb{C}$ with $\operatorname{Re}s>0$ and h be sufficiently small, then summing over r gives

$$\begin{aligned}\sum_{r\geq 1}\frac{m(x,r,h)}{r^{1+s+h}} = \zeta(1+s+h)\zeta(s-h)\frac{\zeta^2(1+h)}{\zeta(2+2h)}x^h(1+x)^h+\\ \zeta(1+s)\zeta(s)\frac{\zeta(1+h)\zeta(1-h)}{\zeta(2)}(x^h+(1+x)^h)+\\ \zeta(1+s-h)\zeta(s+h)\frac{\zeta^2(1-h)}{\zeta(2-2h)}\,.\end{aligned}$$

Setting $i(\alpha_j-\beta_j)=h$ and $i(\alpha_1-\alpha_2)=s=i\delta$ in (2.15) and applying the functional equation (2.30) we recognise the off-diagonal term $H_\alpha(i\delta,t)$ (2.20) in $Q(t,\alpha,\beta)$. In particular, taking the limit as $h\to 0$ yields

$$\frac{d^2}{dz^2}\left(h_0(i\delta,z)e^{z(2\gamma_0+\xi)}\right)\Big|_{z=0} - \frac{1}{4}h_0(i\delta,0)\log^2\left(1+\frac{1}{x}\right)\,,$$

where

$$h_0(z,s)=\frac{\zeta(1+z+s)\zeta(1+z-s)}{\zeta(2+2s)}$$

and

$$\xi=\frac{1}{2}\log x(x+1)\,.$$

This is of course exactly the point of Motohashi's work.

2.5 Explicit expression

In this section we compute

$$\int_{-\infty}^{+\infty}W(t)J_{\alpha,\beta}(nm,t)\big(\frac{m}{n}\big)^{it}\kappa(t)dt. \tag{2.52}$$

We show that Motohashi's formula indeed holds in a wide range of α and β with a power saving. We start with the off-diagonal term $n\neq m$ when $\delta=\beta-\alpha$ is small.

2.5.1 Small shifts

In this subsection we consider $\delta = \beta - \alpha \ll T^{1/2-\varepsilon}$. Let $\tilde{T} = \sqrt{(T+\alpha)(T+\beta)}$, and assume that $Q \ll T$ as well as $\delta \ll Q^{1-\varepsilon}$ and $\Delta Q^{1/2} \ll \tilde{T}$. In fact, we can set $\Delta Q^{1/2} = \tilde{T}$, though this is not necessary. Finally, set

$$W(t) = G_\Delta * \mathbf{I}_{[T,2T]}(t)\,, \tag{2.53}$$

where

$$G_\Delta(t) = (\sqrt{\pi}\Delta)^{-1} e^{-t^2/\Delta^2}\,.$$

Now let us simplify (2.52) in a manner somewhat similar to Lemma 2.4.4. The error term for the fourth moment of the Riemann zeta is at least of size $T^{1/2+\varepsilon}$, so we do not expect to be able to capture any terms of smaller size making this a natural first step.

In this range let us show that we can replace $J_{\alpha,\beta}(mn,t)$ by

$$I(w) := \frac{1}{2\pi i}\int_{(1)} w^{2z} G(z)\frac{dz}{z}\,, \tag{2.54}$$

where $w = (t+\alpha)/2\pi n$.

Lemma 2.5.1. *Let $0 \le \alpha \le \beta \ll T^{2-\varepsilon}$ such that $\delta = \beta - \alpha \ll T^{1/2-\varepsilon}$. Set $\tilde{T} = \sqrt{(T+\alpha)(T+\beta)}$ .Let*

$$U(x) = 2\pi \int W(2\pi x w - \alpha) e^{2\pi i r w} w^{i\delta} I(w) dw\,.$$

Then

$$\sum_{n\neq m} \frac{d(n)d(m)}{n^{1/2+i\alpha} m^{1/2-i\beta}} \int W(t) J_{\alpha,\beta}(nm,t) \big(\frac{m}{n}\big)^{it} \kappa(t) dt =$$

$$2\,\mathrm{Re} \sum_{n,r\ge 1} d(n)d(n+r)U(n) + O\big(\frac{\tilde{T}^2}{\Delta^2} + \delta^{3/2}\big)\,.$$

Proof. Without loss of generality set $\alpha = 0$ and $\delta = \beta$ through an obvious change of variables. Thus we start with considering the integral

$$\int W(t)\big(\frac{m}{n}\big)^{it} \kappa(t)(J(nm,t) - I(t/2\pi n))dt\,. \tag{2.55}$$

Set $a := 2\pi n$ and $\eta_y(a) := y + \phi'(a)$. When $y = \log m/n$ we will omit the index and simply write $\eta(a)$. Let $w = t/2\pi n = t/a$. First, observe that

$$\frac{t(t+\beta)}{4\pi^2 mn} = \frac{t^2}{4\pi^2 n^2}\frac{1+\beta/t}{1+(m-n)/n}$$

and

$$\eta(a) = \log \frac{m}{n} + \log \frac{t}{t+\beta} = \log \frac{1+(m-n)/n}{1+\beta/t}\,.$$

Then

$$J(nm,t) = \frac{1}{2\pi i}\int_{(1)} \left(\frac{t(t+\beta)}{4\pi^2 mn}\right)^z G(z)\frac{dz}{z} = \frac{1}{2\pi i}\int_{(1)} w^{2z} e^{-z\eta(t)} G(z)\frac{dz}{z}\,,$$

and whenever $\eta(t) \ll (Q\log T)^{-1/2}$ we have

$$J(nm,t) - I(t/2\pi n) = \eta(t)\hat{G}(2\log w)(1+E(t))\,,$$

where $E(t)$ is a convergent series in t such that $E(t) \ll Q^{1/2}\eta^2(t) \ll Q^{-1/2}$. The rate of decay of $\hat{G}(2\log t/2\pi n)$ restricts us to

$$|t - 2\pi n| \ll \frac{n\log T}{Q^{1/2}}\,,$$

thus changing variables to $t = u + a$ yields

$$\begin{aligned} F(y) := \mathcal{F}(\kappa(J_{\alpha,\beta} - I))(y) = \int \kappa(t) e^{iyt}(J_{\alpha,\beta}(nm,t) - I((t+\alpha)/a))dt = \\ \kappa(a-\alpha)\eta(a)\int e^{i\eta_y(a)u}\hat{G}(u/a)(1+E(a+u))du \ll a|\eta(a)|G(a\eta_y(a))\,. \end{aligned} \tag{2.56}$$

Then by the Fourier convolution theorem (2.55) is

$$\begin{aligned} (\hat{W} * F)(\log m/n) \ll a|\eta(a)|\int \hat{W}(\log m/n - y)G(a\eta_y(a))dy \ll \\ T^{\varepsilon} a|\eta(a)|\mathbf{I}\{\eta(a) \ll \Delta^{-1}\log T\} \end{aligned} \tag{2.57}$$

since we need support of $\hat{W}(\log m/n - y)$ and $G(a\eta_y(a))$ to overlap. Thus the contribution of $\log m/n \ll (Q\log T)^{-1/2}$ is

$$T^{\varepsilon}\sum_{nm \ll \tilde{T}^2} |\eta(a)|\mathbf{I}\{|\eta(a)| \ll \Delta^{-1}\log T\} \ll \frac{\tilde{T}^2}{\Delta^2}\,. \tag{2.58}$$

On the other hand, if $\log m/n \gg (Q\log^2 T)^{-1/2}$, integrate by parts with

$$\begin{aligned} \int W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)(J_{\alpha,\beta}(nm,t) - I(t/2\pi n))dt = \\ -\int \big(\frac{m}{n}\big)^{it}\kappa(t)\frac{1}{i\eta(t)}(W(J-I)(t))'dt + \\ \int W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)(J-I)(t)\frac{\eta'(t)}{\eta^2(t)}dt\,. \end{aligned} \tag{2.59}$$

Computing Fourier transforms of W' and $(J-I)'$ we again observe that the first term is negligible when $\log m/n \gg (Q\log^2 T)^{-1/2}$. The second term is bounded by $|\delta|/(\tilde{T}^2\log^3 m/n)$ by the oscillation lemma. Its contribution is then

$$\frac{|\delta|}{\tilde{T}^2}\sum_{nm\ll\tilde{T}^2}\frac{1}{(nm)^{1/2}\log^3 m/n}\mathbf{I}\{\log m/n \gg (Q\log^2 T)^{-1/2}\} \ll \frac{\delta Q}{\tilde{T}}\,. \tag{2.60}$$

Thus the total error from (2.58) and (2.60) is

$$\frac{\tilde{T}^2}{\Delta^2}+\frac{\delta Q}{\tilde{T}}\,. \tag{2.61}$$

Now let us similarly replace

$$\int W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)I(t/2\pi n)dt\,,$$

with

$$\int W(t)e^{itr/n}t^{-i\delta}I(t/2\pi n)dt\,, \tag{2.62}$$

where $r=m-n$. Let $k(t)=(t+\beta)\log(1+r/n)-tr/n+\phi(t)+\beta\log t$, and consider

$$\int W(t)e^{itr/n}t^{-i\delta}(e^{ik(t)}-1)I(t/2\pi n)dt\,.$$

We have that

$$k(t)=(t+\beta)\log\frac{1+\frac{r}{n}}{1+\frac{\delta}{t}}-t(\frac{r}{n}-\frac{\delta}{t})$$

is divisible by the derivative of the new phase. When $\log m/n \gg T^{-1/2}$ similarly to (2.59) by partial integration

$$\int W(t)e^{itr/n}t^{-i\beta}(e^{ik(t)}-1)I(t/2\pi n)dt \ll \frac{\delta}{\tilde{T}^2\log^3 m/n}\,, \tag{2.63}$$

where we again used the rate of decay of $\mathcal{F}W'$ and $\mathcal{F}(J-I)'$. When $\log m/n \ll T^{-1/2}$ and thus $\eta(t)\ll T^{-1/2}$ we have

$$k(t)=-\frac{1}{2}\eta(t)\left(t\eta(t)+\beta(\frac{r}{n}+\frac{\beta}{t})+O\left(\frac{(T+\beta)r^2}{n^2}+\frac{(T+\beta)\beta^2}{T^2}\right)\right)\ll 1\,.$$

Integrating by parts

$$\int e^{itr/n}t^{i\delta}\left(W(t)I(t/2\pi n)\frac{e^{ik(t)}-1}{\eta(t)}\right)'dt=\int e^{itr/n}t^{i\delta}\left(W(t)I(t/2\pi n)\right)'\frac{e^{ik(t)}-1}{\eta(t)}dt+$$
$$\int e^{itr/n}t^{i\delta}W(t)I(t/2\pi n)\left(\frac{e^{ik(t)}-1}{\eta(t)}\right)'dt\,.$$

The first integral similarly to (2.57) is

$$\int e^{itr/n}t^{-i\delta}\left(W(t)I(t/2\pi n)\right)'\frac{e^{ik(t)}-1}{\eta(t)}dt \ll I(\frac{T}{2\pi n})(T|\eta(T)|+\frac{\beta r}{n}+\frac{\delta^2}{T})\hat{W}(\eta(T))+ \\ W(a)G(a\eta(a))(a|\eta(a)|+|\delta|r/n+\delta^2/n)$$

and using oscillation lemma

$$\int e^{itr/n}t^{i\delta}W(t)I(t/2\pi n)\left(\frac{e^{ik(t)}-1}{\eta(t)}\right)'dt \ll T^{\varepsilon}+\frac{\delta}{\tilde{T}}\min\{\frac{\tilde{T}}{(1+\delta)^{1/2}},\frac{\delta^{1/2}}{\log m/n}\}\,.$$

Summing over m,n the above gives total error

$$\frac{\delta}{\tilde{T}}\sum_{nm\ll\tilde{T}^2}\frac{1}{(nm)^{1/2}}\min\{\frac{\tilde{T}}{(1+\delta)^{1/2}},\frac{\delta^{1/2}}{\log m/n}\}\mathbf{I}\{\log m/n\ll T^{-1/2}\}\ll T^{\varepsilon}+|\delta|^{3/2}$$

and

$$\sum_{nm\ll\tilde{T}^2}\frac{1}{(nm)^{1/2}}(T|\eta(T)|+\frac{\delta^2}{\tilde{T}})\mathbf{I}\{\eta(T)\ll\Delta^{-1}\log T\}\ll\frac{\tilde{T}^2}{\Delta^2}+\frac{\delta^2}{\Delta}\,;$$

$$\sum_{nm\ll\tilde{T}^2}\frac{1}{(nm)^{1/2}}(n|\eta(a)|+\frac{\delta^2}{n})\mathbf{I}\{\eta(a)\ll Q^{1/2}/n\wedge T^{-1/2}\}\ll\frac{\delta^2}{\tilde{T}^{1/2}}\,.$$

This together with (2.61) implies the statement of the lemma. □

Now let us evaluate the integral

$$2\pi\sum_{n,r\geq 1}d(n)d(n+r)\int W(2\pi nw-\alpha)e^{2\pi irw}w^{i\delta}I(w)dw$$

In order to use the asymptotic of divisor sum correlations, we need $|m-n|$ to be quite small. This is essentially the purpose of the parameter Δ for W.

There are two ways to approach this integral. We can choose W and G in such a way that the internal integral gives a smooth weight function of compact support and apply Motohashi's result in its full strength. Alternatively, we can use this result as a black box, but with a cleaner computation. For small δ we do the latter.

Let

$$\sum_{n\leq X}d(n)d(n+r)=S(X,r)+E(X,r)\,,$$

where $S(X,r)=X\sum_{j=0}^{2}b_j(r)\log^j X$ is the main term, and $E(X,r)=o(X)$ is the error term. Then we have the following.

Lemma 2.5.2. *Let $0 \le \alpha \le \beta \ll T^{2-\varepsilon}$ such that $\delta = |\alpha - \beta| \ll T^{1/2-\varepsilon}$. Set $\tilde{T} = \sqrt{(T+\alpha)(T+\beta)}$, and assume $\Delta^2\delta \ll \tilde{T}^{2-\varepsilon}$. Let*

$$U(x) = 2\pi \int_{-\infty}^{+\infty} W(2\pi xw - \alpha) e^{2\pi irw} w^{i\delta} I(w) dw \,.$$

Then

$$\sum_{n,r\ge 1} d(n)d(n+r)U(n) = -\sum_{r\ge 1} \int_0^\infty S(x,r)U'(x)dx + O\left(\Delta + \frac{\delta\Delta^2}{\tilde{T}} + \frac{\tilde{T}^{5/3}}{\Delta}\right),$$

where $\tilde{T}^{5/3}/\Delta$ can be replaced with $\tilde{T}^{4/3}\Delta^{2/3}$ if $\Delta\delta \ll \tilde{T}^{1-\varepsilon}$.

Proof. Again, without loss of generality set $\alpha = 0$. Let $U(x)$ be as defined above, then $U(x)$ is smooth and for any $k \in \mathbb{N}$

$$U^{(k)}(x) = 2\pi \int_{-\infty}^{+\infty} (2\pi w)^k W^{(k)}(2\pi xw) e^{2\pi irw} w^{i\delta} I(w) dw \,.$$

Function $I(w)$ is similar to the indicator function on the interval $[1, +\infty)$, and $W^{(k)}(2\pi xw)$ is (approximately) supported on $[T_1/2\pi x, T_2/2\pi x]$. When $x \gg T^{1+\varepsilon}$, these intervals do not overlap. In particular, for $0 < w < 1$

$$I(w) = \frac{1}{2\pi i}\int_{(1)} w^{2z} G(z) \frac{dz}{z} \ll e^{Q^{1/2}\log w},$$

thus when $x \gg T^{1+\varepsilon}$ for any $k \in \mathbb{Z}^{\ge 0}$

$$U^{(k)}(x) \ll \left(\frac{T}{x}\right)^{\sqrt{Q}}.$$

Using this fact and applying partial summation rewrite

$$\sum_{n\ge 1} d(n)d(n+r)U(n) = -\int_1^\infty (S(x,r) + E(x,r))U'(x)dx \,.$$

Set $\eta(x) = r/x + \delta/T$. For $W(t) = (G_\Delta * \mathbf{I}_{[T_1,T_2]})(t)$ we have

$$W'(t) = \frac{1}{\Delta\pi^{1/2}} e^{-\frac{(t-T_2)^2}{\Delta^2}} - \frac{1}{\Delta\pi^{1/2}} e^{-\frac{(t-T_1)^2}{\Delta^2}}.$$

Substituting this into the integral and changing variables to $y = 2\pi xw - T_j$ each term gives

$$\frac{4\pi^{3/2}}{\Delta}\int e^{-\frac{(2\pi xw - T_j)^2}{\Delta^2}} e^{2\pi irw} w^{1+i\delta} I(w)dw =$$
$$\frac{e^{\frac{iT_j r}{x} + i\phi(T_j)}}{\sqrt{\pi}\Delta x^{2+i\delta}} \int e^{-\frac{y^2}{\Delta^2}} e^{iy\eta(x)} (T_j + y) I((T_j+y)/2\pi x)(1 + E(y))dy \,,$$

where $E(y) \ll y^2\delta/T_j^2$ is a convergent series. Expanding $I(w)$ into a Taylor series near $T_j/2\pi x$ and computing the Fourier transform above we see that the above equals

$$T_j \frac{e^{\frac{iT_j r}{x}+i\phi(T_j)}}{(2\pi)^{i\delta}x^{2+i\delta}} e^{-\frac{\Delta^2}{4}\eta^2(x)} I(T_j/2\pi x)(1+E_1(x))$$

where $E_1(x)$ is again a convergent series in x with

$$E_1(x) \ll \frac{\Delta^2\delta}{\tilde{T}^2} + \frac{\Delta}{xT} \max_{|t-T_j|\le \Delta\log T} |\hat{G}(2\log t/2\pi x)|\,.$$

This implies that

$$\int_1^\infty E(x,r)U'(x) \ll \tilde{T}\int_1^\infty \frac{|E(x,r)|}{x^2} e^{-\frac{\Delta^2}{4}\eta^2(x)} I(T_j/2\pi x)dx \ll$$
$$\tilde{T}\int_1^{3\tilde{T}} \frac{|E(x,r)|}{x^2}\mathbf{I}\{|\eta(x)|\le \log T/\Delta\}dx\,,$$

which also restricts r to the range $r \le \delta + \tilde{T}/\Delta$. If $\delta \ll \tilde{T}/\Delta$, we have

$$\sum_{r\ge 1}\int_{r\Delta}^{3\tilde{T}} \frac{|E(x,r)|}{x^2}dx \ll \int_\Delta^{3\tilde{T}} \sum_{r\le x/\Delta} |E(x,r)|\frac{dx}{x^2} \ll \int_\Delta^{3\tilde{T}} \left(\frac{x^2}{\Delta}\right)^{2/3}\frac{dx}{x^2} \ll \left(\frac{\tilde{T}}{\Delta^2}\right)^{1/3+\varepsilon}$$

using (2.48). If $\tilde{T}/\Delta \ll \delta \ll (\tilde{T}/\Delta)^2$ and $\delta \ll T^{1/2-\varepsilon}$

$$\sum_{r\ge 1}\int_1^{3\tilde{T}} \frac{|E(x,r)|}{x^2}\mathbf{I}\{|\eta(x)|\le \log T/\Delta\}dx \ll \int_{T/\delta}^{3\tilde{T}} \sum_{r=\frac{\delta x}{T}+O(\frac{x}{\Delta})} |E(x,r)|\frac{dx}{x^2} \ll$$
$$\int_{T/\delta}^{3\tilde{T}} \frac{x^{5/3}}{\Delta}\frac{dx}{x^2} \ll \frac{\tilde{T}^{2/3}}{\Delta}\,.$$

This is acceptable when $\Delta \gg \tilde{T}^{2/3}$. Finally, integrating by parts and using the oscillation lemma

$$\sum_{r\ge 1}\int_0^\infty S(x,r)e^{\frac{iT_j r}{x}} e^{-\frac{\Delta^2}{4}\eta^2(x)} I(T_j/2\pi x)E_1(x)\frac{dx}{x^{2+i\delta}} \ll \frac{\delta\Delta^2}{\tilde{T}} + \Delta\,.$$

Combining these errors yields the statement of the lemma. □

Now let us compute

$$-\sum_{r\ge 1}\int_0^\infty S(x,r)U'(x)dx\,.$$

In fact, this is somewhat unnecessary since the essence of the problem is in bounding the error term, and we already know the Dirichlet series of $S(x,r)$. But let us finish the computation nonetheless.

We can remove I and W at the cost of an acceptable error and essentially follow [38]. However, let us do this using contour integration and Mellin transforms instead.

Theorem 2.5.3 (Off-diagonal term)**.** *Let* $0 \le \alpha \le \beta \ll T^{3/2-\varepsilon}$ *such that* $\delta = \beta - \alpha \ll T^{1/2-\varepsilon}$*. Let* $\tilde{T} := \sqrt{(T+\alpha)(T+\beta)} \ll T^{5/4-\varepsilon}$*. Let*

$$h(z,s) = \frac{1}{\zeta(2+2s)} \left(\zeta(1+z+s)\zeta(1-z+s) - \frac{2s\zeta(1+2z)}{s^2-z^2} \right).$$

Then

$$2\,\mathrm{Re} \sum_{n \neq m} \frac{d(n)d(m)}{n^{1/2+i\alpha} m^{1/2-i\beta}} \int W(t) \big(\frac{m}{n}\big)^{it} \kappa(t) V_{\alpha,\beta}(nm,t) dt = \mathrm{Re}(2TF(2T) - TF(T)) + E(T)\,,$$

where

$$E(T) \ll \begin{cases} \tilde{T}^{4/5+\varepsilon}, & \text{if } \delta \ll \tilde{T}^{1/3}\,; \\ \tilde{T}^{5/6}\delta^{1/4}, & \text{if } \tilde{T}^{1/3} \ll \delta \ll T^{1/2-\varepsilon}\,, \end{cases}$$

and

$$F(X) = \frac{d^2}{dz^2} \frac{h(i\delta,s)}{z+1} \left(\frac{(X+\alpha)(X+\beta)}{4\pi^2} e^{2\gamma_0} \right)^s \Big|_{s=0} - \frac{1}{4} h(i\delta,0) \log^2 \frac{X+\alpha}{X+\beta}\,.$$

We also recognise $h(z,s)$ *as defined in* (2.19) *and* (2.23)*.*

Proof. Without loss of generality set $\alpha = 0$. Let

$$U'(x) = (2\pi)^2 \int_0^\infty W'(2\pi x w) e^{2\pi i r w} w^{1+i\delta} I(w) dw$$

as before. Let us show that $F(T)$ is given by

$$T \sum_{r=1}^{\infty} \int_1^\infty S(x,r) e^{\frac{iTr}{x} + i\phi(T/x)} e^{-\frac{\Delta^2}{4}\eta^2(x)} I(T/2\pi x) \frac{dx}{x^2}\,, \tag{2.64}$$

or equivalently after changing variables to $u = Tr/2\pi x$

$$\sum_{r=1}^{\infty} \frac{1}{r^{1-i\delta}} \int_0^\infty e^{2\pi i u} S(Tr/2\pi u, r) I(u/r) u^{-i\delta} e^{-\frac{\pi^2\Delta^2}{T^2}(u-\delta/2\pi)^2} du\,.$$

Let us rewrite this expression in a more convenient manner. First, let

$$Y(u,s) := \frac{1}{2\pi i} \int_{(2)} \zeta(1+s+z-i\delta)\zeta(z-s-i\delta) u^{z-i\delta-s-1} \frac{G(z)dz}{z}\,.$$

For further computations assume that $0 < s < 1$ is a real number. This does not affect the value of the integral and is only relevant when shifting the line of integration and picking up poles.

Second, using the expansion 2.47 we have

$$S(x,r) = x\sum_{k=1}^{\infty}\frac{\mu(k)}{k^2}\sum_{d|r}\frac{1}{d}\left(\left(\log\frac{xe^{2\gamma_0-1}}{d^2k^2}\right)^2+1\right) = x\left(\sum_{k=1}^{\infty}\frac{\mu(k)}{k^2}\sum_{d|r}\frac{1}{d}\right)\frac{d^2}{ds^2}\frac{B^s}{s+1}\Big|_{s=0},$$

where $B = xe^{2\gamma}/(dk)^2$.

Using the above and changing the order of summation and integration due to absolute convergence rewrite the integral in question as

$$\sum_{r=1}^{\infty}\frac{1}{r^{1-i\delta}}\int_0^{\infty}e^{2\pi iu}S(Tr/2\pi u,r)I(u/r)u^{-i\delta}e^{-\frac{\pi^2\Delta^2}{T^2}(u-\delta/2\pi)^2}du = T\frac{d^2}{ds^2}\frac{A^s\zeta^{-1}(2+2s)}{s+1}\int_0^{\infty}e^{2\pi iu}e^{-\frac{\pi^2\Delta^2}{T^2}(u-\delta/2\pi)^2}Y(u,s)du\Big|_{s=0},$$

where $A = Te^{2\gamma_0}/2\pi$. Moving the line of integration to $(s+\varepsilon)$ we pick up the pole at $z = 1+s+i\delta$ with residue

$$\zeta(2+2s)\frac{G(1+s+i\delta)}{1+s+i\delta}\int_0^{\infty}e^{2\pi iu}e^{-\frac{\pi^2\Delta^2}{T^2}(u-\delta/2\pi)^2}du\,,$$

which gives error

$$T\frac{d^2}{ds^2}\frac{A^s}{s+1}\frac{G(1+s+i\delta)}{1+s+i\delta}\int_0^{\infty}e^{2\pi iu}e^{-\frac{\pi^2\Delta^2}{T^2}(u-\delta/2\pi)^2}du\Big|_{s=0}$$

,.

If $\delta \gg Q^{1/2}$, then $G^{(l)}(1+i\delta) \ll |\delta/\sqrt{Q}|^{-b}$, otherwise when $\delta \ll Q^{1/2} \ll T/\Delta$

$$\int_0^{\infty}e^{2\pi iu}e^{-\frac{\pi^2\Delta^2}{T^2}(u-\delta/2\pi)^2}du \ll \frac{T}{\Delta}e^{-T^2/\Delta^2},$$

so this is negligible in both cases.

Note that we initially set $r > 0$. Combining with $r < 0$ replaces $e^{2\pi iu}$ with $\cos(2\pi u)$. Changing the order of integration and essentially computing the Mellin transform of $\cos(2\pi u)$

$$\int_0^{\infty}\cos(2\pi u)u^{w-1}e^{-\frac{\pi^2\Delta^2}{T^2}(u-\delta/2\pi)^2}du = \frac{1}{(2\pi)^w}\cos(\pi w/2)\Gamma(w)+E(w)\,,$$

where $E(w) \ll (\Delta/T)^{1+\varepsilon}$, together with

$$\frac{1}{(2\pi)^w}\zeta(w)\cos(\pi w/2)\Gamma(w) = \zeta(1-w)\,,$$

gives

$$\frac{1}{2\pi i}\int_{(s+\varepsilon)}\zeta(1+s+z-i\delta)\zeta(1+s-z+i\delta)\frac{G(z)}{z}dz\,.$$

Moving the line of integration to (0) with a small half-circle around 0 gives 1/2 of the residue at $z=0$, the residue at $z=s-i\delta$ and

$$\operatorname{Re}\frac{1}{2\pi i}\int_b^\infty \zeta(1+s+iy+i\delta)\zeta(1+s-iy-i\delta)\frac{G(iy)}{y}dy=0\,. \tag{2.65}$$

Finally, the residues give

$$\frac{1}{2}\zeta(1+s-i\delta)\zeta(1+s+i\delta)-\frac{\zeta(1+2s)G(s+i\delta)}{s+i\delta}\,. \tag{2.66}$$

If $\delta\gg Q^{1/2}$, the second term is negligible. Otherwise

$$\operatorname{Re}\frac{G(s-i\delta)}{s-i\delta}=\frac{2s}{s^2+\delta^2}+R(s)\,,$$

where $R(s)$ is an absolutely convergent series with largest terms

$$\frac{s}{Q}+\frac{s\delta^2}{Q^2}\,.$$

Hence we can replace h_G with h at the cost of an error of size $O(T/Q)$.

Putting together errors from Lemmas 2.4.3, 2.5.1 and 2.5.2 we have error

$$\frac{Q^{1/2}(\delta+Q^{1/2})}{\Delta}+\frac{\tilde{T}^2}{\Delta^2}+\Delta+|\delta|^{3/2}+\frac{\delta\Delta^2}{\tilde{T}}+\frac{\tilde{T}^{4/3}}{\Delta^{2/3}}+\frac{\tilde{T}^{5/3}}{\Delta}\mathbf{I}\{\delta\Delta\gg\tilde{T}\}\,.$$

If $\delta\ll T^{1/3-\varepsilon}$ and $\tilde{T}\ll T^{5/4}$, set $Q=(\tilde{T}/\Delta)^2$ and $\Delta=\tilde{T}^{4/5}$, then this is $O(\tilde{T}^{4/5})$. If $T^{1/3}\ll\delta\ll T^{1/2}$, set $Q=(\tilde{T}/\Delta)^2$ and $\Delta=\tilde{T}^{5/6}/\delta^{1/4}$, which yields total error $O(\tilde{T}^{5/6}\delta^{1/4})$. □

Now let us move on to the complementary case.

2.5.2 Large shifts

Let $\tilde{T}=\sqrt{(T+\alpha)(T+\beta)}$. Let $\delta=\beta-\alpha$ be such that $\delta\gg\tilde{T}^{1/3}$ and $\delta\gg(\tilde{T}/\Delta)^{2+\varepsilon}$. Set $Q=T^\varepsilon$, and

$$W(t)=\begin{cases}p(|t-T_j|/\Delta), & |t-T_j|\le\Delta\,;\\ 1, & T_1+\Delta\le t\le T_2-\Delta\,;\\ 0, & \text{otherwise}\,,\end{cases} \tag{2.67}$$

where $T_j = jT$, and

$$p(x) = \exp\left(-\frac{1}{x}\exp\left(-\frac{1}{1-x}\right)\right).$$

Then

$$\int W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)J_{\alpha,\beta}(nm,t)dt \tag{2.68}$$

can be evaluated using stationary phase as in Lemma 2.2.3 and 2.2.4.

Lemma 2.5.4. *Let $m(x,r)$ be as in* (2.51)*, and let*

$$W_1(x) := e^{-i\delta\log\delta/2\pi e}I(\delta/2\pi r)\delta^{1/2}W(\delta x-\alpha)\,.$$

Then

$$\sum_{n,m\geq 1}\frac{d(n)d(m)}{n^{1/2+i\alpha}m^{1/2-i\beta}}\int_{-\infty}^{+\infty}W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)J_{\alpha,\beta}(nm,t)dt = \sum_{r=1}^{\infty}\frac{1}{r^{1+i\delta}}\int_0^{\infty}m(x,r)W_1(\delta x-\alpha)dx\ + O\left(\frac{\tilde{T}^2}{\Delta\sqrt{\delta}}\right).$$

Proof. Let $f(t) = t\log m/n + \phi(t)$, and let s be its stationary point, that is

$$f'(s) = \log m/n + \phi'(s) = \log m/n + \log\frac{s+\alpha}{s+\beta} = 0\,.$$

Then $s = n\delta/r - \alpha$, where $r = m-n$. We also $f''(s) = r^2/(mn\delta)$, and

$$J_{\alpha,\beta}(nm,s) = I(\delta/2\pi r)$$

as in (2.54) without having to do the reduction as in Lemma 2.5.1. Then by Lemma 2.2.3

$$\int_{-\infty}^{+\infty}W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)J_{\alpha,\beta}(nm,t)dt = e^{if(s)+\pi i/4}\frac{(\delta mn)^{1/2}}{r}W(s)I(\delta/2\pi r) + R_1(m,n)\,, \tag{2.69}$$

where $R_1(m,n)$ is given by the lower order terms in the asymptotic expansion. This term satisfies

$$R_1(m,n) \ll |f''(s)|^{-3/2}(WJ)''(s)$$

and gives total error

$$\delta^{3/2}T^{\varepsilon}\sum_{n,r\geq 1}\frac{n(n+r)}{r^3}(WJ)''(n|\delta|/r-\alpha) \ll \frac{(T+\alpha)(T+\beta)}{\Delta\sqrt{\delta}} = \frac{\tilde{T}^2}{\Delta\sqrt{\delta}}\,.$$

Observing that

$$f(s) - \alpha\log n + \beta\log m = -\delta\log\delta/2\pi r + \delta$$

and setting

$$W_1(x) = e^{i\delta \log 2\pi e/\delta} I(\delta/2\pi r)\delta^{1/2} W(\delta x - \alpha)$$

we then reduce our problem to evaluating

$$\sum_{r=1}^{\infty} \frac{1}{r^{1-i\delta}} \sum_{n,r\geq 1} d(n)d(n+r)W_1(n/r)\,, \tag{2.70}$$

which we can do using Motohashi's result 2.4.6. □

Recall that for a function $U(x)$

$$\Xi(z;U) = \frac{\Gamma^2(1/2+z)}{\Gamma(1+2z)} \int_0^\infty U(x) x^{-1/2-z} F(1/2+z, 1/2+z, 1+2z; -1/x) dx\,,$$

where $F(a,b,c;x)$ is the hypergeometric function, and

$$\Theta(y;U) = \frac{1}{2}\operatorname{Re}\left(\left(1 + \frac{i}{\sinh \pi y}\right)\Xi(iy;U)\right)\,.$$

Note that $\Xi(y;W)$ and $\Theta(y;W)$ do not depend on r at all, so we can bound it separately from the sum over r.

Lemma 2.5.5. *Let $W_2(x) = \delta^{1/2}W(\delta x - \alpha)$. Then for $y \in \mathbb{R}$ and any $l \in \mathbb{N}$*

$$\Xi(iy;W_2) \ll_l \frac{\tilde{T}^{l+1/2}}{(1+|y|)^{l+3/2}\Delta^l},\ \Theta(y;W_2) \ll \frac{\tilde{T}^{1/2}(1+1/y)}{(1+|y|)^{3/2}}\,.$$

In addition, for $k \in \mathbb{N}$

$$\Xi(k-1/2;W_2) \ll \delta^{1/2}\begin{cases} \frac{(k-1)!^2}{(2k-1)!}\left(\frac{\delta}{T+\alpha}\right)^{k-1}, & \text{if } \delta \leq T+\alpha\,; \\ \min\{\frac{1}{k^2}, \frac{1}{k}e^{-Ck\sqrt{(T+\alpha)/\delta}}\}, & \text{otherwise}\,. \end{cases}$$

Proof. By definition,

$$\Xi(iy;W_2) = \int_0^\infty W_2(x) \int_0^1 (w(1-w))^{-1/2+iy}(x+w)^{-1/2-iy} dw dx\,,$$

and integrating by parts $l+1$ times with respect to x gives

$$\Xi(iy;W_2) \ll_l (1+|y|)^{-l-1}\left|\int_0^\infty W_2^{(l+1)}(x)dx \int_0^1 (w(1-w))^{-1/2+iy}(x+w)^{l+1/2-iy} dw\right|.$$

We can bound the internal integral using the oscillation lemma for the second derivative, which is slightly different depending on whether $|x| \ll 1$ or $|x| \gg 1$. Noting that the stationary point is at $y\log(w/(1-w) - y\log(x+w)$ is at $\sqrt{x^2+x} - x$ gives

$$\int_0^1 (w(1-w))^{-1/2+iy}(x+w)^{l+1/2-iy} dw \ll (1+|y|)^{-1/2}\min\{|x|, |x|^{1/2}\}^{l+1/2}\,.$$

Thus

$$\Xi(iy;W_2) \ll_l (1+|y|)^{-l-3/2}\int_0^\infty |W_2^{(l+1)}(x)|\min\{|x|,|x|^{1/2}\}^{l+1/2}dx \ll_l$$
$$\frac{\delta^{l+1/2}}{(1+|y|)^{l+3/2}\Delta^l}\min\{(T+\alpha)/\delta,\sqrt{(T+\alpha)/\delta}\}^{l+1/2} \ll_l \frac{\tilde{T}^{l+1/2}}{(1+|y|)^{l+3/2}\Delta^l}\,.$$

In particular, this implies that we can restrict to $y \ll X$, where

$$X = X(T,\alpha,\beta,\Delta) := \tilde{T}/\Delta\,,$$

and take $l=0$. A similar bound to above applies to $\Theta(y;W_2)$ with

$$\Theta(y;W_2) = \frac{1}{2}\operatorname{Re}\left(\left(1+\frac{i}{\sinh\pi y}\right)\Xi(iy;W_2)\right) \ll \frac{\tilde{T}^{1/2}(1+1/y)}{(1+|y|)^{3/2}}\,. \tag{2.71}$$

When $z = k-1/2$ and $k\in\mathbb{N}$, we have

$$\Xi(k-1/2;W_2) \ll \delta^{1/2}\begin{cases}\frac{(k-1)!^2}{(2k-1)!}\left(\frac{\delta}{T+\alpha}\right)^{k-1}, & \text{if } \delta \le T+\alpha\,;\\ \min\{\frac{1}{k^2},\frac{1}{k}e^{-Ck\sqrt{(T+\alpha)/\delta}}\}, & \text{otherwise}\,,\end{cases} \tag{2.72}$$

where $C>0$ is fixed. In the first case, this follows directly from the series representation of the hypergeometric function. The second case follows after integrating by parts as above and observing that for $w\in(0,1)$ and $x \asymp (T+\alpha)/\delta < 1$

$$\frac{w(1-w)}{x+w} = \frac{1-w}{1+x/w} = \exp\left(\log\frac{1-w}{1+x/w}\right) \le \begin{cases}e^{-C_1(w+x/w)}, & \text{if } w>x\,;\\ e^{-C_2(w+w/x+\log w/x)}, & \text{if } w\le x\,.\end{cases}$$

In both cases, this is $O(e^{-C_3\sqrt{x}})$ with some fixed constants $C_j>0$. □

Lemma 2.5.6. *Let $e(W_1,r) = e_1(W_1,r)+e_2(W_1,r)+e_3(W_1,r)$, where*

$$e_1 = \frac{r^{1/2}}{\pi}\int_{-\infty}^{\infty}\frac{r^{-iy}\sigma_{2iy}(r)|\zeta(1/2+iy)|^4}{|\zeta(1+2iy)|^2}\Theta(y;W_1)dy\,;$$
$$e_2 = r^{1/2}\sum_{j=1}^{\infty}\alpha_j t_j(r)H_j^2(1/2)\Theta(\kappa_j;W_1)\,;$$
$$e_3 = \frac{1}{4}r^{1/2}\sum_{k=6}^{\infty}\sum_{j=1}^{\vartheta(k)}(-1)^k\alpha_{j,k}t_{j,k}(r)H_{j,k}^2(1/2)\Xi(k-1/2;W_1)\,.$$

Then

$$\sum_{r\ge1}\frac{e(W_1,r)}{r^{1+i\delta}} \ll \frac{\tilde{T}^{1+\varepsilon}}{\Delta^{1/2}} + \tilde{T}^{1/2+\varepsilon}\delta^{1/3} + \delta^{5/6+\varepsilon}.$$

Proof. Let $E_j = \sum_{r\geq 1} e_j(W_1, r)/r^{1+i\delta}$ for $j = 1, 2, 3$. Recall that

$$W_1(x) = e^{i\delta \log 2\pi e/\delta} I(\delta/2\pi r)\delta^{1/2} W(\delta x - \alpha) = e^{i\delta \log 2\pi e/\delta} I(\delta/2\pi r) W_2(x)\,,$$

so the error terms factorise conveniently. Let

$$\begin{aligned}
Z_{iy}(1/2 + i\delta) &= e^{i\delta \log 2\pi e/\delta} \sum_{r=1}^{\infty} \frac{\sigma_{2iy}(r) I(\delta/2\pi r)}{r^{1/2+i\delta+iy}}\,;\\
L_j(1/2 + i\delta) &:= e^{i\delta \log 2\pi e/\delta} \sum_{r=1}^{\infty} \frac{t_j(r) I(\delta/2\pi r)}{r^{1/2+i\delta}}\,;\\
L_{j,k}(1/2 + i\delta) &:= e^{i\delta \log 2\pi e/\delta} \sum_{r=1}^{\infty} \frac{t_{j,k}(r) I(\delta/2\pi r)}{r^{1/2+i\delta}}\,.
\end{aligned}$$

As $I(\delta/2\pi r)$ is just a weight function essentially truncating the sum at $r \leq \delta/2\pi$, the functions above are approximations of $\zeta(1/2 + i\delta + iy)\zeta(1/2 + i\delta - iy)$, $H_j(1/2 + i\delta)$ and $H_{j,k}(1/2 + i\delta)$ respectively. Changing the order of summation and integration due to absolute convergence, we then have

$$\begin{aligned}
E_1 &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{Z_{iy}(1/2 + i\delta)|\zeta(1/2 + iy)|^4}{|\zeta(1 + 2iy)|^2} \Theta(y; W_2) dy\,;\\
E_2 &= \sum_{j=1}^{\infty} \alpha_j L_j(1/2 + i\delta) H_j^2(1/2) \Theta(\kappa_j; W_2)\,;\\
E_3 &= \frac{1}{4} \sum_{k=6}^{\infty} \sum_{j=1}^{\vartheta(k)} (-1)^k \alpha_{j,k} L_{j,k}(1/2 + i\delta) t_{j,k}(r) H_{j,k}^2(1/2) \Xi(k - 1/2; W_2)\,.
\end{aligned}$$

Let us first bound E_1. Applying (2.71),

$$E_1 \ll \tilde{T}^{1/2} \int_{y \ll X^{1+\varepsilon}} \frac{|Z_{iy}(1/2 + i\delta)||\zeta(1/2 + iy)|^4(1 + 1/|y|)}{|\zeta(1 + 2iy)|^2(1 + |y|)^{3/2}} dy\,,$$

so it remains to bound $Z_{iy}(1/2 + i\delta)$. Recall the subconvexity bound

$$\zeta(1/2 + it) \ll (1 + |t|)^{13/84+\varepsilon} \ll (1 + |t|)^{1/6}\,. \tag{2.73}$$

The approximate functional equation for the Riemann zeta implies that

$$|\zeta(1/2 + it)| \leq 2\Big| \sum_{n \leq \sqrt{t/2\pi}} n^{-1/2-it} \Big| + O(1)\,,$$

and in fact, going back to the proof [22], the bound (2.73) is a consequence of a corresponding bound for exponential sums $\sum_{n\leq N} e^{-it\log n}$. Similarly in our case, $Z_{iy}(1/2 + i\delta)$ has the same upper bound. Let us give a quick proof of this fact.

Recall the definition of $I(x)$, then

$$\sum_{r=1}^{\infty}\frac{\sigma_{2iy}(r)I(\delta/2\pi r)}{r^{1/2+i\delta+iy}}=\frac{1}{2\pi i}\int_{(1)}\zeta(1/2+i\delta+iy+s)\zeta(1/2+i\delta-iy+s)\left(\frac{\delta}{2\pi}\right)^{s}G(s)\frac{ds}{s}\,.$$

Moving the line of integration to (0) with a small semi-circle around (0) we pick up poles at $s=1/2-i\delta\pm iy$ with residue

$$\zeta(1\pm 2iy)\left(\frac{\delta}{2\pi}\right)^{1/2-i\delta\pm iy}\frac{G(1/2-i\delta\pm iy)}{1/2-i\delta\pm iy}\ll\frac{\delta^{1/2}}{1+|\delta\pm y|}|\zeta(1\pm 2iy)|G(i|\delta\pm y|)$$

as well as half the pole at $s=0$ with residue $\zeta(1/2+i\delta+iy)\zeta(1/2+i\delta-iy)$. Finally,

$$\int_{|w|\geq 1}\zeta(1/2+i\delta+iy+iw)\zeta(1/2+i\delta-iy+iw)\left(\frac{\delta}{2\pi}\right)^{iw}G(iw)\frac{dw}{w}\ll$$
$$\int_{1\ll|w|\ll Q^{1/2+\varepsilon}}|\zeta(1/2+i\delta+iy+iw)\zeta(1/2+i\delta-iy+iw)|\frac{dw}{w}\,.$$

Thus if $|\delta\pm y|\gg Q^{1/2+\varepsilon}\gg T^{\varepsilon}$, the bound (2.73) applies to $Z_{iy}(1/2+i\delta)$. Otherwise we have the trivial bound $Z_{iy}(1/2+i\delta)\ll\delta^{1/2+\varepsilon}$.

Using this bound together with $\zeta(1+it)\gg 1+1/|t|$ and $\int_0^N|\zeta(1/2+it)|^4dt\ll N^{1+\varepsilon}$,

$$\int_{|\delta\pm y|\gg Q^{1/2+\varepsilon}}\frac{|Z_{iy}(1/2+i\delta)||\zeta(1/2+iy)|^4(1+1/|y|)}{|\zeta(1+2iy)|^2(1+|y|)^{3/2}}dy\ll$$
$$\int_{|\delta\pm y|\gg Q^{1/2+\varepsilon}}\frac{|\delta^2-y^2|^{1/6}|\zeta(1/2+iy)|^4}{(1+|y|)^{3/2}}dy\ll\delta^{1/3+\varepsilon}$$

and

$$\int_{|\delta\pm y|\ll Q^{1/2+\varepsilon}}\frac{|Z_{iy}(1/2+i\delta)||\zeta(1/2+iy)|^4(1+1/|y|)}{|\zeta(1+2iy)|^2(1+|y|)^{3/2}}dy\ll$$
$$\frac{1}{\delta}\int_{|\delta\pm y|\ll Q^{1/2+\varepsilon}}\frac{|\zeta(1/2+iy)|^4}{|\zeta(1+2iy)|}dy\ll T^{\varepsilon}\,.$$

Thus

$$E_1\ll\tilde{T}^{1/2}\int\frac{|Z_{iy}(1/2+i\delta)||\zeta(1/2+iy)|^4(1+1/|y|)}{|\zeta(1+2iy)|^2(1+|y|)^{3/2}}dy\ll\tilde{T}^{1/2+\varepsilon}\delta^{1/3}\,.$$

To bound E_2 and E_3 we will use bounds from Lemma 2.2.10. Same bounds apply if one replaces S_j and $S_{j,k}$ with L_j and $L_{j,k}$ respectively via a standard partial summation argument, thus we also have

$$\begin{aligned}&\sum_{K\leq\kappa_j\leq 2K}\alpha_j|L_j(1/2+it)|^2\ll K^{2+\varepsilon}+K^{\varepsilon}t^{2/3+\varepsilon}\,,\\&\sum_{K\leq k\leq 2K}\sum_{j=1}^{\vartheta(k)}\alpha_{j,k}|L_{j,k}(1/2+it)|^2\ll K^{2+\varepsilon}+K^{\varepsilon}t^{2/3+\varepsilon}\,.\end{aligned}\tag{2.74}$$

Plugging (2.71) into E_2, and using Cauchy's inequality,

$$E_2 \ll \tilde{T}^{1/2} \sum_{\kappa_j \ll X^{1+\varepsilon}} \frac{\alpha_j |L_j(1/2+i\delta)| H_j^2(1/2)}{\kappa_j^{3/2}} \ll$$
$$\tilde{T}^{1/2} \sqrt{\sum_{\kappa_j \ll X^{1+\varepsilon}} \frac{\alpha_j |L_j(1/2+i\delta)|^2}{\kappa_j} \sum_{\kappa_j \ll X^{1+\varepsilon}} \frac{\alpha_j H_j^4(1/2)}{\kappa_j^2}}\,.$$

Split the sums into dyadic intervals. Applying (2.9) to the first sum then gives

$$\sum_{\kappa_j \ll X^{1+\varepsilon}} \alpha_j H_j^4(1/2) \kappa_j^{-2} \ll X^{\varepsilon}\,,$$

while (2.74) similarly gives

$$\sum_{\kappa_j \leq X^{1+\varepsilon}} \alpha_j \kappa_j^{-1} |L_j(1/2+i\delta)|^2 \ll X^{1+\varepsilon} + \delta^{2/3+\varepsilon} X^{\varepsilon}\,.$$

Thus, substituting $X = \tilde{T}/\Delta$,

$$E_2 \ll \tilde{T}^{1+\varepsilon}/\Delta^{1/2} + \tilde{T}^{1/2+\varepsilon}\delta^{1/3}\,.$$

Similarly for E_3, if $\delta \leq T + \alpha$, applying the first case in (2.72) together with (2.74), and then Cauchy's inequality,

$$E_3 \ll \delta^{1/2} \left(\frac{\delta}{T+\alpha}\right)^5 \sum_{6 \leq k \leq X^{1+\varepsilon}} \sum_{j=1}^{\vartheta(k)} \alpha_{j,k} |L_{j,k}(1/2+i\delta)| H_{j,k}^2(1/2) 2^{-k} \ll \delta^{5/6+\varepsilon}\,.$$

If $\delta > T + \alpha$, then applying the second case in (2.72) instead,

$$E_3 \ll \delta^{1/2} \sum_{6 \leq k \ll T^{\varepsilon}\sqrt{\frac{\delta}{T+\alpha}}} \sum_{j=1}^{\vartheta(k)} \alpha_{j,k} |L_{j,k}(1/2+i\delta)| H_{j,k}^2(1/2) k^{-2} \ll \delta^{5/6+\varepsilon}\,.$$

Thus the total error is

$$E_1 + E_2 + E_3 \ll \tilde{T}^{1+\varepsilon}/\Delta^{1/2} + \tilde{T}^{1/2+\varepsilon}\delta^{1/3} + \delta^{5/6+\varepsilon}\,.$$

In particular, this restricts us to $\delta \ll T^{6/5-\varepsilon}$. □

Theorem 2.5.7 (Off-diagonal term)**.** *Let $0 \leq \alpha \leq \beta \ll T^{8/5}$. Set $\delta = \beta - \alpha$ and $\tilde{T} := \sqrt{(T+\alpha)(T+\beta)}$. Additionally, assume that $\delta \gg (\tilde{T}/\Delta)^{2+\varepsilon}$. Finally, let*

$$h(z,s) = \frac{\zeta(1+z+s)\zeta(1-z+s)}{\zeta(2+2s)}\,,$$

and

$$H_{\alpha,\beta}(t) = \frac{d^2}{ds^2}\left(h(i\delta,s)\frac{(t+\alpha)^{s/2}(t+\beta)^{s/2}}{(2\pi)^s}e^{2\gamma_0 s}\right)\Big|_{s=0} - \frac{1}{4}h(i\delta,0)\log^2\frac{t+\alpha}{t+\beta}\,.$$

Then

$$2\,\mathrm{Re}\sum_{n\neq m}\frac{d(n)d(m)}{n^{1/2+i\alpha}m^{1/2-i\beta}}\int W(t)\big(\frac{m}{n}\big)^{it}\kappa(t)V_{\alpha,\beta}(nm,t)dt = \int_0^\infty W(t)H_{\alpha,\beta}(t)dt + O\left(\frac{\tilde{T}^2}{\Delta\sqrt{\delta}} + \delta^{5/6+\varepsilon} + \delta^{1/3+\varepsilon}\tilde{T}^{1/2+\varepsilon}\right).$$

In particular, $\Delta = \tilde{T}^{1-\varepsilon}/\delta^{1/4}$ *gives error term of size*

$$\frac{\tilde{T}^{1+\varepsilon}}{\delta^{1/4}} + \delta^{5/6+\varepsilon} + \delta^{1/3+\varepsilon}\tilde{T}^{1/2+\varepsilon} = o(T)$$

when $\delta \ll T^{6/5-\varepsilon}$ *and* $(T+\alpha)(T+\beta) \ll T^{2-\varepsilon}\delta^{1/2}$.

Proof. The main term is given by

$$\sum_{r=1}^{\infty}\frac{1}{r^{1-i\delta}}\int_0^\infty S(x,r)W_1(\delta x-\alpha)dx\,. \tag{2.75}$$

Changing the order of summation and integration due to absolute convergence, and summing over r, we have

$$\sum_{r=1}^{\infty}\frac{S(x,r)}{r^{1-i\delta}}I(\delta/2\pi r) = \frac{1}{2\pi i}\int_{(2)}L(x,1-i\delta+s)\left(\frac{\delta}{2\pi}\right)^s\frac{G(s)}{s}ds\,, \tag{2.76}$$

where $L(x,1+z) = \lim_{h\to 0}L_h(x,1+z)$ and

$$L_h(x,1+z) = \frac{\zeta^2(1+h)}{\zeta(2+2h)}\zeta(1+z)\zeta(z-2h)x^h(1+x)^h + \frac{\zeta(1+h)\zeta(1-h)}{\zeta(2)}\zeta(1+z-h)\zeta(z-h)(x^h+(1+x)^h) + \frac{\zeta^2(1-h)}{\zeta(2-2h)}\zeta(1+z-2h)\zeta(z)\,.$$

Moving the line of integration to $\mathrm{Re}\,s = 0$ with a small half-circle around 0 gives $1/2$ of the residue at 0, residues at $s = \pm i\delta$ with

$$\frac{1}{2}L(x,1-i\delta) + O(e^{-\delta^2/Q}) + \frac{1}{2\pi i}\int_{|\Im s|\geq a}L(x,1-i\delta+s)\left(\frac{\delta}{2\pi}\right)^s\frac{G(s)}{s}ds\,. \tag{2.77}$$

Similarly to (2.65) in Theorem 2.5.3 observe that

$$\operatorname{Re}\frac{1}{2\pi i}\int_{|y|\geq a} L(x,1-i\delta+s)\left(\frac{\delta}{2\pi}\right)^{iy}\frac{G(iy)}{y}dy \ll \frac{Q^{1+\varepsilon}}{\delta^{1/2}} \ll \delta^{-1/2+\varepsilon}\,.$$

Finally, changing variables to $t=\delta x-\alpha$ in (2.75) we directly obtain

$$\int_0^\infty W(t)H_{\alpha,\beta}(t)dt\,.$$

The error term follows after collecting errors from Lemma 2.4.3, 2.5.4 and 2.5.6 with total

$$\frac{\tilde{T}^2}{\Delta\sqrt{\delta}}+\frac{\tilde{T}^{1+\varepsilon}}{\Delta^{1/2}}+\delta^{1/2+\varepsilon}+\frac{T+\beta}{\delta^{1/2}}+\delta^{1/3}(T+\beta)^{1/2+\varepsilon} \ll \frac{\tilde{T}^2}{\Delta\sqrt{\delta}}+\delta^{1/3}(T+\beta)^{1/2+\varepsilon}\,.$$

This concludes the proof. □

2.5.3 Diagonal term and main theorem

To conclude our computation, let us finally evaluate the diagonal term.

Theorem 2.5.8 (Diagonal term)**.** *Let W be an infinitely differentiable function supported on $[T,2T]$ chosen with parameter Δ. Let $\alpha,\beta\ll T^{2-\varepsilon}$, and set $\delta=|\alpha-\beta|$. Then*

$$2\operatorname{Re}\sum_{n\geq 1}\frac{d^2(n)}{n^{1+i\delta}}\int_{-\infty}^{+\infty}W(t)\kappa(t)V_{\alpha,\beta}(n^2,t)dt = 2\operatorname{Re}\operatorname*{Res}_{s\in\{0,i\delta\}}L(1+s)\frac{\tilde{W}(1+s)G(s-i\delta)}{s-i\delta}+O(T^{1/2+\varepsilon}) = 2T\operatorname{Re}(F(2T)-F(T))+O(T^{\varepsilon}\Delta+T^{1+\varepsilon}/Q)\,,$$

where

$$F(X)=\operatorname*{Res}_{s\in\{0,i\delta\}}\frac{\zeta^4(1+s)}{\zeta(2+2s)}\frac{(X/2\pi)^s}{s-i\delta}\,.$$

In fact, if $\delta\gg(T+\alpha)^{1/3}$, the whole expression is $O((T+\alpha)^{2/3})$.

Proof. If $\delta\gg T^{1/3}$ using oscillation lemma

$$\int W(t)\kappa(t)V_{\alpha,\beta}(n^2,t)dt \ll \frac{1}{\log\frac{T+\alpha}{T+\beta}}\max_t|W(t)V_{\alpha,\beta}(n^2,t)|$$

thus the whole expression is bounded as follows:

$$\frac{1}{\log\frac{T+\alpha}{T+\beta}}\sum_{n\ll T^4}\frac{d^2(n)}{n}\ll T^{\varepsilon}\log^{-1}\frac{T+\alpha}{T+\beta}\ll T^{\varepsilon}\left(\frac{T+\alpha}{\delta}+1\right)\ll(T+\alpha)^{2/3}\,.$$

If $\delta \ll (T+\alpha)^{1/3}$, set $\tilde{W}(s) = \mathcal{M}W(s)$, and

$$L(s) = \frac{\zeta^4(s)}{\zeta(2s)}\,.$$

Then using Lemma 2.5.1

$$\sum_{n\geq 1} d^2(n) \int W(2\pi nw - \alpha) w^{i\delta} I(w) dw = \frac{1}{2\pi i}\int_{(2)} L(1+s)\tilde{W}(s+1)G(s-i\delta)\frac{ds}{s-i\delta} =$$
$$\operatorname*{Res}_{s\in\{0,i\delta\}} L(1+s)\frac{\tilde{W}(1+s)G(s-i\delta)}{s-i\delta} + O(T^{1/2+\varepsilon})\,.$$

Further, computing the residues and taking real part we easily see that the above equals

$$\operatorname*{Res}_{s\in\{0,i\delta\}} \frac{\zeta^4(1+s)}{\zeta(2+2s)}\frac{((T+\alpha)/2\pi)^s}{(s+1)(s-i\delta)} + O(\Delta + T/Q + T^{1/2+\varepsilon})\,.$$

If $\delta = 0$ we only have one pole with

$$\operatorname{Res}_{s=0}\frac{\zeta^4(1+s)}{\zeta(2+2s)}\frac{(T/2\pi)^s}{s(s+1)}\,.$$

This corresponds to the diagonal term of the 4-th moment of the Riemann zeta. □

Putting together Theorems 2.5.8, 2.5.3 and 2.5.7, and summing over dyadic intervals yields Theorem 2.1.1. Since α and β are fixed, and T varies, we apply the appropriate version of the main theorem to each range. Note that when $T^{1/3} \ll \delta \ll T^{1/2-\varepsilon}$, the bound from Theorem 2.5.7 is better than the one in Theorem 2.5.3. If $E(X)$ is the error term on the interval $[X, 2X]$ and $\tilde{X} = \sqrt{(X+\alpha)(X+\beta)}$, then

$$E(X) \ll T^{\varepsilon}\begin{cases} \tilde{X}^{4/5}, & \text{if } \delta \ll \tilde{X}^{1/3}\,; \\ \frac{\tilde{X}}{\delta^{1/4}} + \delta^{1/3}\tilde{X}^{1/2} + \delta^{5/6}, & \text{if } \delta \gg \tilde{X}^{1/3}\,. \end{cases}$$

Thus

$$\sum_{j\leq \log T} E(T/2^j) \ll T^{\varepsilon}\begin{cases} \tilde{T}^{4/5} + \delta^{11/4}, & \text{if } \delta \ll \tilde{T}^{1/3}\,; \\ \frac{\tilde{T}}{\delta^{1/4}} + \delta^{1/3}\tilde{T}^{1/2} + \delta^{5/6}, & \text{otherwise}\,. \end{cases}$$

In particular, this is bounded by $O((T+\alpha)^{11/12+\varepsilon} + \delta^{5/6+\varepsilon})$. This bound is not optimal, and can be further improved by doing the stationary phase analysis more carefully, except for the biggest contribution coming from E_2 and E_3 in Lemma 2.5.6.

2.6 Discussion and prospective work

In this chapter we presented two results, though their significance is not necessarily proportional to the amount of work required.

First, we computed the correlations of $|\zeta|^2$

$$\frac{1}{T}\int_0^T |\zeta(\frac{1}{2}+it+i\alpha)|^2|\zeta(\frac{1}{2}+it+i\beta)|^2dt$$

up to a power saving for a wide range of shifts. We avoided the issue of generalised divisor functions by restricting to two shifts from the start. A similar idea was used in the work of Ng [109] and Ng, Shen and Wong [110], though their goal was to evaluate the corresponding moments of zeta and not correlations.

We did not attempt to optimise the errors since our interest was more in terms of quality rather than quantity. For example, Section 2.5 with small δ is very "lossy". Still, one may wonder what is the largest range for $\delta = |\alpha - \beta|$. For $\delta \gg T^{1/2}$ the largest error came from

$$\sum_{j=1}^{\infty} \alpha_j H_j(1/2+i\delta)H_j^2(1/2)\Theta(\kappa_j;W)\,;$$
$$\sum_{k=6}^{\infty}\sum_{j=1}^{\vartheta(k)}(-1)^k\alpha_{j,k}H_{j,k}(1/2+i\delta)H_{j,k}^2(1/2)\Xi(k-1/2;W)\,,$$

so improving the estimates on these sums would automatically extend the range. However, naively, it seems that $(T+\alpha)(T+\beta) \ll T^3$ is the limit of this method, so it might not be possible to go beyond this range without any new ideas, if at all.

Now let us discuss moments of moments. Though a simple corollary of the result above, we initially found the explicit expression for $\mathcal{M}_{2,2}$ surprising when c is a constant, but this perhaps is an important lesson to be careful around critical points.

Further, we may observe that "moments of moments" are in fact usual moments of an extended family. This becomes more apparent as $c = c(T)$ grows especially comparing to [43] and [31].

It would be interesting to see how our result translates to other families of L-functions. We expect that if one can compute the k-th moment of a family, then one should at least be able to compute the corresponding moment of moment. This can

be in either of the parameters of the corresponding L-functions:

$$\sum_{\chi \bmod q} \left(\int_{-1}^{1} L(1/2 + it + ih, \chi)^l dh \right)^k ;$$

$$\sum_{\chi \bmod q} \left(\sum_{\psi \bmod c} L(1/2 + it, \chi\psi)^l \right)^k .$$

Finally, one may hope to obtain the following results. It seems relatively straightforward to prove upper and lower bounds for $\mathcal{M}_{k,\gamma}$ of the right order adapting the approach of Harper [68]. In particular, the intermediate case has recently been proved by Ng, Shen and Wong [111]. In a slightly different direction, one may also ask what conjectures on the error term for generalised divisor function correlations would allow us to compute higher order moments of moments, say when $k = 3$. The most technically challenging but general option is of course a conjectural error term for

$$\sum_{n \leq X} \tau_\alpha(n) \tau_{-\beta}(n + r) .$$

Chapter 3

Equivalence classes of quadratic forms

In this chapter we consider random symmetric matrices with independent entries distributed according to the Haar measure on $\mathbb{Z}_p$ for odd primes p and derive the distribution of their canonical form with respect to several equivalence relations. We give a few examples of applications including an alternative proof for the result of Bhargava, Cremona, Fisher, Jones, and Keating on the probability that a random quadratic form over $\mathbb{Z}_p$ has a non-trivial zero. This chapter is based on [95].

3.1 Introduction

Let p be prime, and let $\mathbb{Z}_p$ be the ring of p-adic integers. Denote the set of $n \times n$ matrices over $\mathbb{Z}_p$ by $\mathrm{M}_n(\mathbb{Z}_p)$, and its subset of invertible matrices by $K := \mathrm{GL}_n(\mathbb{Z}_p)$.

For any matrix $X \in \mathrm{M}_n(\mathbb{Z}_p)$ there exist matrices $U, V \in K$ such that there is a decomposition

$$X = U\Sigma V\,,$$

where $\Sigma = \mathrm{diag}(p^{k_1}, \ldots, p^{k_n})$ and k_i are the unique integers satisfying

$$0 \leq k_1 \leq \cdots \leq k_n \leq +\infty\,.$$

The numbers p^{k_j} define the equivalence class of X with respect to left and right multiplication by $\mathrm{GL}_n(\mathbb{Z}_p)$. These numbers called the elementary divisors of X are analogous to the singular values of complex matrices.

Let $X \in \mathrm{M}_n(\mathbb{Z}_p)$ be a random matrix with independent entries distributed according to the Haar measure on $\mathbb{Z}_p$. This entry-wise measure on $\mathrm{M}_n(\mathbb{Z}_p)$ is in fact the (multiplicative and additive) Haar measure on $\mathrm{M}_n(\mathbb{Z}_p)$ as a group. The elementary

divisors of X are then random variables, and the probability that a random matrix X is equivalent to Σ is the measure of the double coset $K\Sigma K$.

Elementary divisors come up in many different contexts, and in each one of them they exhibit a different kind of structure. First, their joint distribution was studied by Macdonald [97] from the point of view of symmetric functions and the Hecke ring of GL_n. It follows from his work that for $\mathbf{k} = (k_1, \ldots, k_n)$ and $\Sigma = \mathrm{diag}(p^{\mathbf{k}})$ the measure of the double coset $K\Sigma K$ equals

$$\pi_n \; p^{-\sum_i (n-i)k_i} P_{\mathbf{k}}(p^{-1}, p^{-2}, \ldots, p^{-n}; p^{-1}) \,, \tag{3.1}$$

where $\pi_n := p^{-n^2} |\mathrm{GL}_n(\mathbb{F}_p)|$ is the density of invertible $n \times n$ matrices over a finite field $\mathbb{F}_p$ and $P_{\mathbf{k}}(x_1, \ldots, x_n; t)$ is the Hall-Littlewood polynomial of the partition $\mathbf{k}$.

If matrix X has elementary divisors $p^{k_1}, \ldots, p^{k_n}$, then its cokernel $\mathrm{cok}\, X$ satisfies

$$\mathrm{cok}\, X \cong \bigoplus_{i=1}^{n} \mathbb{Z}/p^{k_i}\mathbb{Z} \,,$$

so studying elementary divisors is equivalent to studying cokernels. Friedman and Washington [56] were the first to draw a connection between the distribution of cokernels of random matrices and the Cohen-Lenstra heuristics in their work on the distribution of class groups of curves over a finite field. They also showed that for a finite abelian p-group H of p-rank r

$$\mathbb{P}\{\mathrm{cok}\, X \cong H\} = \frac{\pi_n^2}{\pi_r \, |\mathrm{Aut}\,(H)|} \,.$$

This was generalised to other distributions of entries in the asymptotic case as n grows by Maples [99] and Matchett Wood [133].

Evans [54] studied the distribution of elementary divisors from the probabilistic point of view similar to Brent and McKay [24], and Fulman [57]. He showed that one can construct a Markov chain from the multiplicities m_k of elementary divisors, and represented the distribution

$$\mathbb{P}\{X \in K\Sigma K\} = \frac{\pi_n^2}{\pi_{m_0}\pi_{m_1}\cdots\pi_{m_k}\cdots} p^{-(n-m_0)^2-(n-m_0-m_1)^2-\ldots-(n-m_0-\ldots-m_k)^2-\ldots} \,.$$

in the form

$$\mathbb{P}\{X \in K\Sigma K\} = \prod_{j=0}^{\infty} \frac{\pi_{l_j}^2}{\pi_{m_j}\pi_{l_j-m_j}^2} p^{-(l_j-m_j)^2}$$

where $l_0 = n$ and $l_j = n - m_0 - \ldots - m_{j-1}$ for $j \geq 1$.

Finally, for a matrix $X \in K\Sigma K$ the matrix Σ is sometimes referred as its Smith normal form, in this context see, for example, the work of Wang and Stanley [130].

Restricting to symmetric matrices allows us to consider other similar types of decompositions. As general matrices represent bilinear forms, symmetric matrices represent quadratic forms, and this connection provides a helpful way to interpret the decompositions for symmetric matrices.

Let $S_n(\mathbb{Z}_p)$ be the set of symmetric $n \times n$ matrices over $\mathbb{Z}_p$. For any matrix $X \in S_n(\mathbb{Z}_p)$ there exists a matrix $U \in \mathrm{GL}_n(\mathbb{Z}_p)$ such that there is a decomposition

$$X = U\Sigma S U^\top,$$

where $\Sigma = \mathrm{diag}(p^{k_1}, \ldots, p^{k_n})$ is the matrix of elementary divisors with $k_1 \leq \ldots \leq k_n$, and $S = \mathrm{diag}(a_1, \ldots, a_n)$ is the matrix of signatures. Let $r \in \mathbb{Z}_p^\times$ is a fixed non-square in $\mathbb{Z}_p$, then $a_i \in \{1, r\}$. Once put in the canonical form (depending on Σ), the signatures a_i together with the elementary divisors define the $\mathbb{Z}_p$-equivalence class of X. This decomposition of course corresponds to diagonalising quadratic forms.

Studying $\mathbb{Z}_p$-equvalence classes is almost equivalent to studying the cokernel of X with a given duality pairing. Clancy, Kaplan, Leake, Payne, and Matchett Wood [32] show that the probability that X has a cokernel H of p-rank r with a fixed duality pairing δ equals

$$\frac{\pi_n}{\beta_{n-r}|H||\mathrm{Aut}\,(H,\delta)|}, \tag{3.2}$$

where $\beta_t = \prod_{i=1}^{\lfloor \frac{t}{2} \rfloor}(1 - p^{-2i})$. The asymptotic distribution of cokernels of random symmetric p-adic matrices with independent entries was further studied by Matchett Wood [132].

Let $\mathbb{Q}_p$ denote the field of p-adic numbers, and $\mathrm{GL}_n(\mathbb{Q}_p)$ denote the set of all invertible matrices over $\mathbb{Q}_p$. For any matrix $X \in S_n(\mathbb{Z}_p)$ there exists a matrix $U \in \mathrm{GL}_n(\mathbb{Q}_p)$ such that there is a decomposition $X = UDU^\top$, where D is the canonical form for the $\mathbb{Q}_p$-equivalence class of X. This corresponds to the canonical form of quadratic forms over $\mathbb{Q}_p$. The classification of quadratic forms over $\mathbb{Q}_p$ is well known, see Cassels [29], but the distribution of such classes is much less studied. A related work in this direction is the paper of Bhargava, Cremona, Fisher, Jones, and Keating [15] on the probability of isotropy of integral quadratic forms.

In this chapter we consider random symmetric matrices with entries distributed independently according to the Haar measure on $\mathbb{Z}_p$ for odd primes p, and compute the distribution of equivalence classes with respect to conjugation by $\mathrm{GL}_n(\mathbb{Z}_p)$ and $\mathrm{GL}_n(\mathbb{Q}_p)$. This is a natural choice of measure for two reasons. First, this is simply the

counting measure on $S_n(\mathbb{Z}_p)$. Second, the Haar measure on the ring of integers of a local field is analogous to the standard Gaussian measure over $\mathbb{R}$ due to its invariance properties (for exposition see Evans [53]). This makes such a random matrix the analogue of the Gaussian Ensemble of Random Matrix Theory, so it is reasonable to expect that the asymptotic behaviour of such matrices is universal i.e. showcases the asymptotic behaviour of a larger class of random matrices. This turns out to be true as shown in the work of Matchett Wood [132]. Additionally, due to the invariance of this measure on $S_n(\mathbb{Z}_p)$ with respect to conjugation by $\mathrm{GL}_n(\mathbb{Z}_p)$, it is an analogue of the Gaussian Orthogonal Ensemble (GOE).

Our approach to studying symmetric matrices was motivated by classical Random Matrix Theory and the "local GOE" analogy rather than the Cohen-Lenstra heuristics. We believe that the idea of reducing a certain class of problems to studying the joint distribution of eigenvalues or singular values may provide a more unified framework over different fields.

We present the following results. First, we give an elementary Random Matrix Theory inspired proof for the distribution of $\mathbb{Z}_p$-equivalence classes. We note again that this distribution is essentially the same as (3.2) with a small difference of taking into account the signature for $k_j = 0$, so this result though independent is not entirely new.

Theorem 3.1.1. *Let p be an odd prime, and X be a random $n \times n$ symmetric matrix with independent entries distributed uniformly on $\mathbb{Z}_p$. Let $\Sigma = diag(p^{k_1}, \ldots, p^{k_n})$ be its matrix of elementary divisors and $S = diag(1, \ldots, 1, a_0, \ldots, 1, \ldots, 1, a_k, \ldots)$ be an admissible matrix of signatures in the canonical form. Let $m_k = \#\{i : k_i = k\}$ and s_k be the quadratic character mod p at a_k. Then*

$$\mathbb{P}\{X \in K\Sigma S K^\top\} = \frac{\pi_n}{\alpha_{m_0}^{s_0}\alpha_{m_1}^{s_1}\cdots\alpha_{m_k}^{s_k}\cdots}\prod_{j=1}^{n} p^{-k_j(n-j+1)},$$

where $\alpha_j^s = p^{-j(j-1)/2}|O_j^s(\mathbb{F}_p)|$ is the density of the orthogonal group over $\mathbb{F}_p$ for $j \geq 1$, and $\alpha_0^+ = 1$.

Corollary 3.1.2. *For a random symmetric matrix defined as above*

$$\mathbb{P}\{X \in K\Sigma K\} = \frac{\pi_n}{\beta_{m_0}\cdots\beta_{m_k}\cdots}\prod_{j=1}^{n} p^{-k_j(n-j+1)},$$

where $\beta_t = ((\alpha_j^+)^{-1} + (\alpha_j^-)^{-1})^{-1} = \prod_{i=1}^{\lfloor \frac{t}{2} \rfloor}(1 - p^{-2i})$.

Note that $\sum_k m_k = n$, so only finitely many m_k are positive, and the product in the denominator is effectively finite. Moreover, if $m_k = 0$, then $s_k = +$.

There are multiple elementary ways to prove Theorem 3.1.1. The one we suggest is as follows. Fix Σ and reduce to a counting problem over $\mathbb{Z}/p^K\mathbb{Z}$ for some big K. Compute the size of the stabiliser of ΣS in $\mathbb{Z}/p^K\mathbb{Z}$ by counting solutions to $U\Sigma SU^\top = \Sigma S$. This system is easy to solve over $\mathbb{F}_p$, then we can lift the solutions back to $\mathbb{Z}/p^K\mathbb{Z}$ and eventually $\mathbb{Z}_p$ via Hensel's lemma. In spirit, this is similar to computing the Jacobian of the transformation from the entry-wise to the eigenvalue-eigenvector coordinates as done for the Gaussian Orthogonal Ensemble.

Further, we use the distribution of $\mathbb{Z}_p$-classes to obtain a closed expression for the distribution of $\mathbb{Q}_p$-classes.

Theorem 3.1.3. *Let p be an odd prime, and let ε be the quadratic character mod p at -1. Let Q be a random quadratic form in n variables with independent coefficients distributed according to the Haar measure on $\mathbb{Z}_p$. Let $d(Q) \in \mathbb{Z}_p/(\mathbb{Z}_p^2)$ be its discriminant and $c(Q) \in \{\pm 1\}$ be its Hasse invariant. Then if n is odd,*

$$\mathbb{P}\{d(Q) = a, c(Q) = b\} = \begin{cases} \frac{1}{4(1+1/p)} + b\,\frac{\pi_n}{4\beta_{n+1}\beta_{n-1}}\,, & \text{if } a \in \{1, r\}\,; \\ \\ \frac{1}{4p(1+1/p)} + b\,\varepsilon\left(\frac{\varepsilon}{p}\right)^{\frac{n+1}{2}}\frac{\pi_n}{4\beta_{n+1}\beta_{n-1}}, & \text{if } a \in \{p, pr\}\,; \end{cases}$$

and if n is even and s is the quadratic character mod p at a,

$$\mathbb{P}\{d(Q) = a, c(Q) = b\} = \begin{cases} \frac{(1+s\left(\frac{\varepsilon}{p}\right)^{n/2})(1-\frac{s}{p^2}\left(\frac{\varepsilon}{p}\right)^{n/2})}{4(1+1/p)(1-1/p^{n+1})} + b\,\frac{\pi_n}{2\beta_n\alpha_n^s}\,, & \text{if } a \in \{1, r\}\,; \\ \\ \frac{1-1/p^n}{4p(1+1/p)(1-1/p^{n+1})}\,, & \text{if } a \in \{p, pr\}\,. \end{cases}$$

We also provide a few examples of applications for the densities obtained. First, we use the distribution of elementary divisors from Cor 3.1.2 to derive the distribution of ranks and determinants over $\mathbb{Z}/p^k\mathbb{Z}$. Second, we use the local-global principle to compute the probability that a random integer matrix has square-free determinant, and more generally, a square-free minor of size m.

Corollary 3.1.4. *Let $r, n \in \mathbb{N}$ with $1 \le r \le n$. Then as $H \to \infty$ the proportion of integer $n \times n$ symmetric matrices with entries between 1 and H that have a square-free minor of size $n - r$ tends to*

$$\prod_p \left(\sum_{t=0}^{r+1} \frac{\pi_n}{\beta_{n-t}\pi_t} p^{-\frac{t(t+1)}{2}} - \frac{\pi_n}{\beta_{n-r-1}\pi_{r+1}} p^{-(r+1)(r+2)} \right) > 0\,.$$

In particular, if $r = 0$, the proportion of matrices with square-free determinant is

$$\prod_p \left(\frac{\pi_n}{\beta_n} + p^{-1} \frac{\pi_n}{\beta_{n-1}} \right) > 0 \,.$$

As $n \to \infty$ this expression tends to $(\zeta(2)\zeta(3)\zeta(5)\cdots\zeta(2k+1)\cdots)^{-1} \approx 0.4824$.

We note that the statement for $r \leq n-1$ follows the infinite version of the Chinese remainder theorem known as Ekedahl's sieve [51], while $r = 0$ requires us to first prove that such a local product applies. In general, unconditionally one can only count square-free values of multivariate polynomials when the number of variables is exponential in the degree of the polynomial with the best bound on the number of variables given by Destagnol and Sofos [48]. There are also exceptional "nice" polynomials, where such a restriction does not apply. This includes several univariate invariant polynomials due to Bhargava [11], the discriminant of a general univariate polynomial with coefficients ordered by height, and the discriminant of binary n-ic forms ordered as above due to Bhargava, Shankar and Wang [17, 18]. Poonen [114] proved that such a local product applies under the ABC-conjecture, but in a weaker sense when one variable is much larger than the others, and this statement extends to multivariate polynomials with a variable of degree ≤ 3 without assuming the ABC-conjecture. In the case of the determinant of either a general or a symmetric matrix, one can notice that it is also an example of an invariant polynomial, and utilise its symmetry properties to prove the corresponding statement.

Finally, we show several ways of deriving the probability of isotropy i.e. existence of a non-trivial zero for random quadratic forms in $n \leq 4$ variables. A quadratic form over $\mathbb{Z}_p$ in $n \geq 5$ is always isotropic and for the cases of $n = 2, 3, 4$ the probability of isotropy was obtained by Bhargava, Cremona, Fisher, Jones, and Keating [15] through a recursive computation over finite fields. Though their method lies in a different framework of recursive computation, and is capable of solving a much bigger class of problems (see Bhargava [12], Bhargava, Cremona, and Fisher [13] [14]), ours is more straightforward specifically for quadratic forms.

Corollary 3.1.5. *Let p be an odd prime, and let Q be a random quadratic form in n variables with coefficients distributed independently according to the Haar measure on $\mathbb{Z}_p$. Then*

$$\mathbb{P}\{Q \text{ is isotropic}\} = \begin{cases} \frac{1}{2}\,, & \text{if } n = 2\,; \\ 1 - \frac{1}{2(1+1/p)^2}\,, & \text{if } n = 3\,; \\ 1 - \frac{1-1/p}{4(1+1/p)^2(1-1/p^5)}\,, & \text{if } n = 4\,. \end{cases}$$

Finally, we would like to add a bit more historical context on the development of Random Matrix Theory over different fields. Random Matrix Theory over fields other than complexes has mostly been evolving independently of the theory over the complex numbers. For an overview of Random Matrix Theory over finite fields and in particular cycle index techniques see Fulman [57], and for an overview of rank problems and Stein's method see Fulman and Goldstein [58]. One of the recent results exploring the similarities between theory over different fields is the work of Gorodetsky and Rodgers [64], on the finite field analogue of the fundamental result of Diaconis and Shahshahani [49] on the convergence of traces of unitary matrices to the normal distribution.

Random p-adic matrices have been studied to a smaller extent. Apart from the ones already mentioned, some of the recent developments in this area are the works of Ellenberg, Jain, and Venkatesh [52] on modelling λ-invariants, and Poonen and Rains [116], Bhargava, Kane, Lenstra, Poonen, and Rains [16], Poonen [115], Park, Poonen, and Matchett Wood [113] on the heuristics for ranks of elliptic curves based on random symmetric and alternating matrices.

3.2 Preliminaries

Let us first discuss general p-adic matrices without the symmetry constraint. Let $\mathrm{M}_n(\mathbb{Q}_p)$ be the set of all $n \times n$ matrices with entries in p-adic numbers $\mathbb{Q}_p$, and $\mathrm{M}_n(\mathbb{Z}_p)$ be its subset of matrices with integral entries. Let ν be the Haar measure on $\mathbb{Q}_p$ normalised so that $\nu(\mathbb{Z}_p) = 1$, and let μ be the measure on $\mathrm{M}_n(\mathbb{Q}_p)$ such that it is the direct product of n^2 measures ν on its entries. We also trivially have that $\mu\left(\mathrm{M}_n(\mathbb{Z}_p)\right) = 1$.

The general linear group $\mathrm{GL}_n(\mathbb{Q}_p)$ is the set of invertible $n \times n$ matrices with entries in $\mathbb{Q}_p$. The general linear group $K := \mathrm{GL}_n(\mathbb{Z}_p)$ is its subgroup of invertible $n \times n$ matrices over $\mathbb{Z}_p$, or

$$\mathrm{GL}_n(\mathbb{Z}_p) := \{A \in \mathrm{M}_n(\mathbb{Z}_p) : \ \det(A) \in \mathbb{Z}_p^\times\} = \{A \in \mathrm{M}_n(\mathbb{Z}_p) : \ |\det(A)|_p = 1\}.$$

As a set $\mathrm{GL}_n(\mathbb{Z}_p)$ is both open and compact of measure

$$\pi_n := \mu\big(\mathrm{GL}_n(\mathbb{Z}_p)\big) = \prod_{k=1}^{n} \left(1 - p^{-k}\right).$$

As a group $\mathrm{GL}_n(\mathbb{Z}_p)$ is the largest compact subgroup of $\mathrm{GL}_n(\mathbb{Q}_p)$, and in this sense it corresponds to the unitary group over Euclidean spaces. Unitary matrices represent

isometries, and preserve the Frobenius norm of a complex matrix, while the matrices from $\mathrm{GL}_n(\mathbb{Z}_p)$ instead preserve the max-norm $\|A\|_p = \max |a_{ij}|_p$.

Lemma 3.2.1. *Let $A \in M_n(\mathbb{Q}_p)$ and $\gamma \in GL_n(\mathbb{Z}_p)$. Then $\|\gamma A\|_p = \|A\gamma\|_p = \|A\|_p$.*

Proof. For any $\gamma \in \mathrm{GL}_n(\mathbb{Z}_p)$ we have $\|\gamma\|_p = \|\gamma^{-1}\|_p = 1$. Let $B = \gamma A$, then for any $b_{ij} = \sum_{k=1}^n \gamma_{ik} a_{kj}$ we have

$$|b_{ij}|_p \leq \max_k |\gamma_{ik} a_{kj}|_p \leq \max |\gamma_{ik}|_p \max |a_{kj}|_p \leq \|\gamma\|_p \|A\|_p = ||A||_p \,,$$

hence $\|B\|_p \leq \|A\|_p$. Conversely, for $A = \gamma^{-1} B$, we also have

$$\|A\|_p \leq \|\gamma^{-1}\|_p \|B\|_p = \|B\|_p$$

so $\|B\|_p = \|A\|_p$. □

Any complex matrix has a canonical form obtained via the singular value decomposition. This corresponds to factoring out by the unitary group $\mathrm{U}_n(\mathbb{C})$, which is a compact maximal subgroup of $\mathrm{GL}_n(\mathbb{C})$.

3.2.1 General matrices

For any matrix $A \in \mathrm{M}_n(\mathbb{C})$ there exist matrices $U, V \in \mathrm{U}_n(\mathbb{C})$ such that $A = USV^*$, where S is the diagonal matrix with entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$ defined uniquely. These are called the singular values of A. Similarly, for any $A \in \mathrm{M}_n(\mathbb{Q}_p)$ there exists a decomposition $A = U\Sigma V$, where $U, V \in \mathrm{GL}_n(\mathbb{Z}_p)$, and $\Sigma = \mathrm{diag}(p^{k_1}, \ldots, p^{k_n})$ with k_i integers such that $-\infty < k_1 \leq \ldots \leq k_n \leq +\infty$. This sequence is unique and is called the sequence of elementary divisors. Note that $k_i = +\infty$ is possible with $p^{+\infty} = 0$ since $|p^{+\infty}|_p = p^{-\infty} = 0$. In this case the matrix is singular. Such a decomposition exists more generally over principal ideal rings, and is sometimes called the Smith normal form. For general theory see Brown's book [26].

If $X \in \mathrm{M}_n(\mathbb{Z}_p)$, then all k_i are non-negative. Moreover, we can also assume that $k_i < +\infty$, since the set of singular matrices

$$V = \{\mathbf{X} \in \mathrm{M}_n(\mathbb{Q}_p) : \det \mathbf{X} = 0\}$$

has measure zero. The proof of this fact is analogous to the case of $\mathbb{R}$ and Lebesgue's measure and relies on the property that the graph of a rational function defined over an open set has measure zero. So further we will restrict to the group of non-singular matrices $\tilde{\mathrm{M}}_n(\mathbb{Z}_p) := \mathrm{M}_n(\mathbb{Z}_p) \cap \mathrm{GL}_n(\mathbb{Q}_p)$. Now let us give a simple proof for the distribution of elementary divisors for general matrices over $\mathbb{Z}_p$.

For $\gamma \in \tilde{\mathrm{M}}_n(\mathbb{Z}_p)$ define $C_\gamma := K \cap \gamma K \gamma^{-1}$. In particular, for $\Sigma = \mathrm{diag}(p^{k_1}, \dots, p^{k_n})$ with multiplicities $m_k = \#\{i : k_i = k\}$ we have

$$C_\Sigma = \begin{bmatrix} \mathrm{GL}_{m_0}(\mathbb{Z}_p) & M_{m_0\times m_1}(\mathbb{Z}_p) & M_{m_0\times m_2}(\mathbb{Z}_p) & & \cdots & \\ M_{m_1\times m_0}(p\mathbb{Z}_p) & \mathrm{GL}_{m_1}(\mathbb{Z}_p) & M_{m_1\times m_2}(\mathbb{Z}_p) & & \cdots & \\ M_{m_2\times m_0}(p^2\mathbb{Z}_p) & M_{m_2\times m_1}(p\mathbb{Z}_p) & \mathrm{GL}_{m_2}(\mathbb{Z}_p) & & \cdots & \\ \vdots & \vdots & & \ddots & & \\ M_{m_k\times m_0}(p^k\mathbb{Z}_p) & M_{m_k\times m_1}(p^{k-1}\mathbb{Z}_p) & \cdots & \cdots & \mathrm{GL}_{m_k}(\mathbb{Z}_p) & \\ \vdots & \vdots & & & \vdots & \ddots \end{bmatrix}.$$

This set is also open and compact, and has measure

$$\mu\big(C_\Sigma\big) = p^{-D(\Sigma)} \prod_{k=0}^{\infty} \pi_{m_k}\,, \tag{3.3}$$

where $D(\Sigma) = \sum_{0\le i<j} m_i m_j (j-i) = \sum_{1\le i<j\le n} (k_j - k_i)$.

Proposition 3.2.2. *Let X be a random $n \times n$ matrix with independent entries distributed according to the Haar measure on $\mathbb{Z}_p$. Let $\Sigma = diag\{p^{k_1}, \dots, p^{k_n}\}$, and $m_k = \#\{k_i : k_i = k\}$. Then*

$$\mathbb{P}\{X \in K\Sigma K\} = \frac{\pi_n^2}{\pi_{m_0} \cdots \pi_{m_k} \cdots} |\det \Sigma|^n p^{D(\Sigma)}\,.$$

Proof. The group $K = \mathrm{GL}_n(\mathbb{Z}_p)$ is a subgroup of the group $\tilde{\mathrm{M}}_n(\mathbb{Z}_p)$, so $\tilde{\mathrm{M}}_n(\mathbb{Z}_p)$ is a disjoint union of double K-cosets. For each such coset we have

$$K\Sigma K = \bigcup_{\gamma\in K} \gamma\Sigma K = \bigsqcup_{\gamma\Sigma K \in K\Sigma K/K} \gamma\Sigma K\,.$$

The number of left cosets of K equals $[K : C_\Sigma]$ with $C_\Sigma = K \cap \Sigma K \Sigma^{-1}$, not necessarily finite. Let F_Σ be a fundamental domain for right multiplication by C_Σ, then

$$K = \bigsqcup_{\gamma\in F_\Sigma} \gamma C_\Sigma\,.$$

Sets γC_Σ are open and compact, so $\bigsqcup_{\gamma\in F_\Sigma} \gamma C_\Sigma$ is a disjoint open-compact cover of the open-compact set K. Then this cover has a finite subcover, and it coincides with the cover itself. Using that F_Σ is finite and $\gamma \in \mathrm{GL}_n(\mathbb{Z}_p)$ are isometries we get

$$\mu(K) = \sum_{\gamma\in F_\Sigma} \mu(\gamma C_\Sigma) = \sum_{\gamma\in F_\Sigma} \mu(C_\Sigma) = |F_\Sigma|\ \mu(C_\Sigma)\,.$$

Thus $|F_\Sigma| = [K : C_\Sigma] = [K : K \cap \Sigma K \Sigma^{-1}] = \mu(K)/\mu(C_\Sigma)$ and

$$\mu(K\Sigma K) = \sum_{\gamma\Sigma K \in K\Sigma K/K} \mu(\gamma\Sigma K) = \sum_{\gamma\Sigma K \in K\Sigma K/K} \mu(\Sigma K) = |F_\Sigma|\ \mu(\Sigma K) = \frac{\mu(K)\mu(\Sigma K)}{\mu(C_\Sigma)}\,.$$

Finally, for any measurable subset $\mathcal{A} \subseteq \mathrm{M}_n(\mathbb{Q}_p)$ with finite measure and $\gamma \in \mathrm{M}_n(\mathbb{Q}_p)$ we have $\mu(\gamma\mathcal{A}) = \mu(\mathcal{A})|\det(\gamma)|^n$. This fact together with (3.3) implies the statement of the proposition. □

The proof easily extends to rectangular matrices. Let $n > m$ and let $K_{n,m} \subseteq \mathrm{M}_{n\times m}(\mathbb{Z}_p)$ denote the set of $n \times m$ matrices consisting of m linearly independent mod p columns. Then for $X \in \mathrm{M}_{n\times m}(\mathbb{Z}_p)$ there exist $U \in K_{n,m}$ and $V \in K_m = \mathrm{GL}_m(\mathbb{Z}_p)$ such that $X = U\Sigma V$ with $\Sigma = \mathrm{diag}(p^{\mathbf{k}})$. In this case

$$\mu(K_{n,m}\Sigma K_m) = \frac{\mu(K_{n,m})\mu(\Sigma K_m)}{\mu(C_\Sigma)} = \frac{\pi_n\pi_m}{\pi_{n-m}\pi_{m_0}\pi_{m_1}\cdots\pi_{m_k}\cdots} p^{D(\Sigma)}|\det\Sigma|^m\,.$$

Moreover, the argument works for any p-adic number field with $\mathbb{Z}_p$ replaced by the corresponding ring of integers and $\mathbb{F}_p$ replaced by the corresponding residue field. This has a nice immediate consequence.

Corollary 3.2.3. *Let $n, m \in \mathbb{N}$ with $n \le m$. Let X be a random $n \times m$ matrix with independent entries distributed uniformly on $\mathbb{F}_q$, then*

$$\mathbb{P}\{rk\ X = n - r\} = q^{-r(m-n+r)} \frac{\pi_n(q)\pi_m(q)}{\pi_r(q)\pi_{m-n+r}(q)\pi_{n-r}(q)}\,,$$

where $\pi_j(q) = \prod_{i=1}^{j}(1 - q^{-i})$. In particular, when $n = m$,

$$\mathbb{P}\{rk\ X = n - r\} = q^{-r^2} \frac{\pi_n(q)^2}{\pi_r(q)^2\pi_{n-r}(q)}\,.$$

The proof of Proposition 3.2.2 relies on the invariance of μ under multiplication by $\gamma \in \mathrm{GL}_n(\mathbb{Z}_p)$. The Haar measure on $p^N\mathbb{Z}_p$ is the only entry-wise measure with this property, just like the Gaussian measure over $\mathbb{R}$ [53]. However it is possible to define a measure on the whole group $\mathrm{M}_n(\mathbb{Z}_p)$ with this property e.g. the Haar measure on $\mathrm{M}_n(\mathbb{Z}_p)$ with $d_H X = \prod \mu(dx_{ij})/|\det X|^n$, or simply define a measure that only depends on Σ, see Fulman and Kaplan [59].

Macdonald [97] showed that for the (multiplicative) Haar measure μ_H on $\mathrm{M}_n(\mathbb{Z}_p)$ with $\mu_H(K) = 1$

$$\mu_H(K\Sigma K) = p^{-\sum_i (n-i)k_i} P_{\mathbf{k}}(p^{n-1}, p^{n-2}, \ldots, 1; p^{-1})\,,$$

where $P_\lambda(x_1, \ldots, x_n; t)$ is the Hall-Littlewood polynomial defined by

$$P_\lambda(x_1, \ldots, x_n; t) = \frac{(1-t)^n}{\prod_{i=0}^{\infty} \prod_{j=1}^{m_i(\mathbf{k})}(1-t^j)} \sum_{\sigma \in S_n} \sigma\left(\mathbf{x}^{\mathbf{k}} \prod_{i<j} \frac{x_i - t x_j}{x_i - x_j}\right).$$

We have $\mu(K\Sigma K) = \pi_n |\det \Sigma|^n \mu_H(K\Sigma K)$, and it is also easy to see from the definition above that $A^{-|\lambda|} P_\lambda(x_1, \ldots, x_n; t) = P_\lambda(x_1/A, \ldots, x_n/A; t)$. Then

$$\mu(K\Sigma K) = \pi_n \, p^{-\sum_i (n-i)k_i} P_{\mathbf{k}}(p^{-1}, p^{-2}, \ldots, p^{-n}; p^{-1}).$$

We find it helpful to think of the joint distribution both in terms of the measure of the quotient, and in terms of symmetric functions. Both of these approaches are mirrored in Random Matrix Theory. The first one corresponds to computing the Jacobian of the change of variables from the entries of the matrix to the parameters of the decomposition. The second one is essentially the Representation Theory approach, which is more recent, but has led to many deep insights in Random Matrix Theory. For reference, see work of Diaconis and Shahshahani [49], and the exposition by Gamburd [62].

3.2.2 Quadratic forms

From now on, we restrict to odd primes p.

A quadratic form over $\mathbb{Z}_p$ in n variables is a homogeneous multivariate polynomial of degree 2, i.e.

$$Q(x_1, \ldots, x_n) = \sum_{1 \le i \le j \le n} q_{ij} x_i x_j, \quad q_{ij} \in \mathbb{Z}_p.$$

Let $S_n(\mathbb{Z}_p)$ be the set of symmetric $n \times n$ matrices over $\mathbb{Z}_p$. This is an $n(n+1)/2$-dimensional module over $\mathbb{Z}_p$ isomorphic to the module of quadratic forms over $\mathbb{Z}_p$ in n variables. We will use symmetric matrices and quadratic forms interchangeably. Due to homogeneity it does not matter if we consider quadratic forms over $\mathbb{Z}_p$ or $\mathbb{Q}_p$, but the former is more convenient.

We will use the following notation throughout the paper. Let $r \in \mathbb{Z}_p^\times$ be a fixed non-square in $\mathbb{Z}_p$; it is also a non-square mod p. Define

$$\chi(x) := \begin{cases} +1, & \text{if } x \in \mathbb{Z}_p^\times \cap (\mathbb{Z}_p)^2; \\ -1, & \text{if } x \in \mathbb{Z}_p^\times \setminus (\mathbb{Z}_p)^2; \\ 0, & \text{if } x \in p\mathbb{Z}_p. \end{cases}$$

In particular, $\chi(1) = +1$ and $\chi(r) = -1$. Sometimes we will write $\chi(1) = +$ and $\chi(r) = -$ to match the notation of $\alpha_n^\pm$. Denote $\chi(-1)$ by ε.

Let $\langle a, b\rangle$ denote the Hilbert symbol

$$\langle a, b\rangle = \begin{cases} 1\,, & ax^2 + by^2 = z^2 \text{ has a non-trivial solution over } \mathbb{F}_p\,; \\ 0\,, & \text{otherwise}\,. \end{cases}$$

A non-singular p-adic quadratic form Q has two invariants: the discriminant

$$d(Q) \in \mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^2 \cong \{1, r, p, pr\}\,,$$

and the Hasse invariant

$$c(Q) \in \{\pm 1\}\,.$$

If $Q(\mathbf{x}) = a_1x_1^2 + \ldots + a_nx_n^2$, then

$$d(Q) = \prod_{i=1}^{n} a_i\,, \quad c(Q) = \prod_{1\le i<j\le n} \langle a_i, a_j\rangle\ .$$

These two invariants completely define the equivalence class of Q with respect to a non-degenerate change of variables. Translating this into the matrix language, for any matrix $Q \in S_n(\mathbb{Q}_p)$ there exists a matrix $U \in \mathrm{GL}_n(\mathbb{Q}_p)$ such that $Q = UDU^\top$, where D is the representative of the equivalence class defined by $d(Q)$ and $c(Q)$. There are $2|\mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^2| = 8$ classes altogether.

Note that this is very different from the spectral decomposition of a real symmetric matrix, where the diagonal entries of D are the eigenvalues of the matrix, and U is the unitary matrix of eigenvectors. It is possible to consider a different decomposition which would be analogous to the spectral decomposition (for example, see [57]), but this is not the subject of this paper.

If we restrict to $U \in \mathrm{GL}_n(\mathbb{Z}_p)$, we have to account for scaling by powers of p. For any matrix $Q \in S_n(\mathbb{Z}_p)$ there exists $U \in \mathrm{GL}_n(\mathbb{Z}_p)$ such that $Q = U\Sigma SU^\top$, where Σ is the matrix of elementary divisors, and $S = S(\Sigma)$ is the matrix of signatures. Each diagonal block in Σ corresponding to p^k with multiplicity m_k has its own quadratic class 1 or r, which is reflected in the corresponding diagonal block of S.

The canonical form for a block of S with signature $s_k = +$ meaning quadratic class 1 is the $m_k \times m_k$ identity matrix $\mathbf{1}_{m_k}^+$. For a block with signature $s_k = -$ meaning quadratic class r it is the matrix $\mathbf{1}_{m_k}^-$, the $m_k \times m_k$ identity matrix with last entry replaced by r. If $m_k = 0$, set $s_k = +$. This is exactly the same as the canonical form for quadratic forms, for more detail see Cassels' book [29]. Further we will write $K\Sigma SK^\top$ for the $\mathbb{Z}_p$-equivalence class of ΣS in $S_n(\mathbb{Z}_p)$.

3.3 Distribution of equivalence classes of symmetric matrices

In this section we consider symmetric matrices with independent entries distributed according to the Haar measure on $\mathbb{Z}_p$ for odd p; denote this measure on $S_n(\mathbb{Z}_p)$ by μ. We derive the distribution of equivalence classes with respect to conjugation by $\mathrm{GL}_n(\mathbb{Z}_p)$ and $\mathrm{GL}_n(\mathbb{Q}_p)$. It is possible to consider $S_n(\mathbb{Z}_2)$ and obtain a similar answer with the principal difference coming from the canonical form and the orthogonal group over $\mathbb{F}_2$.

3.3.1 $\mathbb{Z}_p$-equivalence

First, let us show that we can reduce the problem over $\mathbb{Z}_p$ to computing $K\Sigma S K^\top$ mod p^k for some large k. For convenience set

$$R_k := \mathbb{Z}/p^k\mathbb{Z}\,.$$

If two matrices are equivalent over $\mathbb{Z}_p$, their reductions are equivalent over R_k for all $k \in \mathbb{N}$. The converse is only true for k large enough in terms of Σ.

Lemma 3.3.1 (Chapter 8, Lemma 5.1 [29])**.** *Let $A, B \in S_n(\mathbb{Z}_p)$ such that $|\det A|_p = |\det B|_p = p^{-m}$. Then $A \sim_{GL_n(R_k)} B \bmod p^k$ if and only if $A \sim_{GL_n(\mathbb{Z}_p)} B$ for all $k \geq m+1$.*

Let

$$C_R(A) := \{\gamma \in \mathrm{GL}_n(R) : \gamma A \gamma^\top = A\} \subseteq \mathrm{GL}_n(R)$$

denote the stabiliser of $A \in \mathrm{M}_n(R)$, and let

$$\mathrm{O}_n^+(R) := C_R(\mathbf{1}_n^+)$$

denote the orthogonal group over a ring R. We only consider $R = \mathbb{Z}_p$ or $R_k = \mathbb{Z}/p^k\mathbb{Z}$, so R always contains a non-square, and we can define the twin brother of the orthogonal group $\mathrm{O}_n^-(R) = C_R(\mathbf{1}_n^-)$. We will call both of them orthogonal groups. If $\mathbb{F}_p$ is a finite field with p odd, $\varepsilon = \chi(-1)$ and $s \in \{\pm\}$, then

$$|\mathrm{O}_n^s(\mathbb{F}_p)| = \begin{cases} 2p^{n(n-1)/2}\prod_{i=1}^t (1-p^{-2i})\,, & \text{for } n = 2t+1\,; \\ \\ \frac{2p^{n(n-1)/2}}{1+s\varepsilon^t p^{-t}}\prod_{i=1}^t (1-p^{-2i})\,, & \text{for } n = 2t\,. \end{cases}$$

The size of the orthogonal groups over $\mathbb{Z}/p^k\mathbb{Z}$ can be obtained via multivariate Hensel's lemma, and in particular, their "densities"

$$\alpha_n^s := |\mathrm{O}_n^s(\mathbb{F}_p)|/p^{n(n-1)/2}$$

are preserved.

Lemma 3.3.2 (Multivariate Hensel's lemma for systems [55])**.** *Let* $\mathbf{f} = (f_1, \ldots, f_m) \in \mathbb{Z}_p[x_1, \ldots, x_n]^m$ *with* $n \geq m$*, and let* $\mathbf{J}_\mathbf{f} = (\partial f_i/\partial x_j)$ *be the Jacobian of the system* $\mathbf{f} = \mathbf{0}$*. Let* $\mathbf{t} \in \mathbb{Z}_p^n$ *satisfy*

$$\|\mathbf{f}(\mathbf{t})\|_p < 1\,, \text{ and } \max_{S\subseteq[n],|S|=m} |\det \mathbf{J}_{\mathbf{f},S}(\mathbf{t})|_p = 1\,,$$

where $\mathbf{J}_{\mathbf{f},S}(\mathbf{t})$ *is the restriction of the Jacobian matrix to a subset of columns* S *of size* m*. With variables outside of* S *fixed, there exists a unique* $\mathbf{y} \in \mathbb{Z}_p^n$ *such that* $\mathbf{f}(\mathbf{y}) = 0$ *with* $\|\mathbf{y} - \mathbf{t}\|_p < 1$*.*

Lemma 3.3.3. *Let* $n, k \in \mathbb{N}$ *and* $s \in \{\pm\}$*. Then* $p^{-kn(n-1)/2}|O_n^s(R_k)| = \alpha_n^s$*.*

Proof. Without loss of generality, set $s = +$ and write $\mathbf{1}$ for $\mathbf{1}_n^+$ to simplify notation in this lemma. If $U \in \mathrm{O}_n^+(R_k)$, then U is a solution to the equation $XX^\top = \mathbf{1}$.

First, we will need to prove that each solution of $XX^\top - \mathbf{1} = 0 \bmod p$ has exactly $p^{(k-1)n(n-1)/2}$ lifts to R_k. Let $X = (x_{ij})$, then $XX^\top = \mathbf{1}$ is a system of $m = n(n+1)/2$ equations in n^2 variables. Enumerate the variables along the rows so that

$$y_1 = x_{11}, \ \ldots\ , y_n = x_{1n}, x_{n+1} = x_{21}, \ldots, y_{2n} = x_{2n}, \ \ldots\ , y_{n^2} = x_{nn}\,.$$

Enumerate the equations along the rows starting from the diagonal so that

$$f_1 = (XX^\top - \mathbf{1})_{11}, \ \ldots\ , f_n = (XX^\top - \mathbf{1})_{1n}, f_{n(n+1)/2} = (XX^\top - \mathbf{1})_{nn}\,.$$

Let A be a solution to $XX^\top = \mathbf{1} \bmod p$. The condition $\max_S |\det \mathbf{J}_{\mathbf{f},S}(\mathbf{t})|_p = 1$ is equivalent to $\mathrm{rk}_p \mathbf{J}_\mathbf{f}(\mathbf{t}) := \mathrm{rk}\ \mathbf{J}_\mathbf{f}(\mathbf{t}) \bmod p = m$, or that all m rows of the Jacobian matrix at A are linearly independent mod p.

Let $a_1, \ldots, a_n$ be the rows of A. Then $\mathbf{J}_\mathbf{f}(A)$ is block-upper-triangular. Let $A \setminus a_1 \ldots a_j$ be the truncated A without first j rows. As $p \neq 2$, $\mathrm{rk}_p \mathbf{J}_\mathbf{f}(\mathbf{t}) = \mathrm{rk}_p B$ for

$$B = \begin{bmatrix} A & & & & & \\ 0 & A \setminus a_1 & & & & \\ 0 & 0 & A \setminus a_1, a_2 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ 0 & 0 & 0 & \cdots & 0 & a_n \end{bmatrix}.$$

Let b_j denote the rows of B. Suppose $\sum_j \gamma_j b_j = 0$ for some $\gamma_j \in \mathbb{F}_p$. Looking only at the first n columns this implies $\sum_{j\leq n} \gamma_j a_j = 0$, but the rows of A are linearly

independent mod p, hence $\gamma_j = 0$ for $j \le n$. Repeating the argument for the next n columns and so on conclude that $\gamma_j = 0$ for all j. Hence rows of B are linearly independent and $\operatorname{rk}_p \mathbf{J_f}(\mathbf{t}) = \operatorname{rk}_p B = m$.

For $A \in \mathrm{O}_n^s(\mathbb{F}_p)$ there is a subset of columns S of $\mathbf{J_f}(A)$ such that these columns are independent mod p. This corresponds to choosing $|S| = m$ variables. Lift all the variables outside of S arbitrarily with $p^{(k-1)n(n-1)/2}$ options. The m variables we chose have a unique lift by Hensel's lemma. □

Theorem 3.3.4. *Let p be an odd prime, and let $X \in S_n(\mathbb{Z}_p)$ be a random symmetric $n \times n$ matrix with independent entries distributed according to the Haar measure on $\mathbb{Z}_p$. Let $\Sigma = diag(p^{k_1}, \dots, p^{k_n})$ be a matrix of elementary divisors with multiplicities $m_k = \#\{i : k_i = k\}$, and $S = diag(\mathbf{1}_{m_0}^{s_0}, \dots, \mathbf{1}_{m_k}^{s_k}, \dots)$ be a matrix of signatures, then the probability that X is in the (Σ, S)-class equals*

$$\mu(K\Sigma SK^\top) = \frac{\pi_n}{\alpha_{m_0}^{s_0} \cdots \alpha_{m_l}^{s_l} \cdots} \prod_{j=1}^{n} p^{-k_j(n-j+1)} .$$

Proof. Let $\Sigma = \operatorname{diag}(p^{k_1}, \dots, p^{k_n})$ with multiplicities $m_k = \#\{i : k_i = k\}$. Let S be the matrix of signatures with $s_0, s_1, \dots$ corresponding to blocks of size m_0, m_1 and so on. Let $|\det \Sigma|_p = p^{-M}$ and $k_n = t$. Then by Lemma 3.3.1 for $m \ge M + 1$

$$\mu(K\Sigma SK^\top) = \frac{\#\{\gamma \in S_n(R_m) :\ \gamma \in K\Sigma SK^\top \bmod p^m\}}{p^{mn(n+1)/2}} .$$

By the Orbit-Stabiliser Theorem,

$$\#\{\gamma \in \mathrm{M}_n(R_m) :\ \gamma \in K\Sigma SK^\top \bmod p^m\} = \frac{|\mathrm{GL}_n(R_m)|}{|C_{R_m}(\Sigma S)|} .$$

We have $C_{\mathbb{Z}_p}(\Sigma S) \subseteq K \cap \Sigma K \Sigma^{-1}$ and $C_{R_m}(\Sigma S) \subseteq (K \cap \Sigma K \Sigma^{-1}) \bmod p^m$. Consider the equation $U\Sigma SU^\top = \Sigma S$. Divide U into blocks according to m_k so that

$$U = \begin{bmatrix} U_{0,0} & U_{0,1} & U_{0,2} & \cdots & U_{0,t} \\ pU_{1,0} & U_{1,1} & U_{1,2} & \cdots & U_{1,t} \\ p^2U_{2,0} & pU_{2,1} & U_{2,2} & \cdots & U_{2,t} \\ \vdots & \vdots & & \ddots & \vdots \\ p^tU_{t,0} & p^{t-1}U_{t,1} & \cdots & pU_{t,t-1} & U_{t,t} \end{bmatrix}$$

with $U_{i,j} \in M_{m_i \times m_j}(\mathbb{Z}_p)$, $U_{i,i} \in \mathrm{GL}_{m_i}(\mathbb{Z}_p)$ and $S_{ii} = \mathbf{1}_{m_i}^{s_i}$. Rewrite $U\Sigma SU^\top = \Sigma S$ block-wise for a system of matrix equations with (i, j)-th block satisfying

$$\sum_{k=i}^{j} p^j U_{i,k} S_{kk} U_{j,k}^\top + \sum_{k=0}^{i-1} p^{i+j-k} U_{i,k} S_{kk} U_{j,k}^\top + \sum_{k=j+1}^{\infty} p^k U_{i,k} S_{kk} U_{j,k}^\top = \mathbf{0} \bmod p^m .$$

Divide out (i,j)-th block by p^j and separate the diagonal and off-diagonal conditions. The diagonal conditions are

$$U_{ii}S_{ii}U_{ii}^\top + \sum_{k=0}^{i-1} p^{i-k}U_{ik}S_{kk}U_{ik}^\top + \sum_{k=i+1}^{\infty} p^{k-i}U_{ik}S_{kk}U_{ik}^\top = S_{ii} \mod p^{m-i},\ i \le t\,,$$

and the off-diagonal are

$$\sum_{k=i}^{j} U_{ik}S_{kk}U_{jk}^\top + \sum_{k=0}^{i-1} p^{i-k}U_{ik}S_{kk}U_{jk}^\top + \sum_{k=j+1}^{\infty} p^{k-j}U_{ik}S_{kk}U_{jk}^\top = \mathbf{0} \mod p^{m-j},\ i < j \le t\,.$$

Solve this system mod p, where it reduces to

$$\begin{cases} U_{ii}S_{ii}U_{ii}^\top = S_{ii} & \text{for } i \le t\,; \\ \\ \sum\limits_{k=i}^{j} U_{ik}S_{kk}U_{jk}^\top = \mathbf{0} & \text{for } i < j \le t\,, \end{cases}$$

and lift the solutions to R_m as follows.

Fix the lower blocks with $i > j$ arbitrarily. Solve the diagonal equations with $U_{ii} \in \mathrm{O}^{s_i}_{m_i}(\mathbb{F}_p)$, then solve the remaining linear system for $j = i+1, \dots, t$ consecutively ("diagonal by diagonal"). The blocks U_{ik} with $i < k$ are determined uniquely from U_{ik} with $i \ge k$.

It now remains to show that this system is "well-defined" i.e. its Jacobian matrix is full-rank. To avoid writing out each element-wise derivative separately, use matrix derivatives. The Jacobian of the system mod p consists of the following matrix derivatives:

$$\begin{aligned} B^{ii}_{ii} &= \frac{\partial(U_{ii}S_{ii}U_{ii}^\top)}{\partial U_{ii}} = E\left((U_{ii}S_{ii} \otimes \mathbf{1}^+_{m_i})K + (\mathbf{1}^+_{m_i} \otimes U_{ii}S_{ii})\right) && \text{for } i \le t\,; \\ B^{ij}_{jj} &= \frac{\partial(\sum_{k=i}^{j} U_{ik}S_{kk}U_{jk}^\top)}{\partial U_{jj}} = \frac{\partial(U_{ij}S_{jj}U_{jj}^\top)}{\partial U_{jj}} = U_{ij}S_{jj} \otimes \mathbf{1}^+_{m_j} && \text{for } i < j \le t\,; \\ B^{ij}_{ik} &= \frac{\partial(\sum_{k=i}^{j} U_{ik}S_{kk}U_{jk}^\top)}{\partial U_{ik}} = \frac{\partial(U_{ik}S_{kk}U_{jk}^\top)}{\partial U_{ik}} = S_{kk}U_{jk}^\top \otimes \mathbf{1}^+_{m_i} && \text{for } k = i \dots j\,. \end{aligned}$$

Here $\otimes$ is the tensor product, E and K are the elimination and the commutation matrix. The diagonal derivatives B^{ii}_{ii} are as in Lemma 3.3.3. The Jacobian matrix mod p is block-lower-triangular. On the diagonal blocks B^{ii}_{ii} for $i = 0, \dots, t$ come first, then B^{ij}_{ij} for $i = 0, \dots, t$ and $j = i+1, \dots, t$. This corresponds to how we solved the system mod p. Blocks $B^{ij}_{ij} = S_{jj}U_{jj}^\top \otimes \mathbf{1}^+_{m_i}$ are full-rank as $\det B^{ij}_{ij} = (\det S_{jj}U_{jj})^{m_i} \in \mathbb{Z}_p^\times$, and so are B^{ii}_{ii} by Lemma 3.3.3. Thus the Jacobian has full row rank.

Blocks U_{ii} have $|\mathrm{O}^{s_i}_{m_i}(\mathbb{Z}/p^{m-i}\mathbb{Z})|p^{im_i^2} = \alpha^{s_i}_{m_i}(p^{m-i})^{m_i(m_i-1)/2}p^{im_i^2}$ lifts mod p^{m-i}. For $i > j$ blocks $U_{ij} \in \mathrm{M}_{m_i\times m_j}(p^m\mathbb{Z}/p^{i-j}\mathbb{Z})$ with $p^{(m-(i-j))m_im_j}$ options mod $p^{m-(i-j)}$, and the upper blocks are determined uniquely mod p^{m-j} by Hensel's lemma. Hence

$$|C_{R_m}(\Sigma S)| = \prod_{i=0}^{\infty}\alpha^{s_i}_{m_i}p^{(m-i)m_i(m_i-1)/2+im_i^2}\prod_{i<j}p^{(m-(j-i))m_im_j+jm_im_j} = p^{mn(n-1)/2}\prod_{i=0}^{\infty}\alpha^{s_i}_{m_i}\prod_{j=1}^{n}p^{k_j(n-j+1)}.$$

This then implies the statement of the theorem. □

Corollary 3.3.5. *Let $\Sigma = diag(p^{k_1},\dots,p^{k_n})$ be a matrix of elementary divisors with multiplicities $m_k = \#\{i : k_i = k\}$, then*

$$\mathbb{P}\{Q \in K\Sigma K\} = \mu(K\Sigma K \cap S_n(\mathbb{Z}_p)) = \frac{\pi_n}{\beta_{m_0}\cdots\beta_{m_k}\cdots}\prod_{i=1}^{n}p^{-k_j(n-j+1)},$$

where $(\beta_t)^{-1} = (\alpha_t^+)^{-1} + (\alpha_t^-)^{-1} = \prod_{j=1}^{\lfloor\frac{t}{2}\rfloor}(1-p^{-2j})$.

Now let us give a brief discussion regarding the connection between the distributions in question and symmetric functions. Recall that the density of elementary divisors (3.1) of general random matrices over $\mathbb{Z}_p$ can be written in terms of symmetric functions. Here we similarly have

$$\mu(K\Sigma SK^\top) = \frac{\pi_{m_0}Q_{\mathbf{k}}(p^{-\mathbf{n}};p^{-1})}{\alpha^{s_0}_{m_0}\cdots\alpha^{s_k}_{m_k}\cdots},$$

where $p^{-\mathbf{n}} = (p^{-1},p^{-2},\dots,p^{-n})$ and $Q_{\mathbf{k}}(\mathbf{x};t)$ is the Hall-Littlewood Q polynomial [97]

$$Q_{\mathbf{k}}(x_1,\dots,x_n;t) = \frac{(1-t)^n}{\pi_{m_0}}\sum_{\sigma\in S_n}\sigma\left(\mathbf{x}^{\mathbf{k}}\prod_{i<j}\frac{x_i-tx_j}{x_i-x_j}\right);$$
$$Q_{\mathbf{k}}(p^{-\mathbf{n}};p^{-1}) = \frac{\pi_n}{\pi_{m_0}}p^{-\sum_i(n-i+1)k_i}.$$

Let $d_{\mathbf{k}}(t) := \prod_{i\geq 1}\prod_{j=1}^{\lfloor\frac{m_i}{2}\rfloor}(1-t^{2j})$, and rewrite the density in Corollary 3.3.5 as

$$\mu\left(K\Sigma K\right) = \frac{\pi_{m_0}}{\beta_{m_0}}\frac{Q_{\mathbf{k}}(p^{-\mathbf{n}};p^{-1})}{d_{\mathbf{k}}(p^{-1})} = \mu(S_{m_0}(\mathbb{Z}_p)\cap\mathrm{GL}_{m_0}(\mathbb{Z}_p))\frac{Q_{\mathbf{k}}(p^{-\mathbf{n}};p^{-1})}{d_{\mathbf{k}}(p^{-1})}.$$

This form is helpful when considering the limit $n\to\infty$.

Let $\lambda = (\lambda_1,\dots,\lambda_l,0,\dots)$ be a partition of length $\ell(\lambda) = l$ with $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_l > 0$ and $m_j(\lambda) = \#\{i : \lambda_i = j\}$. Partitions are defined up to an arbitrary string

of zeros at the end. Let $f_n(\lambda)$ be the probability that the exponents of elementary divisors are equivalent to λ as a partition, then

$$f_n(\lambda) = \begin{cases} \frac{\pi_n}{\beta_{n-\ell(\lambda)} d_\lambda(p^{-1})} \prod_{i=1}^{\ell(\lambda)} p^{-i\lambda_i} = \frac{\pi_{n-\ell(\lambda)}}{\beta_{n-\ell(\lambda)}} \frac{Q_\lambda(p^{-\mathbf{n}}; p^{-1})}{d_\lambda(p^{-1})}, & \ell(\lambda) \leq n\,; \\ 0, & \text{otherwise}\,. \end{cases} \tag{3.4}$$

We have the following statement for such a function.

Proposition 3.3.6. *Let $f_n(\lambda)$ be the distribution on partitions induced by the elementary divisors. Then the limiting distribution on partitions as $n \to \infty$ exists and is equal to*

$$f(\lambda) := \lim_{n\to\infty} f_n(\lambda) = \frac{\pi_\infty}{\beta_\infty d_\lambda(p^{-1})} p^{-\sum i\lambda_i}\,.$$

Proof. It is clear that the pointwise limit of $f_n(\lambda)$ exists and is positive, so it remains to check $\sum_\lambda f(\lambda) = 1$. Function $f_n(\lambda)$ is bounded from above by an integrable function

$$f_n(\lambda) \leq \frac{1}{\beta_\infty d_\lambda(1/p)} p^{-\sum_i i\lambda_i} \leq \frac{1}{\beta_\infty (1-1/p)^{\ell(\lambda)}} p^{-\sum_i i\lambda_i} =: g(\lambda)\,.$$

and

$$\sum_\lambda g(\lambda) \ll \sum_{l=0}^{\infty} (1-1/p)^{-l} \sum_{\lambda:\ell(\lambda)=l} \prod_{i=1}^{l} p^{-i\lambda_i} \ll \sum_{l=0}^{\infty} (1-1/p)^{-l} p^{-\frac{l(l+1)}{2}} < +\infty\,.$$

By dominated convergence $\sum_\lambda f(\lambda) = \lim_{n\to\infty} \sum_\lambda f_n(\lambda) = 1$, thus $f(\lambda)$ defines a distribution. □

Note that $f_n(\lambda)$ is a rational function in $t = p^{-1}$, so $\sum_\lambda f_n(\lambda) = 1$ is a rational identity for all $t \in (0,1)$, and more generally can be seen as a formal identity. The proof by dominated convergence works for all $t \in (0,1)$ and can be interpreted as a probabilistic proof of the identity

$$\sum_\lambda \frac{Q_\lambda(t, t^2, \ldots; t)}{d_\lambda(t)} = \prod_{j=0}^{\infty} (1 - t^{2j+1})^{-1}\,.$$

We now move on to the equivalence classes of quadratic forms with respect to conjugation by a matrix from $\mathrm{GL}_n(\mathbb{Q}_p)$.

3.3.2 $\mathbb{Q}_p$-equivalence

Let p be an odd prime. Let Q be a random quadratic form in n variables with independent coefficients distributed according to the Haar measure on $\mathbb{Z}_p$. This corresponds to the distribution on $S_n(\mathbb{Z}_p)$ we considered earlier. Let $a \in \mathbb{Z}_p/(\mathbb{Z}_p)^2$ and $b \in \{\pm 1\}$, and denote the probability that Q has invariants $d(Q) = a$ and $c(Q) = b$ by $\rho_n(a, b)$ for $n \in \mathbb{N}$. When $n = 0$ set $\rho_0(a, b) = \mathbf{1}\{a = 1, b = 1\}$.

Recall that

$$\chi(x) = \begin{cases} +1\,, & \text{if } x \in \mathbb{Z}_p^\times \cap (\mathbb{Z}_p)^2\,; \\ -1\,, & \text{if } x \in \mathbb{Z}_p^\times \setminus (\mathbb{Z}_p)^2\,; \\ 0\,, & \text{if } x \in p\mathbb{Z}_p\,. \end{cases}$$

In particular, $\chi(1) = +1$ and $\chi(r) = -1$, where $r \in \mathbb{Z}_p^\times$ is a fixed non-square in $\mathbb{Z}_p$. We will write $\chi(1) = +$ and $\chi(r) = -$ to match the notation of $\alpha_n^\pm$. Denote $\chi(-1)$ by ε.

Condition on m_0 and s_0 to obtain the recursive expression for ρ_n:

$$\rho_n(a, b) = \sum_{\substack{l=0 \\ d \in \{1,r\}}}^{n} \mathbb{P}_n\{m_0 = n - l, s_0 = \chi(d)\}\rho_l(a'(l, d), b'(l, d))\,,$$

where $a'(l, d) = adp^l$ and $b'(l, d) = b\varepsilon^{\frac{l(l-1)}{2}} \langle a, d\rangle \langle p, ad\rangle^{l-1}$. This corresponds to separating out the non-zero mod p part, factoring out the rest by p and inferring the invariants of the remaining quadratic form in $n - m_0$ variables. By Theorem 3.3.4

$$\mathbb{P}_n\{m_0 = n - l, s_0 = s\} = p^{-\frac{l(l+1)}{2}} \frac{\pi_n}{\pi_l \alpha_{n-l}^s}\,,$$

hence

$$\begin{aligned} \rho_n(a, b) = p^{-\frac{n(n+1)}{2}} \rho_n(a'(n, 1), b'(n, 1)) + \sum_{l=0}^{n-1} p^{-\frac{l(l+1)}{2}} \frac{\pi_n}{\pi_l \alpha_{n-l}^+} \rho_l(a'(l, 1), b'(l, 1)) + \\ \sum_{l=0}^{n-1} p^{-\frac{l(l+1)}{2}} \frac{\pi_n}{\pi_l \alpha_{n-l}^-} \rho_l(a'(l, r), b'(l, r))\,. \end{aligned} \tag{3.5}$$

First, we will compute

$$\begin{aligned} \sigma_n(a) &:= \rho_n(a, 1) + \rho_n(a, -1) = \mathbb{P}_n\{d(Q) = a\}\,, \\ \Delta_n(a) &:= \rho_n(a, 1) - \rho_n(a, -1)\,. \end{aligned}$$

Then $\rho_n(a, b) = (\sigma_n(a) + b\Delta_n(a))/2$.

We will also need the following analogue of Corollary 3.2.3 for alternating matrices. Alternating, or skew-symmetric, matrices are matrices satisfying $A^\top = -A$. In particular, this implies that they have zero diagonal.

Lemma 3.3.7 (Carlitz [28]). *Let $\mathbb{F}_q$ be a finite field of odd characteristic. Let A be a random alternating $n \times n$ matrix with independent entries distributed uniformly on $\mathbb{F}_q$. Then for all r such that $n - r$ is even*

$$\mathbb{P}\{rk\ A = n - r\} = q^{-r(r-1)/2} \frac{\pi_n(q)}{\pi_r(q)\beta_{n-r}(q)}\,.$$

Let us start with the distribution of the discriminant $\sigma_n(a)$.

Lemma 3.3.8. *Let $n \in \mathbb{Z}^{\geq 0}$, set $t = 1/p$ and $s = \chi(a)$. If n is even, we have*

$$\sigma_n(a) = \begin{cases} \frac{(1+s(\varepsilon t)^{n/2})(1-s\varepsilon^{n/2}t^{n/2+2})}{2(1+t)(1-t^{n+1})}\,, & \text{if } a \in \{1, r\}\,; \\ \\ \frac{t(1-t^n)}{2(1+t)(1-t^{n+1})}\,, & \text{if } a \in \{p, pr\}\,; \end{cases}$$

and in n is odd,

$$\sigma_n(a) = \begin{cases} \frac{1}{2(1+t)}\,, & \text{if } a \in \{1, r\}\,; \\ \\ \frac{t}{2(1+t)}\,, & \text{if } a \in \{p, pr\}\,. \end{cases}$$

Proof. First, condition on the multiplicity m_0 and the signature s_0 and write the recurrence for $\sigma_n(a)$:

$$\sigma_n(a) = \sum_{l=0}^{n-1} p^{-\frac{l(l+1)}{2}} \left(\frac{\pi_n}{\pi_l \alpha^+_{n-l}} \sigma_l(ap^l) \ + \frac{\pi_n}{\pi_l \alpha^-_{n-l}} \sigma_l(arp^l) \right) + p^{-\frac{n(n+1)}{2}} \sigma_n(ap^n)\,. \qquad (3.6)$$

To reduce the number of cases, observe that if $\mathrm{val}_p(\det Q)$ is odd, or if n is odd, then Q has an elementary divisor of odd multiplicity l. Then we can choose $s_l \in \{\pm\}$ with equal probability as $\alpha^+_l = \alpha^-_l$, hence $\sigma_n(1) = \sigma_n(p)$ for odd n, and $\sigma_n(p) = \sigma_n(rp)$ for all n.

The statement holds for $n = 0$ and is also easy to verify for small n. Suppose it holds for all $l < n$ and substitute coefficients with $l < n$ into (3.6).

For even n, and $a = 1$ or r with $s = \chi(a)$ we have

$$\sigma_n(a) = t^{\frac{n(n+1)}{2}} \sigma_n(a) + \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{t\pi_n}{2(1+t)\pi_l \beta_{n-l}} +$$

$$\sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_n}{\pi_{l+1}\beta_{n-l}} \frac{1 - t^{l+2} + s(1-t^2)(\varepsilon t)^{\frac{n}{2}}}{2(1+t)} \ =$$

$$t^{\frac{n(n+1)}{2}} \sigma_n(a) + \frac{1 - t^{n+2} + s(1-t^2)(\varepsilon t)^{\frac{n}{2}}}{2(1+t)(1-t^{n+1})} \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_{n+1}}{\pi_{l+1}\beta_{n-l}}\,.$$

Similarly for $a = p$ or pr

$$\sigma_n(a) = t^{\frac{n(n+1)}{2}}\sigma_n(a) + \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_n}{2(1+t)\pi_l\beta_{n-l}} + \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_n}{\pi_{l+1}\beta_{n-l}} \frac{t(1-t^n)}{2(1+t)} =$$
$$t^{\frac{n(n+1)}{2}}\sigma_n(a) + \frac{1-t^n}{2(1+t)(1-t^{n+1})} \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_{n+1}}{\pi_{l+1}\beta_{n-l}}\,.$$

The sum over even l equals $1 - p^{-n(n+1)/2}$ as it is the sum over the distribution of rank of $(n+1) \times (n+1)$ alternating matrices over $\mathbb{F}_p$ as in Lemma 3.3.7 without the rank zero term. This implies the statement of the lemma for even n. For odd n

$$\sigma_n(1) = t^{\frac{n(n+1)}{2}}\sigma_n(p) + \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{t\pi_n}{2(1+t)\pi_l\beta_{n-l}} + \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_n}{\pi_{l+1}\beta_{n-l}} \frac{1-t^{l+2}}{2(1+t)} =$$
$$t^{\frac{n(n+1)}{2}}\sigma_n(p) + \frac{1}{2(1+t)} \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_n}{\pi_{l+1}\beta_{n-l}} - t^{\frac{n(n+1)}{2}} \frac{t}{2(1+t)}\,;$$

$$\sigma_n(p) = t^{\frac{n(n+1)}{2}}\sigma_n(1) + \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_n}{2(1+t)\pi_l\beta_{n-l}} + \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_n}{\pi_{l+1}\beta_{n-l}} \frac{t(1-t^l)}{2(1+t)} =$$
$$t^{\frac{n(n+1)}{2}}\sigma_n(1) + \frac{t}{2(1+t)} \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_n}{\pi_{l+1}\beta_{n-l}} - t^{\frac{n(n+1)}{2}} \frac{1}{2(1+t)}\,.$$

The sum over the even indices is the sum over the distribution of rank of $n \times n$ alternating matrices over $\mathbb{F}_p$, hence equals 1. Solving this system we obtain $\sigma_n(1) = \sigma_n(r)$ and $\sigma_n(p) = \sigma_n(rp)$. □

Now let us compute the discrepancy between the distribution of the discriminant of quadratic forms with different Hasse invariants.

Lemma 3.3.9. *Let $n \in \mathbb{Z}^{\geq 0}$, and set $t = 1/p$. Then*

$$\Delta_n(a) = \begin{cases} \frac{\pi_n}{\beta_{n+1}\alpha_n^s}\,, & \text{if } a = 1 \text{ or } r, \text{ and } s = \chi(a)\,; \\ \varepsilon^{\frac{n-1}{2}} t^{\frac{n+1}{2}} \frac{\pi_n}{2\beta_{n+1}\beta_{n-1}}\,, & \text{if } a = p \text{ or } pr, \text{ and } n \text{ is odd}; \\ 0\,, & \text{if } a = p \text{ or } pr, \text{ and } n \text{ is even}. \end{cases}$$

Proof. Similarly to the previous lemma, prove the statement by induction. Condi-

tioning on the multiplicity m_0, write the recurrence for $\Delta_n(a)$ using (3.5):

$$\begin{aligned}\Delta_n(a) = \sum_{l=0}^{n} p^{-\frac{l(l+1)}{2}} \varepsilon^{\frac{l(l-1)}{2}} \frac{\pi_n}{\pi_l \alpha^+_{n-l}} \langle p, a\rangle^{l-1} \Delta_l(ap^l) + \\ + \ \langle a, r\rangle \sum_{l=0}^{n-1} p^{-\frac{l(l+1)}{2}} \varepsilon^{\frac{l(l-1)}{2}} \frac{\pi_n}{\pi_l \alpha^-_{n-l}} \langle p, ar\rangle^{l-1} \Delta_l(arp^l)\,.\end{aligned} \tag{3.7}$$

The statement holds for $n = 0$. It is also easy to check for $n \leq 2$. Suppose the statement holds for all $l < n$. Split this sum into the sum over odd and even indices, and substitute $\Delta_l(\cdot)$ for $l < n$ into (3.7).

Consider first $a = 1$ or r with $s = \chi(a)$, then

$$\begin{aligned}\Delta_n(a) - t^{\frac{n(n+1)}{2}} \varepsilon^{\frac{n(n-1)}{2}} s^{n-1} \Delta_n(ap^n) \ = \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{l(l+1)}{2}} \varepsilon^{\frac{l(l-1)}{2}} \frac{\pi_n}{\pi_l} \left(\frac{\Delta_l(ap)}{\alpha^+_{n-l}} + \frac{\Delta_l(apr)}{\alpha^-_{n-l}} \right) + \\ + \ s \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \varepsilon^{\frac{l(l-1)}{2}} \frac{\pi_n}{\pi_l} \left(\frac{\Delta_l(a)}{\alpha^+_{n-l}} - \frac{\Delta_l(ar)}{\alpha^-_{n-l}} \right).\end{aligned}$$

When n is even, we have

$$\begin{aligned}\Delta_n(a) = \Delta_n(a) s t^{\frac{n(n+1)}{2}} \varepsilon^{\frac{n(n-1)}{2}} + \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{(l+1)^2}{2}} \frac{\pi_n}{2\beta_l \beta_{l+1} \beta_{n-l}} + \\ \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \frac{\pi_n}{2\beta_l \beta_{l+1} \beta_{n-l}} \left(t^{\frac{l}{2}} + s\varepsilon^{\frac{n}{2}} t^{\frac{n-l}{2}} \right),\end{aligned}$$

hence

$$\begin{aligned}\Delta_n(a)(1 - s t^{\frac{n(n+1)}{2}} \varepsilon^{\frac{n(n-1)}{2}}) = \frac{\pi_n}{\beta_n \alpha^s_n} \left(\beta_n + \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{(l+1)^2}{2}} \frac{\beta_n^2}{\beta_{l+1}^2 \beta_{n-l-1}} - s\, t^{\frac{n(n+1)}{2}} \varepsilon^{\frac{n(n-1)}{2}} \right) = \\ \frac{\pi_n}{\beta_n \alpha^s_n} \left(\sum_{j=0}^{n/2} t^{2j^2} \frac{\beta_n^2}{\beta_{2j}^2 \beta_{n-2j}} - s\, t^{\frac{n(n+1)}{2}} \varepsilon^{\frac{n(n-1)}{2}} \right).\end{aligned}$$

The sum over j is the sum over the distribution of rank of general square $n/2 \times n/2$ matrices over $\mathbb{F}_{p^2}$, and by Corollary 3.2.3 it equals 1. Hence the statement holds for $\Delta_n(1)$ and $\Delta_n(r)$.

If n is odd,

$$\begin{aligned}\Delta_n(1) - \varepsilon^{\frac{n(n-1)}{2}} t^{\frac{n(n+1)}{2}} \Delta_n(p) = \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{(l+1)^2}{2}} \frac{\pi_n}{2\beta_l \beta_{l+1} \beta_{n-l}} + \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+2)}{2}} \frac{\pi_n}{2\beta_l^2 \beta_{n-l}} = \\ \frac{\pi_n}{2\beta_{n+1}\beta_{n-1}} \left(\beta_{n+1} + \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{(l+1)^2}{2}} \frac{\beta_{n+1}^2}{\beta_{l+1}^2 \beta_{n-l}} \right) = \frac{\pi_n}{2\beta_{n+1}\beta_{n-1}} \sum_{j=0}^{(n-1)/2} t^{2j^2} \frac{\beta_{n+1}^2}{\beta_{2j}^2 \beta_{n-2j+1}}.\end{aligned}$$

The sum over j is the sum over the distribution of rank of $(n+1)/2 \times (n+1)/2$ matrices over $\mathbb{F}_{p^2}$ except rank zero, hence it equals $1 - p^{-(n+1)^2/2}$.

Further, $\Delta_n(p) = \Delta_n(pr)$ satisfies the recurrence

$$\Delta_n(p) = \varepsilon^{\frac{(n-1)(n+2)}{2}} t^{\frac{n(n+1)}{2}} \Delta_n(p^{n+1}) \; + \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{l(l+1)}{2}} \varepsilon^{\frac{l(l-1)}{2}} \frac{\pi_n}{\pi_l} \left(\frac{\Delta_l(1)}{\alpha^+_{n-l}} \; - \frac{\Delta_l(r)}{\alpha^-_{n-l}} \right) + \\ + \varepsilon \sum_{\substack{l \text{ even} \\ l<n}} t^{\frac{l(l+1)}{2}} \varepsilon^{\frac{l(l-1)}{2}} \frac{\pi_n}{\pi_l} \left(\frac{\Delta_l(p)}{\alpha^+_{n-l}} \; + \frac{\Delta_l(pr)}{\alpha^-_{n-l}} \right),$$

where we conveniently have $\Delta_l(p) = \Delta_l(pr) = 0$ for even $l < n$. If n is even, the sum over odd l is also zero as $n - l$ is odd and $\alpha^+_{n-l} = \alpha^-_{n-l}$. Hence $\Delta_n(a) = 0$. For odd n

$$\Delta_n(p) - \varepsilon^{\frac{n(n-1)}{2}} t^{\frac{n(n+1)}{2}} \Delta_n(1) = \varepsilon^{\frac{n-1}{2}} t^{\frac{n+1}{2}} \frac{\pi_n}{2\beta_{n+1}\beta_{n-1}} \sum_{\substack{l \text{ odd} \\ l<n}} t^{\frac{(l+1)(l-1)}{2}} \frac{\beta_{n+1}\beta_{n-1}}{\beta_{l-1}\beta_{l+1}\beta_{n-l}} = \\ \varepsilon^{\frac{n-1}{2}} t^{\frac{n+1}{2}} \frac{\pi_n}{2\beta_{n+1}\beta_{n-1}} \sum_{j=0}^{(n-3)/2} t^{2j(j+1)} \frac{\beta_{n+1}\beta_{n-1}}{\beta_{2j}\beta_{2j+2}\beta_{n-1-2j}}.$$

The sum over j is the sum over the distribution of rank for $(n+1)/2 \times (n-1)/2$ matrices over $\mathbb{F}_{p^2}$ as in Corollary 3.2.3 except the last term corresponding to rank 0, hence it equals $1 - p^{-(n^2-1)/2}$. Solving this simple system we see that the statement holds for odd n. □

Combining Lemma 3.3.8 and 3.3.9, we obtain the following theorem.

Theorem 3.3.10. *Let $n \in \mathbb{Z}^{\geq 1}$, and set $t = 1/p$. Then for odd n*

$$\rho_n(a,b) = \begin{cases} \frac{1}{4(1+t)} + b\frac{\pi_n}{4\beta_{n+1}\beta_{n-1}}, & \text{if } a \in \{1, r\}; \\ \\ \frac{t}{4(1+t)} + b\varepsilon^{\frac{n-1}{2}} t^{\frac{n+1}{2}} \frac{\pi_n}{4\beta_{n+1}\beta_{n-1}}, & \text{if } a \in \{p, pr\}; \end{cases}$$

and for even n and $s = \chi(a)$

$$\rho_n(a,b) = \begin{cases} \frac{(1+s(\varepsilon t)^{n/2})(1-s\varepsilon^{n/2}t^{n/2+2})}{4(1+t)(1-t^{n+1})} + b\frac{\pi_n}{2\beta_n\alpha^s_n}, & \text{if } a \in \{1, r\}; \\ \\ \frac{t(1-t^n)}{4(1+t)(1-t^{n+1})}, & \text{if } a \in \{p, pr\}. \end{cases}$$

This also gives us the asymptotic distribution of quadratic forms which doesn't depend on parity.

Corollary 3.3.11. *As $n \to \infty$*

$$\rho(a,b) := \lim_{n\to\infty} \rho_n(a,b) = \begin{cases} \frac{1}{4(1+t)} + b\frac{\pi_\infty}{4\beta^2_\infty}, & \text{if } a \in \{1, r\}; \\ \frac{t}{4(1+t)}, & \text{if } a \in \{p, pr\}. \end{cases}$$

We expect that this distribution should be universal for random quadratic forms with independent coefficients from a large class of distributions not necessarily uniform.

Now let us give a few examples of problems where the distributions we derived come in handy.

3.4 Applications

In this section we would like to illustrate that the distribution of equivalence classes is a versatile tool for approaching different kinds of questions about random matrices. Some of the results have already been established, and some are new, in both cases we give short straightforward proofs which do not require any additional knowledge other than the formulae we derived and some very general mathematical background.

We will write $\mathbb{P}_n$ for the probability measure on $S_n(\mathbb{Z}_p)$. We will also use symbols $\sim$, $\asymp$ and $\ll$ as in the standard Vinogradov notation.

3.4.1 Determinant and rank

The following result is not new (e.g. see Brent and McKay [25]), but we would like to give a short proof using the elementary divisor distribution.

Corollary 3.4.1. *Let Q_n be a random $n \times n$ symmetric matrix with independent entries distributed according to the Haar measure on $\mathbb{Z}_p$, then*

$$\mathbb{P}_n\{|\det Q_n|_p = p^{-k}\} = p^{-k}(1-p^{-n}) + O(p^{-3k/2})\,.$$

Proof. Let λ be a partition with $0 \le \lambda_1 \le \ldots \le \lambda_n$ and $m_t = \#\{i : \lambda_i = t\}$. Let $|\lambda| = \lambda_1 + \ldots + \lambda_n$ be the weight of λ and $\ell(\lambda) = n - m_0 \le n$ be its length. Then

$$\mathbb{P}_n\{|\det Q_n|_p = p^{-k}\} = \sum_{|\lambda|=k} \prod_{i=1}^{n} p^{-\lambda_i(n-i+1)} \frac{\pi_n}{\beta_{m_0}\cdots\beta_{m_k}\cdots} = \sum_{|\lambda|=k} f_n(\lambda)\,,$$

where $f_n(\lambda)$ is as defined in (3.4). Express the last part $\lambda_n = k - \lambda_1 - \ldots - \lambda_{n-1}$ so that

$$\mathbb{P}_n\{|\det Q_n|_p = p^{-k}\} = p^{-k}(1-p^{-n}) \sum_{\lambda} f_{n-1}(\lambda)\gamma_\lambda\,,$$

where the sum now ranges over λ with the restriction $\lambda_{n-1} \le k - |\lambda|$. If $\lambda_{n-1} \ne k - |\lambda|$, we have $\gamma_\lambda = 1/\beta_1 = 1$, otherwise $\gamma_\lambda = \beta_{\lambda_{n-1}}/\beta_{\lambda_{n-1}+1} \le 1/(1-p^{-2})$. Then

$$\mathbb{P}_n\{|\det Q_n|_p = p^{-k}\} = p^{-k}(1-p^{-n})\left(\mathbb{P}_{n-1}\{\mathcal{A}\} + \mathcal{O}\left(\mathbb{P}_{n-1}\{\mathcal{B}\}\right)\right) \ll p^{-k}\,,$$

where

$$\mathcal{A} = \{\lambda_{n-1} < k - |\lambda|\}; \quad \mathcal{B} = \{\lambda_{n-1} = k - |\lambda|\}\,.$$

For the complement of $\mathcal{A}$ we have $\mathbb{P}_{n-1}\{\overline{\mathcal{A}}\} = \mathbb{P}_{n-1}\{\lambda_{n-1} \geq k - |\lambda|\}$, hence

$$\mathbb{P}_{n-1}\{\mathcal{A}\} + \mathcal{O}(\mathbb{P}_{n-1}\{\mathcal{B}\}) = 1 + \mathcal{O}(\mathbb{P}_{n-1}\{\lambda_{n-1} \geq k - |\lambda|\})\,.$$

The condition $\lambda_{n-1} \geq k - |\lambda|$ implies $|\lambda| \geq (k + \lambda_1 + \ldots + \lambda_{n-2})/2 \geq k/2$, thus

$$\mathbb{P}_{n-1}\{\lambda_{n-1} \geq k - |\lambda|\} \leq \mathbb{P}_{n-1}\{|\lambda| \geq k/2\} = \mathbb{P}_{n-1}\{|\det Q_{n-1}|_p \leq p^{-k/2}\} \ll p^{-k/2}\,,$$

and $\mathbb{P}_n\{|\det Q_n|_p = p^{-k}\} = p^{-k}(1 - p^{-n})(1 + \mathcal{O}(p^{-k/2}))$. □

There are several definitions of rank of a matrix over a ring with zero divisors, here $R_m = \mathbb{Z}/p^m\mathbb{Z}$ when $m > 1$. The standard rank of Q over R_m equals rank of its reduction $\overline{Q}$ mod p. The distribution of rank in this case is given by

$$\mathbb{P}_n\{\text{rk } Q = n - r\} = p^{-r(r+1)/2}\frac{\pi_n}{\pi_r\beta_{n-r}}\,.$$

One can also consider determinantal rank

$$\text{rk}_d\, Q = \max\{j \leq n : \text{there exists a non-zero minor of size } j\}\,.$$

In the following corollary we derive the upper and lower bounds on the distribution of determinantal rank. We are not aware if this result has been written down explicitly, but our aim here is to provide a short straightforward proof.

Corollary 3.4.2. *Let Q be a random $n \times n$ symmetric matrix with entries distributed independently and uniformly on R_m. Then as $m \to +\infty$,*

$$\mathbb{P}_n\{rk_d\, Q \leq n - 1\} \;\sim\; \left(1 - p^{-n}\right) p^{-m}\,,$$

for $2 \leq r \leq \lceil \frac{n}{2} \rceil$,

$$mr^{-2}p^{-m(2r-1)} \ll \mathbb{P}_n\{rk_d\, Q \leq n - r\} \;\ll\; m\log(m \wedge n)\; p^{-m(2r-1)}\,,$$

and for $r > \lceil \frac{n}{2} \rceil$

$$\mathbb{P}_n\{rk_d\, Q \leq n - r\} \;\ll\; p^{-mn}\,.$$

All implied constants are independent of n, and are uniform for $p \geq 2$.

Proof. Let $Q_n \in S_n(\mathbb{Z}_p)$ be a random matrix defined as usual. Its reduction is distributed uniformly on $S_n(R_m)$. Determinantal rank is invariant under conjugation by $\mathrm{GL}_n(\mathbb{Z}_p)$, so the probability that rank mod p^m is at most $j-1$ equals

$$\mathbb{P}_n\{k_1 + \ldots + k_j \geq m\}\,.$$

For $j = n$ the statement follows from Corollary 3.4.1. For $j < n$ set $r := n-j+1 \geq 2$, then for $\mathbf{k}_j = (k_1, \ldots, k_j)$

$$\begin{aligned}\mathbb{P}_n\{|\mathbf{k}_j| = l\} &= \sum_{|\mathbf{k}_j|=l} \prod_{i\leq n} \frac{1}{p^{k_i(n-i+1)}} \frac{\pi_n}{\beta_{m_0}\cdots\beta_{m_k}\cdots} \asymp \\ &\asymp \sum_{|\mathbf{k}_j|=l} \prod_{i\leq j} \frac{1}{p^{k_i(j-i+1)}} \frac{1}{p^{l(r-1)}} \frac{\pi_j}{\beta_{m_0}\cdots\beta_{m_{k_j}}} \prod_{i>j} \frac{1}{p^{k_i(r-(i-j))}} \frac{\pi_{r-1}}{\beta_{m_{k_{j+1}}}\cdots} \asymp \\ &\asymp \frac{1}{p^{l(r-1)}} \sum_{\frac{l}{j}\leq t\leq l} \mathbb{P}_j\{|\det Q_j|_p = p^{-l},\ k_j = t\}\mathbb{P}_{r-1}\{k_1 \geq t\}\end{aligned}$$

using that for all $N \in \mathbb{N}$ and $q \geq 2$

$$0 < \prod_{i=1}^{\infty}(1-2^{-i}) \leq \pi_N(q) \leq 1\,.$$

Conditioning on $s = m_t$,

$$\begin{aligned}\mathbb{P}_j\{|\det Q_j|_p = p^{-l},\ k_j = t\} \ll \\ \sum_{1\leq s\leq \frac{l}{t}} \mathbb{P}_{j-s}\{|\det Q_{j-s}|_p = p^{-(l-ts)}, k_{j-s} < t\}p^{-s(l-ts)-\frac{ts(s+1)}{2}}\,.\end{aligned}$$

The largest term is at $s = \lfloor \frac{l}{t} \rfloor$ and of size $\asymp p^{-A}$ with

$$A = \frac{t}{2}\lfloor\frac{l}{t}\rfloor(\lfloor\frac{l}{t}\rfloor + 1) + (l - t\lfloor\frac{l}{t}\rfloor)(\lfloor\frac{l}{t}\rfloor + 1)\,,$$

so

$$p^{-A} \ll \mathbb{P}_j\{|\det Q_j|_p = p^{-l},\ k_j = t\} \ll lt^{-1}p^{-A}\,.$$

With a bit more work one can actually obtain a better upper bound matching the lower bound in order of magnitude. Together with $\mathbb{P}_{r-1}\{k_1 \geq t\} = p^{-tr(r-1)/2}$ we have

$$\mathbb{P}_n\{|\mathbf{k}_j| = l\} \gg \frac{1}{p^{l(n-j)}} \sum_{\frac{l}{j}\leq t\leq l} p^{-A-\frac{t}{2}r(r-1)}\,.$$

The exponent of p is a piecewise continuous function in t, and as a function in integer t it attains its minimum lr on all $t \in \left[\frac{l}{r}; \frac{l}{r-1}\right]$. If $r \leq \lceil\frac{n}{2}\rceil$, this interval is contained inside the range, and for fixed r

$$\mathbb{P}_n\{|\mathbf{k}_j| = l\} \gg lr^{-2}p^{-l(2r-1)}.$$

Similarly the upper bound satisfies

$$\mathbb{P}_n\{|\mathbf{k}_j| = l\} \ll p^{-l(2r-1)} \sum_{\frac{l}{j} \le t \le l} l/t \ll p^{-l(2r-1)} l \log(l \wedge j)\,.$$

If $r > \lceil \frac{n}{2} \rceil$, the minimum is attained at $t = \lceil \frac{l}{n-r+1} \rceil > \frac{l}{r-1}$, but the upper bound follows trivially. Summing over $l \ge m$ concludes the proof. □

3.4.2 Local-global problems

Let $\mathcal{A}$ be a subset of $\mathbb{Z}^N$ such that it has density in the sense that

$$\gamma_{\mathcal{A}} := \lim_{H \to \infty} \frac{\#\{\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{Z}^N : 1 \le x_i \le H,\ \mathbf{x} \in \mathcal{A}\}}{H^N}$$

exists. We say that the probabilistic local-global principle holds if

$$\gamma_{\mathcal{A}} = \prod_{p \text{ prime}} \gamma_{\mathcal{A}}(p),$$

where $\gamma_{\mathcal{A}}(p)$ is the local density of $\mathcal{A}$ over $\mathbb{Z}_p$.

There is a lot of ongoing work on determining when the local-global principle holds in density; for some general conditions see [51], [117].

The set of integer $n \times n$ symmetric matrices is essentially $\mathbb{Z}^{n(n+1)/2}$, so we can use the local-global principle to compute the density of some interesting sets of integer symmetric matrices. To do that over all $\mathbb{Z}_p$, we will assume that Corollary 3.3.5 holds for $p = 2$ (the proof would go similarly to the proof for odd primes, but we do not wish to deal with the canonical form of quadratic forms over $\mathbb{Z}_2$).

Let us first define the elementary divisors of an integer matrix. Let $X \in \mathrm{M}_n(\mathbb{Z})$, then there exist $U, V \in \mathrm{GL}_n(\mathbb{Z})$ such that $X = U\Sigma V$, where $\Sigma = \mathrm{diag}(a_1, \dots, a_n)$ with $a_i \ge 0$ and $a_1 | a_2 | \cdots | a_n$. The integers a_i defined uniquely are the elementary divisors of X. If we look at X as an element of $\mathrm{M}_n(\mathbb{Z}_p)$, then $\mathrm{val}_p(a_i) = k_i$ as in the definition of the corresponding decomposition over $\mathbb{Z}_p$.

Corollary 3.4.3. *As $H \to \infty$ the proportion of integer $n \times n$ symmetric matrices with entries between 1 and H having its first $n-1$ elementary divisors equal to one tends to*

$$\prod_p \left(\frac{\pi_n}{\beta_n} + p^{-1} \frac{\pi_n}{\beta_{n-1}\pi_1} \right).$$

As $n \to \infty$ this expression tends to

$$\prod_{i=1}^{\infty} \zeta(2i+1)^{-1} = (\zeta(3)\zeta(5)\cdots\zeta(2k+1)\cdots)^{-1} \approx 0.7935\,.$$

Proof. The set in question is

$$V(\mathbb{Z}) := \{Q \in S_n(\mathbb{Z}) : a_1 = \ldots = a_{n-1} = 1\} = \{Q \in S_n(\mathbb{Z}) : a_{n-1} = 1\}$$

since $a_j | a_{n-1}$ for all $j \leq n-1$. Then its complement

$$\overline{V}(\mathbb{Z}) = \{Q \in S_n(\mathbb{Z}) : a_{n-1} > 1\}$$

is a closed subscheme of $\mathbb{A}_{\mathbb{Z}}^{n(n-1)/2}$ of rank 2. We can use the local-global principle for this problem due to Ekedahl [51]. Moreover, this is the symmetric matrix version of the application Ekedahl gives in [51]. Over $\mathbb{Z}_p$ we have

$$\gamma_p(V) = \mathbb{P}\{k_{n-1} = 0\} = \frac{\pi_n}{\beta_n} + p^{-1}\frac{\pi_n}{\beta_{n-1}\pi_1} = \begin{cases} \frac{\pi_n}{\beta_n}\left(1 - \frac{1}{p}\right)^{-1}, & \text{if } n \text{ is odd}; \\ \frac{\pi_n}{\beta_n}\left((1-\frac{1}{p})^{-1} - \frac{1}{p^n}\right), & \text{if } n \text{ is even}. \end{cases}$$

This gives us a convergent product and

$$\left| \prod_p \frac{\pi_\infty}{\beta_\infty \pi_1} - \prod_p \left(\frac{\pi_n}{\beta_n} + \frac{1}{p}\frac{\pi_n}{\beta_{n-1}\pi_1} \right) \right| \ll \sum_p \frac{1}{p^{n+1}} \ll \frac{1}{2^n},$$

thus as n grows

$$\prod_p \gamma_p(V) \to \prod_p \frac{\pi_\infty}{\beta_\infty \pi_1} = \prod_{i=1}^{\infty} \zeta(2i+1)^{-1}$$

as required. □

The condition $k_1 = \cdots = k_{n-1} = 0$ is equivalent to $\operatorname{cok} Q$ being cyclic. The asymptotic in Corollary 3.4.3 is known to be the upper bound on the probability to have a cyclic cokernel in case of a general distribution of the entries, see the work of Matchett Wood [132]. Moreover, in the case of general matrices Nguyen and Matchett Wood [112] proved that the limit as $n \to \infty$ is independent of the distribution of entries. It is reasonable to assume that this should also hold for symmetric matrices.

The following two corollaries are new results on the density of integer symmetric matrices that have a square-free minor of a particular size.

Corollary 3.4.4. *Let $r, n \in \mathbb{N}$ with $1 \leq r \leq n-1$. Then as $H \to \infty$ the proportion of integer $n \times n$ symmetric matrices with entries between 1 and H that have a square-free minor of size $n-r$ tends to*

$$\prod_p \left(\sum_{t=0}^{r+1} \frac{\pi_n}{\beta_{n-t}\pi_t} p^{-\frac{t(t+1)}{2}} - \frac{\pi_n}{\beta_{n-r-1}\pi_{r+1}} p^{-(r+1)(r+2)} \right) > 0.$$

Proof. A square-free number is an integer which is not divisible by p^2 for any prime p. Let $V(\mathbb{Z}) \subseteq S_n(\mathbb{Z})$ be the set of matrices with a square-free $(n-r) \times (n-r)$ minor. All minors are polynomial in the entries of the matrix and are independent of each other, so the complement of V is a closed subscheme of $\mathbb{A}^{n(n+1)/2}_{\mathbb{Z}}$ of rank $(r+1)(r+2)/2 \geq 3$. Let

$$N_{\leq H}(n,r) = \#\{Q \in S_n(\mathbb{Z}) : 1 \leq q_{ij} \leq H,\ Q \text{ has a square-free minor of size } n-r\}\,.$$

Applying Ekedahl's sieve [51] gives

$$\lim_{H\to\infty} \frac{N_{\leq H}(n,r)}{H^{n(n+1)/2}} = \prod_p \gamma_p(V)\,,$$

where $\gamma_n(p)$ is the measure of the set

$$V_p = \{Q \in S_n(\mathbb{Z}_p) : k_1 + \ldots + k_{n-r} \leq 1\}\,.$$

The complement of V_p in $S_n(\mathbb{Z}_p)$ is

$$\overline{V}_p = \{Q \in S_n(\mathbb{Z}_p) : \mathrm{rk}_d\, Q \leq n-r-1 \text{ in } \mathbb{Z}/p^2\mathbb{Z}\}\,.$$

There are two options over each $\mathbb{Z}_p$. Either $k_1 = \cdots = k_{n-r} = 0$ with probability

$$a_p(V) = \sum_{t=0}^{r} \frac{\pi_n}{\beta_{n-t}\pi_t} p^{-\frac{t(t+1)}{2}}\,,$$

or $k_1 = \cdots = k_{n-r-1} = 0$ and $k_{n-r} = 1$ with probability

$$b_p(V) = \frac{\pi_n}{\beta_{n-r-1}\pi_{r+1}} p^{-\frac{(r+1)(r+2)}{2}} (1 - p^{-\frac{(r+1)(r+2)}{2}})\,,$$

so $\gamma_p(V) = a_p(V) + b_p(V)$. We have $\mu(\overline{V}_p) \ll p^{-r^2/2} \ll p^{-2}$, thus the product over primes converges to a positive number. Moreover, if r is fixed, or grows as $r = r(n)$, then as $n \to \infty$

$$\prod_p \gamma_p(V) \to \prod_p \frac{\pi_\infty}{\beta_\infty} \left(\sum_{t=0}^{r+1} \frac{p^{-\frac{t(t+1)}{2}}}{\pi_t} - \frac{1}{\pi_{r+1}} p^{-(r+1)(r+2)} \right)\,.$$

In particular, if $r = 1$,

$$\prod_p \gamma_p(V) \to \frac{\prod_p (1 + p^{-3}(1+1/p)^{-1} + p^{-4})}{\zeta(3)\zeta(5)\cdots\zeta(2k+1)\cdots} \approx 0.9581\,,$$

and if $r = r(n) \to \infty$ as $n \to \infty$, then $\prod_p \gamma_p(V) \to 1$. □

One may also ask if a similar statement holds for the determinant itself. The problem here is the fact that this induces a scheme of rank 1, so Ekedahl's sieve does not apply. Counting square-free values of polynomials is a generally difficult problem, however one can circumvent the general obstacles here by utilising the symmetry of the determinant.

Proposition 3.4.5. *The probabilistic local-global principle holds for* $\det X$*, where* $X \in M_n(\mathbb{Z})$ *or* $X \in S_n(\mathbb{Z})$.

Proof. Let $G_n(\mathbb{Z})$ be either $\mathrm{M}_n(\mathbb{Z})$ or $S_n(\mathbb{Z})$, and N_n be either n^2 or $n(n+1)/2$ respectively. If X is a general square matrix, then $|\det X|$ is symmetric with respect to all permutations of x_{ij}, and linear in all x_{ij}. If X is a symmetric matrix, then it is symmetric with respect to permutations of diagonal and off-diagonal elements separately. It is linear in diagonal entries and quadratic in off-diagonal. The determinant is also irreducible and self-similar in the sense that the Laplace expansion for the determinant gives

$$\det X = \sum_{j=1}^{n} (-1)^{i+j} x_{i,j} M_{i,j}\,,$$

where $M_{i,j}$ is the $(i,j)-$th minor.

Let us adopt an approach similar to the work of Poonen [114] building on the work of Hooley [76] on the square-free values of polynomials. We start with showing that $\det X$ is typically not divisible by a square of a large prime. Let

$$\mu_{\leq H}(M) = \#\{X \in G_n(\mathbb{Z}) : 1 \leq x_{ij} \leq H, \exists\, p > M : \ \det X = 0 \bmod p^2\}$$

and for $K = p$ or p^2

$$\nu_{n,\leq H}(K) := \#\{X \in G_n(\mathbb{Z}) : 1 \leq x_{ij} \leq H, \ \det X = 0 \bmod K\}\,.$$

We need to show that the following limit exists and

$$\limsup_{M\to\infty} \limsup_{H\to\infty} \frac{\mu_{\leq H}(M)}{H^{N_n}} = 0\,. \tag{3.8}$$

Suppose $\det X = 0 \bmod p^2$. Then by the Laplace expansion this is equivalent to

$$x_{11} M_{11} + f(X \setminus x_{11}) = 0 \bmod p^2\,.$$

If $M_{11} \neq 0 \bmod p$, then for each $\{x_{ij}\}_{(i,j)\neq(1,1)}$ there are at most $H/p^2 + O(1)$ values of x_{11}. Since $\det X \leq n! H^n$, there are most

$$a(n,H) := \frac{\log(n!H^n)}{2\log H} \leq \frac{n}{2} + \frac{n\log n}{2\log H}$$

primes larger than H such that their squares divide $\det X$. There are at most H^{N_n-1} solutions to $\det X = 0$ in $\mathbb{Z}$, and for any other $\{x_{ij}\}_{(i,j)\neq(1,1)}$ the remaining variable x_{11} can only be in $a(n,H)$ sets implying

$$\mu_{\leq H}(H) \leq a(n,H) H^{N^n-1}\,.$$

In particular, since in our problem n is fixed and H grows, we can think of $a(n,H)$ as a fixed number, say $a(n,H) = a(n) = n$.

Using the union bound for $p \leq H$ and the estimate above for $p > H$, we have

$$\mu_{\leq H}(M) \leq \sum_{M<p\leq H} \nu_{n,\leq H}(p) + a(n) H^{N^n-1} \leq a(n) H^{N^n-1} + \sum_{M<p\leq H} H^{N_n-1}\left(\frac{H}{p^2}+1\right) + \sum_{M<p\leq H} \nu_{n-1,\leq H}(p)\left(\frac{H}{p}+1\right) H^{N_n-N_{n-1}-1} + \sum_{M<p\leq H} \nu_{n-1,\leq H}(p^2) H^{N_n-N_{n-1}}\,.$$

We have

$$\sum_{M<p\leq H} H^{N_n-1}\left(\frac{H}{p^2}+1\right) \leq \frac{H^{N_n}}{M\log M} + \frac{H^{N_n}}{\log H}\,,$$

thus it remains to bound $\tilde{\nu}_{\leq H}(p)$. This is typically a problem, since if we wanted to continue the same procedure for an arbitrary polynomial f by separating its variables of low degree, there is no guarantee that we again get a polynomial with a variable of low degree. However, we do know that for the determinant since $M_{11} = \det Y$ for $Y \in G_{n-1}(\mathbb{Z})$.

Let us first look at small n. If $n = 1$, we have

$$\#\{m \leq H : p|m\} \leq \frac{H}{p},\ \#\{m \leq H : p^2|m\} \leq \frac{H}{p^2}\,.$$

In particular, we trivially have

$$\sum_{m\leq H} \mathbf{I}\{p^2|m\} = \left\lfloor \frac{H}{p^2} \right\rfloor\,,$$

thus

$$\mu_{\leq H}(M) \leq H \sum_{p>M} \frac{1}{p^2} \ll \frac{H}{M\log M}\,.$$

Assume that for $k \leq n-1$

$$\nu_{k,\leq H}(K) \leq H^{N_k-1}\left(\frac{H}{K}+1\right)\,.$$

If $n \geq 2$, then depending on whether $M_{11} = 0 \bmod p$

$$\nu_{n,\leq H}(p) \leq H^{N_n-1}\left(\frac{H}{p}+1\right) + \nu_{n-1,\leq H}(p) H^{N_n-N_{n-1}} \ll_n H^{N_n-1}\left(\frac{H}{p}+1\right)\,.$$

and

$$\nu_{n,\leq H}(p^2) \leq H^{N_n-1}\left(\frac{H}{p^2}+1\right)+\nu_{n-1,\leq H}(p)\left(\frac{H}{p}+1\right)H^{N_n-N_{n-1}-1}+$$
$$\nu_{n-1,\leq H}(p^2)H^{N_n-N_{n-1}} \ll_n H^{N_n-1}\left(\frac{H}{p^2}+1\right).$$

Then

$$\mu_{\leq H}(M) \ll_n H^{N_n-1}\log\log H + b(n)H^{N_n-1} + \frac{H^{N_n}}{M\log M} + \frac{H^{N_n}}{\log H}$$

which implies (3.8) indeed holds. Applying the Chinese remainder theorem for $p < M$ then yields the statement of the proposition. □

Corollary 3.4.6. *As $H \to \infty$ the proportion of integer $n \times n$ symmetric matrices with entries between 1 and H and square-free determinant tends to*

$$\prod_p \left(\frac{\pi_n}{\beta_n} + p^{-1}\frac{\pi_n}{\beta_{n-1}}\right) > 0\,.$$

As $n \to \infty$ this expression tends to $(\zeta(2)\zeta(3)\zeta(5)\cdots\zeta(2k+1)\cdots)^{-1} \approx 0.4824$.

Proof. Let $X \in S_n(\mathbb{Z}_p)$, then $\det X$ is square-free if $|\det X|_p \geq 1/p$. If $|\det Q|_p = 1$, then all $k_i = 0$; if $|\det Q|_p = p^{-1}$, then $k_1 = \ldots = k_{n-1} = 0$ and $k_n = 1$ with total probability

$$\gamma_p(V) = \mathbb{P}_n\{|\det Q|_p \geq p^{-1}\} = \frac{\pi_n}{\beta_n} + \frac{1}{p}\frac{\pi_n}{\beta_{n-1}\beta_1} = \begin{cases} \frac{\pi_n}{\beta_n}\left(1+\frac{1}{p}\right), & \text{if } n \text{ is odd}; \\ \frac{\pi_n}{\beta_n}\left(1+\frac{1}{p}-\frac{1}{p^{n+1}}\right), & \text{if } n \text{ is even}. \end{cases}$$

As $n \to \infty$ we have $\prod_p \gamma_p(V) \to \prod_p(1-p^{-2})\prod_{i=1}^{\infty}(1-p^{-(2i+1)})$. □

3.4.3 Isotropy problem

In this subsection we compute the probability that a random quadratic form with independent coefficients distributed according to the Haar measure on $\mathbb{Z}_p$ is isotropic. A quadratic form Q over a ring R in n variables is isotropic if there exists $\mathbf{t} \in R^n \setminus \{0\}$ such that $Q(t_1, \ldots, t_n) = 0$. Otherwise call Q anisotropic.

It has been long known that a quadratic form over $\mathbb{Z}_p$ in at least 5 variables always has a non-trivial zero, so we only consider $n \leq 4$. Isotropy is invariant under conjugation by $\mathrm{GL}_n(\mathbb{Q}_p)$, thus to determine if a quadratic form is isotropic it suffices to look at its invariants. In particular, if p is an odd prime and $\varepsilon = \chi(-1)$, then

(a) a binary quadratic form Q over $\mathbb{Q}_p$ is isotropic iff $d(Q) = \varepsilon$;

(b) a ternary quadratic form Q over $\mathbb{Q}_p$ is isotropic iff $c(Q) = \langle -1, d(Q)\rangle$;

(c) a quaternary quadratic form Q over $\mathbb{Q}_p$ is anisotropic iff $d(Q) = 1$ and $c(Q) = -1$.

For more details see Cassels' book [29].

The formulae for the probability of isotropy for all $n \leq 4$ have been obtained by Bhargava, Cremona, Fisher, Jones, and Keating [15]. Here we aim to do two things: first, to give a very short proof using the distribution of $\mathbb{Q}_p$-classes, and second, to reconstruct the recursive procedure provided in [15] from the distribution of $\mathbb{Z}_p$-classes and to illustrate that it can in fact be seen as the "local" GOE computation.

Corollary 3.4.7. *Let q_n be the probability of isotropy of a random quadratic form in n variables with coefficients distributed independently and uniformly on $\mathbb{Z}_p$. Set $t = 1/p$, then*

$$\begin{aligned} q_2 &= \frac{1}{2}\,; \\ q_3 &= 1 - \frac{1}{2(1+t)^2}\,; \\ q_4 &= 1 - \frac{1-t}{4(1+t)^2(1-t^5)}\,. \end{aligned}$$

Proof. By Lemma 3.3.8 and 3.3.9, and Theorem 3.3.10

$$\begin{aligned} q_2 &= \sigma_2(\varepsilon) = \frac{(1+\varepsilon^2 t)(1-\varepsilon^2 t^3)}{2(1+t)(1-t^3)} = \frac{1}{2}\,; \\ q_3 &= \rho_3(1,1) + \rho_3(r,1) + \rho_3(p,\varepsilon) + \rho_3(pr,\varepsilon) = 1 - \frac{1}{2(1+t)^2}\,; \\ q_4 &= 1 - \rho_4(1,-1) = 1 - \frac{1-t}{4(1+t)^2(1-t^5)}\,. \end{aligned}$$

One can avoid using the distribution of classes altogether by analysing which combinations of signs and parities of elementary divisors give the right class, and directly computing several geometric sums using Theorem 3.3.4.

For a quadratic form Q from the equivalence class defined by Σ and S we have $d(Q) = \prod_i p^{k_i} S_{ii}$ and $c(Q) = \prod_{i<j} \langle p^{k_i} S_{ii}, p^{k_j} S_{jj} \rangle$. Reduce the number of cases by only considering primitive quadratic forms, i.e. with $k_1 = 0$. For $n = 2$ we have $\mathrm{val}_p(\det Q)$ even and $\prod s_i = \varepsilon$, so

$$q_2 = \frac{1}{1-t^3}\left(\frac{\pi_2}{\alpha_2^{\varepsilon}} + \sum_{\substack{j>0 \\ j \text{ even}}} t^j \frac{\pi_2}{\alpha_1^s \alpha_1^{\varepsilon s}}\right) = \frac{1}{2}\,.$$

For $n = 3$, a quadratic form Q is anisotropic if either $\mathrm{val}_p(\det Q)$ is even and $c(Q) = -1$, or $\mathrm{val}_p(\det Q)$ is odd and $c(Q) = -\varepsilon$, so

$$1 - q_3 = \frac{1}{1-t^6}\Bigg(\frac{\pi_3}{\alpha_2^{-\varepsilon}\beta_1}\sum_{i \text{ odd}} t^i + \frac{\pi_3}{\beta_1\alpha_2^{-\varepsilon}}\sum_{i \text{ odd}} t^{3i} + \frac{\pi_3}{\beta_1}\sum_{\substack{i<j \\ i,j \text{ odd}}} t^{2i+j}\frac{1}{\alpha_1^s\alpha_1^{-s}} + \\ + \frac{\pi_3}{\beta_1}\sum_{\substack{0<i<j \\ i+j \text{ odd}}} t^{2i+j}\frac{1}{\alpha_1^s\alpha_1^{-s\varepsilon}}\Bigg) = \frac{1}{2(1+t)^2}\,.$$

We omit the bulky computation for q_4 with 7 geometric sums over triple indices. □

Yet another option is to use the recurrence equation (3.5) to reproduce the recursive computation from the work of Bhargava, Cremona, Fisher, Jones, and Keating [15].

Let $\mathcal{I}_n$ denote the set of ordered pairs (a, b) such that a quadratic form Q in n variables with $d(Q) = a$ and $c(Q) = b$ is isotropic. Let $u \in \{1, r\}$ and $j \in \mathbb{N}^0$, and let T_j^u denote the operator acting on such pairs as

$$T_j^u(a, b) = (a'(j, u), b'(j, u)) = (aup^j, b\varepsilon^{j(j-1)/2}\langle a, u\rangle\langle p, au\rangle^{j-1})\,.$$

This corresponds to the change in the discriminant and the Hasse invariant of the rest of the quadratic form under rescaling by p. Note that $T_n^1T_n^1 = id$, and $T_n^1\mathcal{I}_n = \mathcal{I}_n$.

Condition on the multiplicity m_0 and use the same observation as in [15] that Q can only be anisotropic if its reduction $\overline{Q}$ mod p doesn't have a non-trivial zero over $\mathbb{F}_p$. This happens in two cases: if $m_0 = \mathrm{rk}\,\overline{Q} = 1$, or if $m_0 = 2$ and $s_0 = -\varepsilon$, i.e. its reduction mod p is irreducible over $\mathbb{F}_p$. If $m_0 \geq 3$ by the Chevalley-Warning theorem $\overline{Q}$ is always anisotropic over $\mathbb{F}_p$. A non-trivial zero over $\mathbb{F}_p$ can be lifted to $\mathbb{Z}_p$ via Hensel's lemma.

Write $\mathbb{P}_n\{\mathcal{I}_n\}$ for the probability that a quadratic form in n variables is isotropic, then

$$\mathbb{P}_n\{\mathcal{I}_n\} = t^{\frac{(n-1)n}{2}}\frac{\pi_n}{\alpha_1^+\pi_{n-1}}\mathbb{P}_{n-1}\{T_{n-1}^1\mathcal{I}_n\} + t^{\frac{(n-1)n}{2}}\frac{\pi_n}{\alpha_1^-\pi_{n-1}}\mathbb{P}_{n-1}\{T_{n-1}^r\mathcal{I}_n\} \\ t^{\frac{(n-2)(n-1)}{2}}\frac{\pi_n}{\alpha_2^{-\varepsilon}\pi_{n-2}}\mathbb{P}_{n-2}\{T_{n-2}^{r\varepsilon}\mathcal{I}_n\} + t^{\frac{n(n+1)}{2}}\mathbb{P}_n\{\mathcal{I}_n\} + \\ \mathbb{P}_n\{m_0 \geq 3 \cup (m_0 = 2, s_0 = -\varepsilon)\}\,.$$

Comparing to the notation of Bhargava, Cremona, Fisher, Jones, Keating [15] we

recognise

$$\xi_1^{(n)} = t^{\frac{(n-1)(n-2)}{2}} \frac{\pi_n}{\alpha_2^{-\varepsilon}\pi_{n-2}} = t^{\frac{(n-2)(n-1)}{2}} \frac{(1-t^{n-1})(1-t^n)}{2(1+t)} = \mathbb{P}_n\{m_0 = 2, s_0 = -\varepsilon\}\,;$$
$$\xi_2^{(n)} = t^{\frac{n(n-1)}{2}} \frac{\pi_n}{\pi_{n-1}\beta_1} = t^{\frac{(n-1)n}{2}} (1-t^n) = \mathbb{P}_n\{m_0 = 1\}\,;$$
$$\xi_0^{(n)} = 1 - t^{\frac{n(n+1)}{2}} - \xi_1^{(n)} - \xi_2^{(n)} = \mathbb{P}_n\{m_0 \geq 3 \cup (m_0 = 2, s_0 = -\varepsilon)\}\,;$$
$$\alpha_1^{(n)} = \mathbb{P}_{n-2}\{T_{n-2}^{r\varepsilon}\mathcal{I}_n\} = \mathbb{P}_n\{Q \text{ is isotropic given } m_0 = 2,\ s_0 = -\varepsilon\}\,;$$
$$\alpha_2^{(n)} = \frac{1}{2}\left(\mathbb{P}_{n-1}\{T_{n-2}^{1}\mathcal{I}_n\} + \mathbb{P}_{n-1}\{T_{n-2}^{r}\mathcal{I}_n\}\right) = \mathbb{P}_n\{Q \text{ is isotropic given } m_0 = 1\}\,,$$

where $\xi_i^{(n)}$ and $\alpha_i^{(n)}$ are as in [15]. Further conditioning in their recursive procedure similarly corresponds to conditioning on the multiplicity $m_1 = \#\{i : k_i = 1\}$ and signature s_1 of the initial quadratic form in n variables in respective cases. For $\alpha_1^{(n)}$ we have

$$\mathbb{P}_{n-2}\{T_{n-2}^{r\varepsilon}\mathcal{I}_n\} = \mathbb{P}_{n-3}\{T_{n-3}^{1}T_{n-2}^{r\varepsilon}\mathcal{I}_n\}\frac{\pi_{n-2}}{\alpha_1^{+}\pi_{n-3}}t^{\frac{(n-3)(n-2)}{2}} +$$
$$\mathbb{P}_{n-3}\{T_{n-3}^{r}T_{n-2}^{r\varepsilon}\mathcal{I}_n\}\frac{\pi_{n-2}}{\alpha_1^{-}\pi_{n-3}}t^{\frac{(n-3)(n-2)}{2}} + \mathbb{P}_{n-4}\{T_{n-4}^{r\varepsilon}T_{n-2}^{r\varepsilon}\mathcal{I}_n\}\frac{\pi_{n-2}}{\alpha_2^{-\varepsilon}\pi_{n-4}}t^{\frac{(n-4)(n-3)}{2}} +$$
$$t^{\frac{(n-1)(n-2)}{2}}\mathbb{P}_{n-2}\{T_{n-2}^{1}T_{n-2}^{r\varepsilon}\mathcal{I}_n\} = \alpha_1^{(n)}\,.$$

The last term has subsequent conditioning on m_2. By the Chevalley-Warning theorem $m_2 > 0$ implies isotropy since solving $a_1x_1^2 + a_2x_2^2 + a_3(px_3)^2 = 0$ is the same as $a_1x_1^2 + a_2x_2^2 + a_3x_3^2 = 0$. Thus

$$\mathbb{P}_{n-2}\{T_{n-2}^{1}T_{n-2}^{r\varepsilon}\mathcal{I}_n\} = (1 - t^{\frac{(n-1)(n-2)}{2}}) + t^{\frac{(n-1)(n-2)}{2}}\mathbb{P}_{n-2}\{T_{n-2}^{1}T_{n-2}^{1}T_{n-2}^{r\varepsilon}\mathcal{I}_n\}\,.$$

Here $\mathbb{P}_{n-2}\{T_{n-2}^{1}T_{n-2}^{1}T_{n-2}^{r\varepsilon}\mathcal{I}_n\} = \mathbb{P}_{n-2}\{T_{n-2}^{r\varepsilon}\mathcal{I}_n\} = \alpha_1^{(n)}$. The corresponding notation in [15] for this case is

$$\beta_1^{(n)} = \mathbb{P}_{n-4}\{T_{n-4}^{r\varepsilon}T_{n-2}^{r\varepsilon}\mathcal{I}_n\} = \mathbb{P}_n\{Q \text{ is isotropic} \mid m_0 = m_1 = 2,\ s_0 = s_1 = -\varepsilon\}\,;$$
$$\beta_2^{(n)} = \frac{1}{2}\sum_{u\in\{1,r\}}\mathbb{P}_{n-3}\{T_{n-3}^{u}T_{n-2}^{r\varepsilon}\mathcal{I}_n\} = \mathbb{P}_n\{Q \text{ is isotropic} \mid m_0 = 2,\ s_0 = -\varepsilon,\ m_1 = 1\}\,;$$

and

$$\nu_1^{(n)} = t^{\frac{(n-1)(n-2)}{2}} = \mathbb{P}_n\{m_2 = 0 \mid m_0 = 2,\ s_0 = -\varepsilon,\ m_1 = 0\}\,;$$
$$\nu_2^{(n)} = 0\,;$$
$$\nu_0^{(n)} = 1 - t^{\frac{(n-1)(n-2)}{2}} = \mathbb{P}_n\{m_2 \geq 1 \mid m_0 = 2,\ s_0 = -\varepsilon,\ m_1 = 0\}\,.$$

For $u \in \{1, r\}$ we have

$$\mathbb{P}_{n-1}\{T^u_{n-1}\mathcal{I}_n\} = t^{\frac{n(n-1)}{2}}\mathbb{P}_{n-1}\{T^1_{n-1}T^u_{n-1}\mathcal{I}_n\} + \mathbb{P}_{n-2}\{T^1_{n-2}T^u_{n-1}\mathcal{I}_n\}\frac{\pi_{n-1}}{\alpha_1^+\pi_{n-2}}t^{\frac{(n-2)(n-1)}{2}} + \mathbb{P}_{n-2}\{T^r_{n-2}T^u_{n-1}\mathcal{I}_n\}\frac{\pi_{n-1}}{\alpha_1^-\pi_{n-2}}t^{\frac{(n-2)(n-1)}{2}} + \mathbb{P}_{n-3}\{T^{r\varepsilon}_{n-3}T^u_{n-1}\mathcal{I}_n\}\frac{\pi_{n-1}}{\alpha_2^{-\varepsilon}\pi_{n-3}}t^{\frac{(n-3)(n-2)}{2}}.$$

The term $\mathbb{P}_{n-1}\{T^1_{n-1}T^u_{n-1}\mathcal{I}_n\}$ needs further conditioning on m_2. Similarly to the previous case, if $m_2 \geq 2$ or if $m_2 = 1$ and $s_0 s_2 = \varepsilon$, a quadratic form will be isotropic, hence

$$\mathbb{P}_{n-1}\{T^1_{n-1}T^u_{n-1}\mathcal{I}_n\} = t^{\frac{(n-1)(n-2)}{2}}\frac{\pi_{n-1}}{\alpha_1^{-\varepsilon s_0}\pi_{n-2}}\mathbb{P}_{n-2}\{T^{r\varepsilon u}_{n-2}T^1_{n-1}T^u_{n-1}\mathcal{I}_n\} + t^{\frac{n(n-1)}{2}}\mathbb{P}_{n-1}\{T^u_{n-1}\mathcal{I}_n\} + 1 - t^{\frac{n(n-1)}{2}} - t^{\frac{(n-1)(n-2)}{2}}\frac{\pi_{n-1}}{\alpha_1^{-\varepsilon s_0}\pi_{n-2}}.$$

Further, one can check that $T^{r\varepsilon u}_{n-2}T^1_{n-1}T^u_{n-1} = T^{r\varepsilon}_{n-2}$ by the definition of T^u_j. The corresponding notation here is

$$\gamma_1^{(n)} = \frac{1}{2}\sum_{u\in\{1,r\}}\mathbb{P}_{n-3}\{T^{r\varepsilon}_{n-3}T^u_{n-1}\mathcal{I}_n\} = \mathbb{P}_n\{Q \text{ isotropic} \mid m_0 = 1,\, m_1 = 2,\, s_1 = -\varepsilon\};$$

$$\gamma_2^{(n)} = \frac{1}{4}\sum_{u,v\in\{1,r\}}\mathbb{P}_{n-2}\{T^u_{n-2}T^v_{n-1}\mathcal{I}_n\} = \mathbb{P}_n\{Q \text{ isotropic} \mid m_0 = m_1 = 1\};$$

$$\eta_1^{(n)} = \frac{1}{2}t^{\frac{(n-1)(n-2)}{2}}(1 - t^{n-1}) = \mathbb{P}_n\{m_2 = 1, s_2 = -\varepsilon s_0 \mid m_0 = 1,\, m_1 = 0\};$$

$$\eta_2^{(n)} = t^{\frac{n(n-1)}{2}} = \mathbb{P}_n\{m_2 = 0 \mid m_0 = 1,\, m_1 = 0\};$$

$$\eta_0^{(n)} = 1 - \eta_1^{(n)} - \eta_2^{(n)}.$$

The coefficients $\beta_i^{(n)}$ and $\gamma_i^{(n)}$ are easy to compute for small n. In particular, if $n \leq 4$, the respective coefficients come from quadratic forms in ≤ 2 variables.

This way of interpreting the recursive procedure from [15] also brings together two seemingly different approaches for computing probability of isotropy over $\mathbb{Z}_p$ and $\mathbb{R}$ in case of the Gaussian Orthogonal Ensemble. First, both over $\mathbb{Z}_p$ and $\mathbb{R}$ the computation is done for "Gaussian" invariant ensembles. Second, all local computations rely on the joint distribution of the canonical form of a quadratic form over the corresponding local field with the difference that eigenvalues over $\mathbb{R}$ are almost surely distinct, and over $\mathbb{Z}_p$ elementary divisors are allowed to repeat.

This comparison hopefully makes a good case for using joint distributions in such problems both in terms of convenience, and in terms of clarity.

3.5 Discussion and prospective work

Our motivation for the problems presented in this chapter comes from many directions.

From the purely probabilistic side, it is interesting to see which results and phenomena translate from classic random matrix theory to random matrix theory over finite fields, or p-adic numbers, or simply integers. There is prior work on general matrices over $\mathbb{Z}_p$, and in this work we considered symmetric matrices. One may expect that as long as a matrix group has some sort of a decomposition, similar ideas would apply. An obvious example here is the additive group of skew-symmetric matrices

$$A_n(R) = \{X \in \mathrm{M}_n(R) : X = -X^\top\}.$$

It is quite likely that one may adapt the proof for the distribution of $\mathbb{Z}_p$-equivalence classes in a straightforward manner. These ideas do not need to be restricted to singular-value-like decompositions.

For example, one may instead work with analogues of the eigenvalue decomposition which can be formulated in terms of the factorisation of the characteristic polynomial. This would be similar to the cycle index results for random matrices over finite fields, see [57]. Obvious matrix groups to look into here would be Lie-type groups like $\mathrm{O}_n(R)$, $\mathrm{U}_n(R)$ and $\mathrm{Sp}_{2n}(R)$.

Symmetric function theory here may provide a slightly different angle. In particular, we have seen in this chapter that random matrices over $\mathbb{Z}_p$ are intimately connected with Hall-Littlewood polynomials, which are a generalisation of Schur polynomials. We may even expect that with appropriate normalisation results over p-adic should morph into the Euclidean case as $p \to \infty$, see for example [129].

From a completely different direction, one may be interested in solving number-theoretic problems involving matrices as we have illustrated in Section 3.4. Those could be either inherently local-global type problems like isotropy, or purely arithmetic problems. In the first case one may benefit from a unified approach over local fields as opposed to treating $\mathbb{Q}_p$ and $\mathbb{R}$ separately. Moreover, joint densities are helpful for computing the actual Euler product. In the second case the benefits are mostly computational. However, as we have shown for square-free values of the determinant, certain functions of matrices may be much more approachable than general multivariate polynomials.

Finally, in light of the recent efforts to use random matrices over p-adics or finite fields to model λ-invariants or the distribution of ranks of elliptic curves it makes

sense to try and better understand the behaviour of different random matrix models on the structural level. Even though this idea is somewhat obvious, let us give an enlightening example. We have recently seen some major breakthroughs in problems studying the size of $\zeta(1/2 + it)$ as discussed in the previous chapter. Some of them came not directly from the random matrix theory heuristics, but from understanding how log-correlated structure manifests in these two setups. Thus we would like to understand if there is an analogue of log-correlation here, and if not, what structural properties random matrices over $\mathbb{F}_p$ or $\mathbb{Z}_p$ share with the objects they are modelling. This in turn circles back to our first point. Thus we suggest to start with studying, for example,

$$\log_p \det(tI_n - X)\,,$$

where $X \in G_n(\mathbb{Z}_p)$, and $\log_p$ is the p-adic logarithm.

Chapter 4

Traces of random matrices over finite fields

This chapter covers joint work with Ofir Gorodetsky. We prove that if X is a random matrix distributed according to the Haar measure on the finite general linear group $\mathrm{GL}_n(\mathbb{F}_q)$, then any non-trivial linear combination of $\mathrm{Tr}(X^1), \mathrm{Tr}(X^2), \ldots, \mathrm{Tr}(X^k)$ has an asymptotically uniform distribution on $\mathbb{F}_q$ as $n \to \infty$ as long as $\log k = o(n)$. Moreover, this range is optimal in a certain sense. Previously this was only known for $k \ll_q n$. Our method is inspired by a classical result of Montgomery and Vaughan on the cancellation in character sums over the integers under the Riemann Hypothesis.

4.1 Introduction

Let $\mathbb{F}_q$ be the finite field of q elements, where $q = p^m$ is a fixed power of a prime p, and let $X \in \mathrm{GL}_n(\mathbb{F}_q)$ be an invertible $n \times n$ matrix over $\mathbb{F}_q$ chosen according to the uniform probability measure on $\mathrm{GL}_n(\mathbb{F}_q)$.

Let $\mathrm{U}_n(\mathbb{C})$ be the group of $n \times n$ unitary matrices over complex numbers, and let $U \in \mathrm{U}_n(\mathbb{C})$ be a random unitary $n \times n$ matrix drawn according to the Haar measure on $\mathrm{U}_n(\mathbb{C})$. A classical result of Diaconis and Shahshahani [49] states that $(\mathrm{Tr}(U), \mathrm{Tr}(U^2), \ldots, \mathrm{Tr}(U^k))$ converges in distribution to $(Z_1, \sqrt{2}Z_2, \ldots, \sqrt{k}Z_k)$, where $\{Z_j\}_{j=1}^k$ are independent standard complex Gaussians. Further, Rains [120] proved that there is a stabilising phenomenon once $k \geq n$ as the eigenvalues of U^k become independent uniform random variables on the unit circle.

Johansson [85] showed that the rate of convergence is superexponential; and due to work of Johansson and Lambert [86] the same rate of convergence applies in total variation. The unitary group is the only compact Lie group known to have such a

fast rate of convergence, for a more detailed discussion of this phenomenon see recent work of Courteaut, Johansson and Lambert [44].

The group $\mathrm{GL}_n(\mathbb{F}_q)$ is a finite field analogue of $\mathrm{U}_n(\mathbb{C})$ making a random matrix X distributed uniformly on $\mathrm{GL}_n(\mathbb{F}_q)$ a finite field analogue of a random unitary matrix U distributed according to the Haar measure on $\mathrm{U}_n(\mathbb{C})$. The analogue of a normal random variable is of course the uniform random variable over $\mathbb{F}_q$. It is then natural to conjecture that $\mathrm{Tr}(X^k)$ should be asymptotically equidistributed over $\mathbb{F}_q$. This turns out to be true as Gorodetsky and Rodgers [64] showed that the total variation distance between the distribution of $\mathrm{Tr}(X^k)$ and the uniform distribution on $\mathbb{F}_q$ is bounded from above by $q^{1-\frac{n^2}{2k}}2^{2n+k}$. This implies that $\mathrm{Tr}(X^k)$ equidistributes in $\mathbb{F}_q$ as $n \to \infty$, uniformly for $k \leq c_q n$ for some sufficiently small c_q depending on q, and the rate of convergence is superexponential. However nothing beyond $k = O(n)$ was known.

In this work we use Hayes characters [71] and their character sums to prove the following result.

Theorem 4.1.1. *Let X be a random matrix distributed according to the Haar measure on $\mathrm{GL}_n(\mathbb{F}_q)$. Let $k = k(n)$ be a positive integer such that $\log k = o(n)$, then the distribution of $\mathrm{Tr}(X^k)$ tends to the uniform distribution on $\mathbb{F}_q$ as $n \to \infty$. Same holds for $\mathrm{Tr}(\sum_{1\leq j\leq k} c_j X^j)$, where $c_j \in \mathbb{F}_q$ are arbitrary constants, $c_k \neq 0$ and $p \nmid k$.*

This result easily extends to collections of traces. In particular, for any fixed positive integer l and positive integers $k_1 < \ldots < k_l$ depending on n such that $p \nmid k_i$ and $\log k_l = o(n)$, the joint distribution of $(\mathrm{Tr}(X^{k_1}), \ldots, \mathrm{Tr}(X^{k_l}))$ is asymptotically uniform on $\mathbb{F}_q^l$ as $n \to \infty$. Note that $\mathrm{Tr}(X^p) = \mathrm{Tr}(X)^p$ in characteristic p, so the condition $p \nmid k$ is necessary to avoid situations like $\mathrm{Tr}(X^p - X) \equiv 0$.

Let us first explain why the range $\log k = o(n)$ is optimal by giving a counterexample $\mathrm{Tr}(\sum_{1\leq i\leq k,\, p\nmid i} c_i g^i)$ with $\log k \asymp n$. Let $\overline{\mathbb{F}_q}$ denote the algebraic closure of $\mathbb{F}_q$, and let $f \in \mathbb{F}_q[T]$ be a monic polynomial factoring over $\overline{\mathbb{F}_q}$ such that

$$f(T) = \prod_{i=1}^{\deg f} (T - \lambda_i)\,.$$

Then its k-th power sum symmetric polynomial is

$$p_k(f) := \sum_{i=1}^{\deg f} \lambda_i^k \in \overline{\mathbb{F}_q}\,.$$

Sometimes we will write $p_k(\lambda)$ or $p_k(\lambda_1, \ldots, \lambda_n)$ instead of $p_k(f)$ to emphasise the connection with the roots of f. By the Newton-Girard formulae $p_k(f)$ is an integral

multivariate polynomial in coefficients of f, hence $p_k(f)$ is in fact in $\mathbb{F}_q$. Moreover, we have

$$p_k(fg) = p_k(f) + p_k(g)$$

for any $f, g \in \mathbb{F}_q[T]$. If $f(T) = \det(I_n T - X)$ is the characteristic polynomial of a matrix X, then $p_k(f) = \mathrm{Tr}(X^k)$. Consider the homogeneous linear system over $\mathbb{F}_q$ in $k - \lfloor k/p \rfloor$ variables $\{c_i\}_{1 \le i \le k,\, p \nmid i}$

$$\sum_{1 \le i \le k,\, p \nmid i} c_i p_i(P) = 0 \tag{4.1}$$

with equations indexed by the monic irreducible polynomials P of degree at most n. By the Prime Polynomial Theorem, there are $N(q,n) \le Aq^n/n$ equations for some constant A independent of q. Hence if $k = 2\lceil Aq^n/n \rceil$, the system has a non-trivial solution $c_i = \overline{c}_i \neq 0$. Since $p_i(fg) = p_i(f) + p_i(g)$, it follows that $\sum_{1 \le i \le k,\, p \nmid i} \overline{c}_i p_i(f)$ vanishes on all monic polynomials f of degree up to n, as any such polynomial is a product of monic irreducibles of degree at most n. Thus

$$\mathrm{Tr}\Big(\sum_{1 \le i \le k,\, p \nmid i} \overline{c}_i X^i\Big) = \sum_{1 \le i \le k,\, p \nmid i} \overline{c}_i p_i(\det(I_n T - X)) = 0$$

for all $X \in \mathrm{GL}_n(\mathbb{F}_q)$.

This illustrates two ideas. First, $\log k = o(n)$ is sharp for any general argument treating all linear combinations of traces. In particular, this bound is sharp for our character sum argument. Second, this implies that $\mathrm{Tr}(X^k)$ is a linear combination of first $N(n,q)$ traces with coefficients independent of X once $k \gg N(n,q)$.

4.1.1 Symmetric function analogue

Let us also formulate a related problem in symmetric functions. Matrix problems and symmetric function theory are inherently connected. On the one hand, we can construct an orthogonal basis for class functions on a matrix group, either finite, or a compact Lie group, using symmetric functions. This is the central idea of the work of Diaconis and Shahshahani [49]. On the other hand, functions that we are normally interested in such as $\det X$, $\mathrm{Tr}(X^k)$ or $\det(I_n T - X)$ have a formal expression in standard symmetric polynomials.

Let $p_k(x_1, \ldots, x_n)$ be the k-th power sum symmetric polynomial

$$p_k(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i^k \,.$$

Let $e_k(x_1, \ldots, x_n)$ be the k-th elementary symmetric function

$$e_k(x_1, \ldots, x_n) := \sum_{1 \le j_1 < j_2 < \cdots < j_k \le n} x_{j_1} x_{j_2} \cdots x_{j_k} . \tag{4.2}$$

Let $\mathbf{x} = \{x_i\}_{i=1}^n$ be $\overline{\mathbb{F}_q}$-valued random variables such that $e_1(\mathbf{x}), \ldots, e_n(\mathbf{x})$ are independent and uniformly distributed on $\mathbb{F}_q$. By the Newton-Girard formulae, the power sum symmetric polynomials $p_k(\mathbf{x})$ must also be $\mathbb{F}_q$-valued. We may also expect $p_k(\mathbf{x})$ to be asymptotically uniform.

To construct such a system, for example, take $a_1, \ldots, a_n$ to be independent uniform random variables on $\mathbb{F}_q$. Then the univariate polynomial

$$T^n + a_1 T^{n-1} + a_2 T^{n-2} + \ldots + a_n$$

is distributed uniformly among the monic polynomials of degree n. Setting x_i to be its n roots in $\overline{\mathbb{F}_q}$ taken with multiplicity in some order, we have $e_j(\mathbf{x}) = a_j(-1)^j$ satisfy the conditions above.

We prove that equidistribution occurs in this case as well.

Theorem 4.1.2. *Let n and $k = k(n)$ be positive integers such that $\log k = o(n)$, then distribution of $p_k(\mathbf{x})$ tends to the uniform distribution on $\mathbb{F}_q$ as n tends to ∞. The same holds for $\sum_{1 \le j \le k} c_j p_j(\mathbf{x})$, where $c_j \in \mathbb{F}_q$ are arbitrary constants with $c_k \neq 0$ and $p \nmid k$.*

Theorem 4.1.2 is essentially about $p_k(f)$ when we force f to be uniformly distributed on the set of monic polynomials of degree n, while Theorem 4.1.1 is about $p_k(f)$ for a polynomial f drawn from the space of possible characteristic polynomials of a matrix from $\mathrm{GL}_n(\mathbb{F}_q)$ endowed with its Haar measure.

From [64, Thm. 1.4], the total variation distance of the law of the first k next-to-leading coefficients of $\det(I_n T - X)$ from the uniform distribution $\mathbb{F}_q^k$ tends to 0 as $n \to \infty$ for k as large as $n - o(\log n)$, so the setup of Theorem 4.1.2 is not that different from the setup of Theorem 4.1.1.

4.1.2 Character sums

The key observation to tackle our matrix problem is to notice that it can be stated using character sums over function fields. In particular, we utilise the fact that

$$\psi_0(p_k(f)) = e^{\frac{2\pi i}{p} \operatorname{Tr}_{\mathbb{F}_q/\mathbb{F}_p} p_k(f)}$$

is a Hayes character of conductor $k-1$. Hayes characters are a generalisation of Dirichlet characters, and unsurprisingly one can show that their sums satisfy bounds analogous to those of the usual Dirichlet characters.

Another important observation concerns the distribution function of the characteristic polynomial. Let $\mathcal{M}_q \subseteq \mathbb{F}_q[T]$ be the set of all monic polynomials over $\mathbb{F}_q$, and $\mathcal{M}_{n,q}$ be its subset of polynomials of degree n. Let $f \in \mathcal{M}_q$ and set

$$\alpha(f) := \mathbb{P}_{X \in \mathrm{GL}_{\deg f}(\mathbb{F}_q)}(\mathrm{CharPoly}(X) = f)\,,$$

where $\mathrm{GL}_{\deg f}(\mathbb{F}_q)$ is endowed with the uniform measure. When restricted to $\mathcal{M}_{n,q}$, by definition $\alpha(f)$ gives the distribution of the characteristic polynomial of a matrix from $\mathrm{GL}_n(\mathbb{F}_q)$. However if we look at it as a function on $\mathcal{M}_q$, it turns out that $\alpha(f)$ is multiplicative, which follows from the work of Reiner [121] and Gerstenhaber [63].

Using these two facts and adapting an argument of Montgomery and Vaughan [102] we then reduce our problem to studying sums of Hayes characters twisted by a multiplicative function. In particular, let $n \in \mathbb{N}$ and consider the sums

$$S_\alpha(n,\chi) := \sum_{f \in \mathcal{M}_{n,q}} \alpha(f)\chi(f)$$

and

$$S_\alpha(n,m,\chi) := \sum_{\substack{f \in \mathcal{M}_{n,q} \\ m\text{-friable}}} \alpha(f)\chi(f)\,.$$

Here an m-friable (smooth) polynomial is a polynomial divisible only by primes of degree at most m. In [64] Gorodetsky and Rodgers showed that for a Hayes character χ of conductor $d(\chi)$

$$|S_\alpha(n,\chi)| \leq q^{-\frac{n^2}{2d(\chi)}+O(n)}.$$

The proof uses Weil's Riemann Hypothesis in a non-essential way which only affects the Big-O constant in the exponent. By adapting the works on sums of Hayes characters that rely crucially on the Riemann Hypothesis over function field (Weil's RH) we obtain the following result.

Theorem 4.1.3. *Let χ be a non-principal Hayes character of conductor $d(\chi)$. Then for positive integers n and $m = m(n)$ such that $2\log_q(d(\chi)+1) \leq m \leq n$ we have*

$$S_\alpha(n,\chi) = S_\alpha(n,m,\chi) + O\left((d(\chi)+1)q^{-\frac{m}{2}}\right),$$

and if $\log(d(\chi)+1) = o(n)$ we have

$$S_\alpha(n,\chi) = o(1)$$

as $n \to \infty$.

This result is slightly more general than needed for the purposes of Theorem 4.1.1 and thus can be regarded as having independent value. For some general theory of sums of Hayes characters twisted by multiplicative functions we refer the reader to the work of Klurman, Mangerel and Teravainen [94].

4.2 Preliminaries

Let $\mathcal{M}_{n,q} \subseteq \mathbb{F}_q[T]$ be the set of monic polynomials of degree n and $\mathcal{M}_q = \cup_{n\geq 0}\mathcal{M}_{n,q}$ be the set of all monic polynomials. Let $\mathcal{P}_{n,q} \subseteq \mathcal{M}_{n,q}$ be the set of irreducible polynomials (primes) of degree n and $\mathcal{P}_q = \cup_{n\geq 0}\mathcal{P}_{n,q} \subseteq \mathcal{M}_q$ be the set of all primes. The letter P always denotes an element from $\mathcal{P}_q$.

4.2.1 Hayes characters

Hayes characters were first introduced in the work of Hayes [71]. They are essentially a generalisation of Dirichlet characters in the function field setting. There are two ways to define them, both covered in the initial work [71].

The first option is to define them as characters of an equivalence relation, prove that the corresponding quotient group admits a set of characters, and then check that corresponding orthogonal relations hold. Alternatively, one may follow the work of Keating and Rudnick [90, 91] and use an involution on $\mathcal{M}_q$ to express a Hayes character as a product of two Dirichlet characters. We prefer this approach, so let us show how to do that.

Let $f \in \mathcal{M}_q$ and $l \in \mathbb{Z}^{\geq 0}$. Let

$$f^*(T) = T^{\deg f} f(1/T)$$

be the polynomial with the same coefficients as f but in inverted order. Then a short interval character ξ of $l-1$ coefficients is a multiplicative function such that

$$\xi(f) = \psi(f^*)\,,$$

where ψ is some Dirichlet character mod T^l. Let ψ have conductor $d_0(\psi) \leq l$, then set the length of ξ to be $\ell(\xi) := d_0(\psi) - 1$. In particular, if ψ is primitive, then $\ell(\xi) = l - 1$.

Say that two functions $f_1, f_2 \in \mathcal{M}_q$ are equivalent under the relation $R_{l,q}$ if they have the same first l next-to-leading coefficients. Short interval characters of l coefficients are exactly the characters of the group $(\mathcal{M}_q/R_{l,q})^\times$. A short interval character

ξ of l coefficients is always a short interval character of $l' > l$ coefficients, and its length $\ell(\xi)$ equals the smallest l such that ξ is a character of $(\mathcal{M}_q/R_{l,q})^\times$.

Conversely, let $k \in \mathbb{N}$ and let $F : \mathcal{M}_q \to \mathbb{C}^\times$ be a completely multiplicative function such that $F(1) = 1$ and $F(f) = 0$ if f^* is not invertible mod T^{k+1}. Moreover, suppose $F(f) = F(g)$ if f and g have the same first k next-to-leading coefficients. Then $F(f)$ is a short interval character of k coefficients. Here we are interested in a particular example. Let

$$\psi_0(a) = e^{\frac{2\pi i}{p} \operatorname{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(a)}.$$

Then we have the following.

Lemma 4.2.1. *Let $n, k \in \mathbb{N}$, and let $f \in \mathcal{M}_{n,q}$. Let $\{\lambda_j\}_{1\le j\le n}$ be roots of f in $\overline{\mathbb{F}_q}$, not necessarily distinct. Let $p_j(f) = \sum_{i=1}^n \lambda_i^j$ and $h_k(f) = \sum_{j=1}^k c_j p_j(f)$, where $c_k \neq 0$ and $p \nmid k$. Then*

$$\psi_0(h_k(f)) = e^{\frac{2\pi i}{p} \operatorname{Tr}_{\mathbb{F}_q/\mathbb{F}_p} h_k(f)}$$

is a non-principal short interval character of k coefficients.

Proof. Let $f, g \in \mathcal{M}_q$ with $\deg f = n$ and $\deg g = m$. Let $\{\lambda_j\}_{1\le j\le n}$ be roots of f and $\{\mu_l\}_{1\le l\le m}$ be roots of g in $\overline{\mathbb{F}_q}$, both taken with multiplicity. Then roots of fg are $\{\lambda_j\} \cup \{\mu_l\}$, thus for all j

$$p_j(fg) = p_j(f) + p_j(g),$$

which then implies that $\psi_0(h_k(f))$ is completely multiplicative.

We trivially have $p_j(1) = 0$, so $\psi_0(h_k(1)) = 1$. By the definition of the involution since f is monic, f^* is always invertible mod T^j for all j with $f^*(0) = 1$. It remains to check that $\psi_0(h_k(f))$ only depends on the first k next-to-leading coefficients of f.

Let

$$e_j(f) = e_j(\lambda_1, \dots, \lambda_n) := \sum_{1\le i_1<i_2<\cdots<i_j\le n} \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_j}.$$

Then $f(T) = T^n + a_1T^{n-1} + \cdots + a_{n-1}T + a_n$, where $a_j = (-1)^j e_j(f)$. By the Newton identities,

$$p_j(f) = \begin{bmatrix} e_1 & 1 & 0 & \cdots & 0 \\ 2e_2 & e_1 & 1 & \cdots & 0 \\ 3e_3 & e_2 & e_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ je_j & e_{j-1} & \cdots & e_2 & e_1 \end{bmatrix},$$

so $h_k(f) = \sum_{j=1}^k c_j p_j(f)$ indeed only depends on $e_1, \dots, e_k$ and thus $a_1, \dots, a_k$. □

Let $Q \in \mathbb{F}_q[T] \setminus \{0\}$ and $l \in \mathbb{Z}^{\geq 0}$. Then the characters of the multiplicative group $G_{Q,l,q} := (\mathbb{F}_q[T]/Q\mathbb{F}_q[T])^\times \times (\mathcal{M}_q/R_{l,q})^\times$ denoted $\mathcal{X}_{Q,l,q}$ are called Hayes characters. In particular, every Hayes character $\chi \in \mathcal{X}_{Q,l,q}$ can be written in the form

$$\chi := \psi\xi\,,$$

where ψ is a Dirichlet character mod Q and ξ is a short interval character of l coefficients. The conductor of a Hayes character is defined as $d(\chi) := d_0(\psi) + \ell(\xi)$. If at least one of ψ and ξ is non-principal, then $d(\chi) > 1$ and χ is non-principal. There are $|\mathcal{M}_{l+1,q}| = q^l$ short interval characters of l coefficients, and

$$|(\mathbb{F}_q[T]/Q\mathbb{F}_q[T])^\times| = \phi(Q) = q^{\deg Q}\prod_{P|Q}(1 - q^{-\deg P})$$

Dirichlet characters mod Q. Thus $|G_{Q,l,q}| = |\mathcal{X}_{Q,l,q}| = q^l\phi(Q)$.

If χ is defined as above, then

$$\frac{1}{q^l\phi(Q)}\sum_{f\in G_{Q,l,q}}\chi(f) = \begin{cases}1, & \chi \text{ is principal}\,;\\ 0, & \text{otherwise}\,,\end{cases}$$

where the sum can also be taken over a representative set of $G_{Q,l,q}$ in $\mathcal{M}_q$ (a set containing exactly one polynomial from each equivalence class in $G_{Q,l,q}$), and

$$\frac{1}{q^l\phi(Q)}\sum_{\chi\in\mathcal{X}_{Q,l,q}}\chi(f)\bar{\chi}(g) = \begin{cases}1, & f = g \bmod Q,\ f^* = g^* \bmod T^k\,;\\ 0, & \text{otherwise}\,.\end{cases}$$

Let $F : \mathcal{M}_q \to \mathbb{C}$ be a multiplicative function on monic polynomials, then its L-function is

$$L(u, F) := \sum_{f\in\mathcal{M}_q} F(f)u^{\deg f}\,.$$

If χ is a non-principal Hayes character, then $L(u, \chi)$ is a polynomial of degree at most $d(\chi)$. Moreover, Weil's RH applies to $L(u, \chi)$ as proved by Rhin [122].

Theorem 4.2.2 (Weil's RH)**.** *Let χ be a non-principal Hayes character of conductor $d(\chi)$. If its L-function $L(u, \chi)$ factors as*

$$L(u, \chi) = \prod_{i=1}^{d(\chi)}(1 - \gamma_i u) \tag{4.3}$$

then $|\gamma_i| \in \{1, \sqrt{q}\}$ for $1 \leq i \leq d(\chi)$.

4.2.2 Basic bounds

The polynomial von Mangoldt function $\Lambda\colon \mathcal{M}_q \to \mathbb{C}$ is defined as

$$\Lambda(f) = \begin{cases} \deg P & \text{if } f = P^k, P \in \mathcal{P}_q, k \geq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

Gauss' identity [123, Thm. 2.2] states that

$$\sum_{f \in \mathcal{M}_{n,q}} \Lambda(f) = \sum_{d|n} |\mathcal{P}_{d,q}| d = q^n \,, \tag{4.5}$$

and in particular, this implies that $|\mathcal{P}_{n,q}| \leq q^n/n$ for all $n \geq 1$. Using these facts one can obtain the following trivial estimates.

Lemma 4.2.3. *Let χ be a non-principal Hayes character of conductor $d(\chi)$. We have*

$$\left| \sum_{f \in \mathcal{M}_{n,q}} \chi(f)\Lambda(f) \right| \leq q^{\frac{n}{2}} d(\chi) \,. \tag{4.6}$$

Proof. Let $L(u,\chi)$ be the L-function of χ, then

$$L(u,\chi) = \sum_{f \in \mathcal{M}_q} \chi(f) u^{\deg f} = \prod_{P \in \mathcal{P}_q} \left(1 - \chi(P) u^{\deg P}\right)^{-1}$$

Take the logarithmic derivative of the Euler product and (4.3) and compare the coefficients to obtain

$$\sum_{f \in \mathcal{M}_{n,q}} \chi(f)\Lambda(f) = -\sum_{i=1}^{d(\chi)} \gamma_i^n$$

for all $n \geq 1$. Then GRH for $L(u,\chi)$ implies the statement of the lemma. □

Lemma 4.2.4. *Let χ be a non-principal Hayes character of conductor $d(\chi)$, then*

$$\left| \sum_{P \in \mathcal{P}_{n,q}} \chi(P) \right| \leq \frac{q^{\frac{n}{2}}}{n} (d(\chi) + 2).$$

Proof. We have

$$\sum_{P \in \mathcal{P}_{n,q}} \chi(P)\Lambda(P) = n \sum_{P \in \mathcal{P}_{n,q}} \chi(P) = \sum_{f \in \mathcal{M}_{n,q}} \chi(f)\Lambda(f) - \sum_{f \in \mathcal{M}_{n,q} \setminus \mathcal{P}_{n,q}} \chi(f)\Lambda(f) \,.$$

The first sum on the right hand side is bounded by $q^{\frac{n}{2}} d(\chi)$ by the lemma above. The trivial bound on $|\mathcal{P}_{d,q}|$ implies that the second sum is bounded by

$$\sum_{d|n,\ d \neq n} |\mathcal{P}_{d,q}| d \leq \sum_{d|n,\ d \neq n} q^d \leq 2q^{\frac{n}{2}} \,.$$

The statement of the lemma follows after applying the triangle inequality. □

Finally, we will need the following fact about power series.

Lemma 4.2.5. *Let $F(u) = \sum F_i u^i$, $G(u) = \sum G_i u^i$ be power series with non-negative G_i and $|F_i| \le G_i$. Suppose G converges absolutely in $|u| < C$. Then $|[u^j]F(u)| \le G(R)R^{-j}$ for every $0 < R < C$.*

Proof. $G(R)R^{-j} = \sum G_i R^{i-j} \ge G_j R^{j-j} = G_j \ge |F_j| = |[u^j]F(u)|$. □

4.3 From traces to character sums

First, use Fourier analysis on $(\mathbb{F}_q, +)$ to reduce our questions in Theorems 4.1.1 and 4.1.2 to questions about cancellation in character sums over function fields.

Let $\psi_0 \colon \mathbb{F}_q \to \mathbb{C}^\times$ be the character defined by

$$\psi_0(a) = e^{\frac{2\pi i}{p} \operatorname{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(a)},$$

where p is the characteristic of $\mathbb{F}_q$.

Let $f \in \mathcal{M}_{n,q}$ be a polynomial with roots $\lambda = (\lambda_1, \dots, \lambda_n)$ taken with multiplicity, and define

$$p_j(f) := p_j(\lambda) = \sum_{i=1}^{n} \lambda_i^k .$$

We will use $p_j(\lambda)$ as a function of roots and $p_j(f)$ as a function of the polynomial itself interchangeably. Set

$$h_k(f) = \sum_{1 \le j \le k} c_j p_j(f),$$

where $c_j \in \mathbb{F}_q$, $c_k \ne 0$ and $p \nmid k$. Then

$$\mathbb{P}(h_k(f) = a) = \mathbb{P}_{f \in \mathcal{M}_{n,q}}(h_k(f) = a) = \frac{1}{q} \sum_{c \in \mathbb{F}_q} \overline{\psi_0}(ca) \mathbb{E}_{f_n \in \mathcal{M}_{n,q}} \psi_0(c h_k(f)),$$

where f is a random polynomial from $\mathcal{M}_{n,q}$. The contribution of $c = 0$ is precisely $1/q$, so by the triangle inequality we have

$$\sum_{a \in \mathbb{F}_q} \left| \mathbb{P}(h_k(f) = a) - \frac{1}{q} \right| \le q \max_{c \in \mathbb{F}_q^\times} \left| \mathbb{E}_{f \in \mathcal{M}_{n,q}} \psi_0(c h_k(f)) \right| . \tag{4.7}$$

For any $c \in \mathbb{F}_q$ the function taking f to $\psi_0(c h_k(f))$ is completely multiplicative, and is in fact a short interval character, see [64, Lem. 2.2]. Then its L-function is

$$L(u, \chi) = \sum_{f \in \mathcal{M}_q} \chi(f) u^{\deg f} = \prod_{P \in \mathcal{P}_q} (1 - \chi(P) u^{\deg P})^{-1} . \tag{4.8}$$

If χ is non-principal, $L(u, \chi)$ is a polynomial of degree $d(\chi)$. If $\chi(f) = \psi_0(c p_k(f))$, or $\chi(f) = \psi_0(c h_k(f))$ as above, we have

$$d(\chi) = k - 1 . \tag{4.9}$$

4.3.1 Proof of Theorem 4.1.2

When f is distributed uniformly on $\mathcal{M}_{n,q}$, inequality (4.7) reduces our problem to checking if there is *any* cancellation in

$$S(n,\chi) := \sum_{f\in\mathcal{M}_{n,q}} \chi(f) = [u^n]L(u,\chi)$$

over the trivial bound $\#\mathcal{M}_{n,q} = q^n$. Though we state Theorems 4.1.1 and 4.1.2 in a non-quantitative way, to give an idea of the order of the error, let us present what is known about the bounds on $S(n,\chi)$.

By definition, $\sum_{f\in\mathcal{M}_{n,q}} \chi(f)$ vanishes for $n > d(\chi)$, so we may restrict to $d(\chi) \geq n$. First, we have the function field version of the Pólya-Vinogradov inequality due to Hsu [78]

$$S(n,\chi) = O(q^{d(\chi)/2}),$$

so there is cancellation as long as $2n - d(\chi)$ tends to infinity. In the same work Hsu [78] proved the analogue of the Burgess bound for a non-principal character χ of prime conductor,

$$|S(n,\chi)| \leq 2r\sqrt{d(\chi)}q^{n+3/2-3d(\chi)/4r(r+1)}$$

for $r \in \mathbb{N}$ such that $n/d(\chi) > 1/4 + 1/r$. This case allows $4n - d(\chi) \geq C(n\log n)^{1/2}$. Bhowmick, Lê and Liu [20] prove that as $n \to \infty$

$$|S(n,\chi)| \leq q^{\frac{n}{2}} \exp\left(O\left(\frac{n\log\log d(\chi)}{\log d(\chi)} + \frac{d(\chi)}{\log^2 d(\chi)}\right)\right). \tag{4.10}$$

This estimate exhibits cancellation when $d(\chi) = o(n\log^2 n)$. The proof of [20] can also be modified to obtain a stronger bound with cancellation in the range $d(\chi) = n^{1+o(1)}$. Further, adapting an argument of Montgomery and Vaughan [102, Lem. 2] to the function field setting, Bhowmick and Lê [19] proved that for $2\log_q d(\chi) \leq m \leq n$

$$S(n,\chi) = \sum_{\substack{f\in\mathcal{M}_{n,q}\\ f \text{ is } m\text{-friable}}} \chi(f) + O\left(d(\chi)q^{n-\frac{m}{2}}\right),$$

where m-friable, or m-smooth, stands for only having prime factors of degree up to m. If $m = o(n)$ then as $n \to \infty$

$$\sum_{\substack{f\in\mathcal{M}_{n,q}\\ f \text{ is } m\text{-friable}}} 1 = o(q^n),$$

hence taking $m = \lceil 2\log_q d(\chi) + \log n\rceil$ and assuming $\log d(\chi) = o(n)$ we have

$$q^{-n}S(n,\chi) = o(1)$$

as $n \to \infty$. That is, we obtain cancellation in $S(n, \chi)$ as long as $\log d(\chi) = o(n)$. This discussion together with (4.7) and (4.9) yields Theorem 4.1.2.

4.3.2 Proof of Theorem 4.1.1

Let X be a random matrix from $\mathrm{GL}_n(\mathbb{F}_q)$ endowed with the uniform measure, then for $f \in \mathcal{M}_{n,q}$

$$\alpha(f) := \mathbb{P}_{X \in \mathrm{GL}_n(\mathbb{F}_q)}(\mathrm{CharPoly}(X) = f) \tag{4.11}$$

is the probability that the characteristic polynomial of g equals f. If we look at $\alpha(f)$ as a function of $f \in \mathcal{M}_q$, it turns out to be multiplicative, and it only depends on the degree and multiplicity of the irreducible factors of f. Let us also give an explicit expression for $\alpha(f)$. Let

$$(a; b)_k = \prod_{j=0}^{k-1} (1 - ab^j)$$

be the q-Pochhammer symbol. Then for $P \in \mathcal{P}_q \setminus \{T\}$ with $a = |P|^{-1} = q^{-\deg P}$

$$\alpha(P^k) = a^k / (a; a)_k \, .$$

Naturally, we have $\sum_{f \in \mathcal{M}_{n,q}} \alpha(f) = 1$, but one can also check that

$$\sum_{f \in \mathcal{M}_{n,q}} \alpha(f) = [u^n] \sum_{f \in \mathcal{M}_q} \alpha(f) u^{\deg f} = [u^n] \prod_{j \geq 1} \left(1 - \frac{u}{q^j}\right)^{-1} = 1$$

for all $n \geq 1$. In particular, using the q-binomial theorem and the Taylor expansion of the logarithm we have

$$\sum_{j \geq 0} \alpha(P^j) u^{j \deg P} = \sum_{j \geq 0} \frac{a^j u^{j \deg P}}{(a, a)_j} = \prod_{j \geq 1} \left(1 - \frac{u^{\deg P}}{q^{j \deg P}}\right)^{-1} =$$
$$\exp\left(-\sum_{j \geq 1} \log\left(1 - \frac{u^{\deg P}}{q^{j \deg P}}\right)\right) = \exp\left(\sum_{j \geq 1} \frac{u^{j \deg P}}{j(q^{j \deg P} - 1)}\right)$$

thus using Gauss' identity

$$\sum_{f \in \mathcal{M}_q} \alpha(f) u^{\deg f} = \prod_{P \in \mathcal{P}_q} \left(\sum_{j \geq 0} \alpha(P^j) u^{j \deg P}\right) = \exp\left(\sum_{k \geq 1} |\mathcal{P}_{k,q}| \sum_{j \geq 1} \frac{u^{jk}}{j(q^{jk} - 1)}\right) =$$
$$\exp\left(\sum_{j \geq 1} \frac{u^m}{m(q^m - 1)} \sum_{k | m} k |\mathcal{P}_{k,q}|\right) = \exp\left(\sum_{j \geq 1} \frac{(qu)^m}{m(q^m - 1)}\right) = \prod_{j \geq 1} \left(1 - \frac{u}{q^j}\right)^{-1} .$$

Consider the twisted character sums

$$S_\alpha(n,\chi) := \sum_{f\in\mathcal{M}_{n,q}} \alpha(f)\chi(f)$$

and

$$S_\alpha(n,m,\chi) := \sum_{\substack{f\in\mathcal{M}_{n,q}\\ f \text{ is } m\text{-friable}}} \alpha(f)\chi(f)\,.$$

Similarly to the previous case, for $h_k(f) = \sum_{1\le j\le k} c_j p_j(f) = \sum_{1\le j\le k} c_j \operatorname{Tr}(X^j)$

$$\begin{aligned}\mathbb{P}_{X\in\mathrm{GL}_n(\mathbb{F}_q)}(h_k(f)=a) = \frac{1}{q}\sum_{c\in\mathbb{F}_q}\overline{\psi_0}(ca)\mathbb{E}_{g\in\mathrm{GL}_n(\mathbb{F}_q)}\psi_0(ch_k(f)) = \\ \frac{1}{q}\sum_{c\in\mathbb{F}_q}\overline{\psi_0}(ca)\sum_{f\in\mathcal{M}_{n,q}}\alpha(f)\psi_0(ch_k(f)) \\ = \frac{1}{q}\sum_{c\in\mathbb{F}_q}\overline{\psi_0}(ca)S_\alpha(n,\psi_0(ch_k(\cdot))),\end{aligned}$$

thus we have the following analogue of (4.7):

$$\sum_{a\in\mathbb{F}_q}\left|\mathbb{P}_{g\in\mathrm{GL}_n(\mathbb{F}_q)}(h_k(f)=a)-\frac{1}{q}\right| \le q\max_{c\in\mathbb{F}_q^\times}|S_\alpha(n,\psi_0(ch_k(\cdot)))|\,. \tag{4.12}$$

Again, in order to prove convergence to uniform distribution it suffices to show some cancellation in $S_\alpha(n,\chi)$ over the trivial bound $\sum_{f\in\mathcal{M}_{n,q}}\alpha(f)=1$.

Then Theorem 4.1.3, which we prove below, together with (4.12) yields Theorem 4.1.1.

4.3.3 Proof of Theorem 4.1.3

Let m be a parameter such that $1\le m\le n$. For any Hayes character χ, define

$$L_{\alpha,>m}(u,\chi) := \prod_{P\in\mathcal{P}_q,\,\deg P>m}\left(\sum_{k\ge0}\alpha(P^k)\chi(P^k)u^{k\deg P}\right),$$

$$L_{\alpha,\le m}(u,\chi) := \prod_{P\in\mathcal{P}_q,\,\deg P\le m}\left(\sum_{k\ge0}\alpha(P^k)\chi(P^k)u^{k\deg P}\right),$$

and

$$L_\alpha(u,\chi) := L_{\alpha,>0}(u,\chi) = \sum_{f\in\mathcal{M}_q}\alpha(f)\chi(f)u^{\deg f}\,.$$

We have

$$\begin{aligned}S_\alpha(n,\chi)-S_\alpha(n,m,\chi) = [u^n](L_\alpha(u,\chi)-L_{\alpha,\le m}(u,\chi)) = \\ [u^n]L_{\alpha,\le m}(u,\chi)\,(L_{\alpha,>m}(u,\chi)-1)\,.\end{aligned} \tag{4.13}$$

Let $(a;b)_k = \prod_{j=0}^{k-1}(1-ab^j)$ be the q-Pochhammer symbol. Then for $P \in \mathcal{P}_q \setminus \{T\}$ with $a = |P|^{-1} = q^{-\deg P}$ and $\alpha(P^k) = a^k/(a;a)_k$, the q-binomial theorem gives

$$\sum_{k\geq 0} \chi(P^k)\alpha(P^k)u^{k\deg P} = \sum_{k\geq 0} \frac{(\chi(P)au^{\deg P})^k}{(a;a)_k} = \prod_{i\geq 1}\left(1 - \frac{\chi(P)u^{\deg P}}{q^{i\deg P}}\right)^{-1}.$$

Using this together with the fact that

$$\sum_{i\geq 1} \log\left(1 - \frac{\chi(P)u^{\deg P}}{q^{i\deg P}}\right) = \sum_{i\geq 1}\sum_{j\geq 1} \frac{(\chi(P)u^{\deg P})^j}{jq^{ij\deg P}} = \sum_{j\geq 1} \frac{\chi(P)^j u^{j\deg P}}{j(q^{j\deg P}-1)}$$

we get

$$\begin{aligned}\sum_{k\geq 0} \chi(P^k)\alpha(P^k)u^{k\deg P} = \prod_{i\geq 1}(1-\chi(P)(u/q^i)^{\deg P})^{-1} = \\ \exp\left(\sum_{j\geq 1} \frac{u^{j\deg P}}{j(q^{j\deg P}-1)}\chi(P)^j\right).\end{aligned} \tag{4.14}$$

In particular, the coefficients of $L_{\alpha,\leq m}(u,\chi)$ are then bounded from above by those of

$$\exp\left(\sum_{\deg P\leq m}\sum_{j\geq 1} \frac{u^{j\deg P}}{j(q^{j\deg P}-1)}\right) = \exp\left(\sum_{k\geq 1}\frac{u^k}{k(q^k-1)}\sum_{d|k,\, d\leq m} d\,|\mathcal{P}_{d,q}|\right).$$

From (4.14) we also have

$$L_{\alpha,>m}(u,\chi) = \exp\left(\sum_{k\geq 1}\frac{u^k}{k(q^k-1)}\sum_{m<d|k} d\sum_{P\in\mathcal{P}_{d,q}}\chi(P^{k/d})\,.\right), \tag{4.15}$$

Applying Lemma 4.2.4 and the bound $|\mathcal{P}_{n,q}| \leq q^n/n$ to

$$\sum_{m<d|k} d\sum_{P\in\mathcal{P}_{d,q}}\chi(P^{k/d}) = k\sum_{P\in\mathcal{P}_{d,q}}\chi(P) + \sum_{m<d|k,d\neq k}\sum_{P\in\mathcal{P}_{d,q}}\chi(P^{k/d})$$

we see that the coefficients of $L_{\alpha,>m}(u,\chi) - 1$ for non-principal χ are bounded from above by those of

$$\exp\left(\sum_{k>m}\frac{u^k}{k(q^k-1)}q^{k/2}(d(\chi)+4)\right) - 1\,.$$

Hence, by (4.13) and Lemma 4.2.5 with $R = 1$,

$$\left|S_\alpha(n,\chi) - S_\alpha(n,m,\chi)\right| \leq AB\,,$$

where

$$A := \exp\left(\sum_{k\geq 1}\frac{1}{k(q^k-1)}\sum_{d|k,\, d\leq m} d\,|\mathcal{P}_{d,q}|\right)$$

and

$$B := \exp\left(\sum_{k>m} \frac{1}{k(q^k-1)} q^{k/2}(d(\chi)+4)\right) - 1\,.$$

For the first factor the bound $|\mathcal{P}_{n,q}| \le q^n/n$ yields

$$\log A \le \sum_{d\le m} q^d \sum_{\substack{k\ge 1\\ d|k}} \frac{1}{k(q^k-1)} = \sum_{d\le m} q^d \left(\frac{1}{dq^d} + O\left(q^{-2d}\right)\right) \le \log m + O(1) \qquad (4.16)$$

since $\sum_{d\le m} 1/d = \log m + O(1)$ and $\sum_{d\le m} q^{-d} \le C$ independent of q. To bound the second factor, observe that $\sum_{k>m} 1/(kq^{k/2}) = O(q^{-m/2}/m)$, so for $m \ge 2\log_q d(\chi)$

$$B \le \exp\left(\sum_{k>m} \frac{2(d(\chi)+4)}{kq^{k/2}}\right) - 1 = O\left(\frac{d(\chi)+1}{mq^{m/2}}\right)\,.$$

Putting this together, we obtain

$$S_\alpha(n,\chi) - S_\alpha(n,m,\chi) = O\left((d(\chi)+1)q^{-\frac{m}{2}}\right)\,.$$

This proves the first part of Theorem 4.1.3.

To prove the second part, take $m = \lceil 2\log_q(d(\chi)+1) + \log n\rceil$. It then suffices to show that for χ_0 being the constant function 1 on $\mathcal{M}_q$

$$S_\alpha(n,m,\chi_0) = [u^n]L_{\alpha,\le m}(u,\chi_0) = o(1)$$

when $m = o(n)$. Differentiating the identity

$$\sum_{n\ge 0} S_\alpha(n,m,\chi_0)u^n = L_{\alpha,\le m}(u,\chi_0) = \exp\left(\sum_{k\ge 1} \frac{u^k}{k(q^k-1)} \sum_{m<d|k} d\,|\mathcal{P}_{q,d}|\right)$$

we obtain

$$\sum_{n\ge 1} nS_\alpha(n,m,\chi_0)u^{n-1} = \sum_{n\ge 0} S_\alpha(n,m,\chi_0)u^n \cdot \sum_{k\ge 1} \frac{u^{k-1}}{q^k-1} \sum_{m<d|k} d\,|\mathcal{P}_{q,d}|\,.$$

Comparing coefficients on both sides gives the following analogue of Hildebrand's identity [75, Eq. (2)]:

$$nS_\alpha(n,m,\chi_0) = \sum_{\substack{P\in\mathcal{P}_q,\,\deg P\le m\\ 1\le i\le n/\deg P}} \frac{S_\alpha(n-i\deg P,m,\chi_0)}{q^{i\deg P}-1}\deg P\,.$$

Since α is a probability measure when restricted to $\mathcal{M}_{k,q}$, we have $S_\alpha(k,m,\chi_0) \le 1$ for all k, hence

$$nS_\alpha(n,m,\chi_0) \le \sum_{\substack{P\in\mathcal{P}_q\\ \deg P\le m}} \deg P \sum_{i\ge 1} \frac{1}{q^{i\deg P}-1} \le C \sum_{\substack{P\in\mathcal{P}_q\\ \deg P\le m}} \frac{\deg P}{q^{\deg P}} = C\sum_{d=1}^{m} \frac{d}{q^d}\,|\mathcal{P}_{d,q}| \le Cm\,.$$

It now follows that $S_\alpha(n,m,\chi_0) = o(1)$ when $m = o(n)$. □

4.4 Discussion and prospective work

Most of the discussion provided in Section 3.5 is also relevant to this chapter.

In this chapter we considered a problem on the interface of random matrix theory and character sums. This is however not a new idea: characters sums over function fields, L-functions of the corresponding characters and their relation to the statistics of various compact groups were extensively studied by Katz and Sarnak [89]. There has been a lot of effort since then to prove analogues of results for Dirichlet characters in function field setting.

From the random matrix theory side, as mentioned in the introduction, $\mathrm{GL}_n(\mathbb{F}_q)$ is an analogue of $\mathrm{U}_n(\mathbb{C})$, but not the only one. One may also consider the orthogonal group

$$\mathrm{O}_n(\mathbb{F}_q) = \{X \in \mathrm{GL}_n(\mathbb{F}_q) : XX^\top = I_n\}$$

or the unitary group

$$\mathrm{U}_n(\mathbb{F}_q) = \{X \in \mathrm{GL}_n(\mathbb{F}_{q^2}) : XX^* = I_n\}\,,$$

where $\mathbb{F}_{q^2}$ is the extension of $\mathbb{F}_q$ of degree 2, and $x^* = x^q$. These cases were partially treated in [64], but only for $k \ll_q n$.

One may also ask if the range $\log k = o(n)$ in Theorems 4.1.1 and 4.1.2 is optimal for a single trace rather than the linear combination. The exponent killing all eigenvalues of $X \in \mathrm{GL}_n(\mathbb{F}_q)$ is

$$N_n = \prod_{j=1}^{n} \Phi_j(q) = q^{O(n^2)}\,,$$

where $\Phi_j(x)$ is the j-th cyclotomic polynomial. Thus one may ask how close can $\log k$ be to n^2. Heuristically, we may expect that the distribution of $\mathrm{Tr}(X^k)$ is asymptotically uniform as long as

$$\mathbb{P}\{P \in \mathcal{P}_{n,q} : P|f \Rightarrow \gcd(q^{\deg P}-1, k) \ne q^{o(\deg P)}\}$$

is small since the complement of this condition is essentially the local version of $\log k = o(n)$. Given the density of m-friable (smooth) polynomials, we expect the characteristic polynomial to have many factors of degree n^c and thus the range of k to extend to $\log k = o(n^2)$. If this is indeed true, then this would help us understand which Hayes characters have large sums and which are exceptionally nice in this regard.

Bibliography

[1] L.-P. Arguin, *Extrema of Log-correlated Random Variables*, Advances in disordered systems, random processes and some applications, (2016), p. 166.

[2] L.-P. Arguin and E. Bailey, *Large deviation estimates of Selberg's central limit theorem and applications*, arXiv preprint arXiv:2202.06799, (2022).

[3] L.-P. Arguin, D. Belius, P. Bourgade, M. Radziwiłł, and K. Soundararajan, *Maximum of the Riemann zeta function on a short interval of the critical line*, Communications on Pure and Applied Mathematics, 72 (2019), pp. 500–535.

[4] L.-P. Arguin, P. Bourgade, and M. Radziwiłł, *The Fyodorov-Hiary-Keating Conjecture. I*, arXiv preprint arXiv:2007.00988, (2020).

[5] L.-P. Arguin, F. Ouimet, and M. Radziwiłł, *Moments of the Riemann zeta function on short intervals of the critical line*, The Annals of Probability, 49 (2021), pp. 3106–3141.

[6] E. Bailey and J. Keating, *On the moments of the moments of $\zeta(1/2 + it)$*, Journal of Number Theory, (2021).

[7] E. C. Bailey and J. P. Keating, *Maxima of log-correlated fields: some recent developments*, Journal of Physics A: Mathematical and Theoretical, 55 (2022), p. 053001.

[8] S. Baluyot and B. Conrey, *Moments of zeta and correlations of divisor-sums: stratification and Vandermonde integrals*, arXiv preprint arXiv:2206.04821, (2022).

[9] S. Bettin, *The second moment of the Riemann zeta function with unbounded shifts*, International Journal of Number Theory, 6 (2010), pp. 1933–1944.

[10] S. Bettin, H. M. Bui, X. Li, and M. Radziwiłł, *A quadratic divisor problem and moments of the Riemann zeta-function*, Journal of the European Mathematical Society, 22 (2020), pp. 3953–3980.

[11] M. Bhargava, *The geometric sieve and the density of squarefree values of invariant polynomials*, arXiv preprint arXiv:1402.0031, (2014).

[12] ——, *A positive proportion of plane cubics fail the Hasse principle*, arXiv preprint arXiv:1402.1131, (2014).

[13] M. Bhargava, J. Cremona, and T. Fisher, *The proportion of plane cubic curves over $\mathbb{Q}$ that everywhere locally have a point*, International Journal of Number Theory, 12 (2016), pp. 1077–1092.

[14] ——, *The proportion of genus one curves over $\mathbb{Q}$ defined by a binary quartic that everywhere locally have a point*, International Journal of Number Theory, 17 (2021), pp. 903–923.

[15] M. Bhargava, J. E. Cremona, T. Fisher, N. G. Jones, and J. P. Keating, *What is the probability that a random integral quadratic form in n variables has an integral zero?*, International Mathematics Research Notices, 2016 (2016), pp. 3828–3848.

[16] M. Bhargava, D. M. Kane, H. Lenstra, B. Poonen, and E. Rains, *Modeling the distribution of ranks, Selmer groups, and Shafarevich-Tate groups of elliptic curves*, arXiv preprint arXiv:1304.3971, (2013).

[17] M. Bhargava, A. Shankar, and X. Wang, *Squarefree values of polynomial discriminants I*, arXiv preprint arXiv:1611.09806, (2016).

[18] ——, *Squarefree values of polynomial discriminants II*, 2022.

[19] A. Bhowmick and T. H. Lê, *On primitive elements in finite fields of low characteristic*, Finite Fields and their Applications, 35 (2015), pp. 64–77.

[20] A. Bhowmick, T. H. Lê, and Y.-R. Liu, *A note on character sums in finite fields*, Finite Fields and their Applications, 46 (2017), pp. 247–254.

[21] P. Bourgade, *Mesoscopic fluctuations of the zeta zeros*, Probability Theory and Related Fields, 148 (2010), pp. 479–500.

[22] J. Bourgain, *Decoupling, exponential sums and the Riemann zeta function*, Journal of the American Mathematical Society, 30 (2017), pp. 205–224.

[23] J. Bourgain, A. Gamburd, and P. Sarnak, *Affine linear sieve, expanders, and sum-product*, Inventiones mathematicae, 179 (2010), pp. 559–644.

[24] R. P. Brent and B. D. McKay, *Determinants and ranks of random matrices over Z_m*, Discrete Mathematics, 66 (1987), pp. 35–49.

[25] ——, *On determinants of random symmetric matrices over Z_m*, Ars Combinatoria A, 26 (1988), pp. 57–64.

[26] W. C. Brown, *Matrices over commutative rings*, Marcel Dekker, Inc., 1993.

[27] H. M. Bui, B. Conrey, and M. P. Young, *More than 41% of the zeros of the zeta function are on the critical line*, Acta Arithmetica, 150 (2011), pp. 35–64.

[28] L. Carlitz, *Representations by skew forms in a finite field*, Archiv der Mathematik, 5 (1954), pp. 19–31.

[29] J. W. S. Cassels, *Rational quadratic forms*, Courier Dover Publications, 2008.

[30] V. Chandee, *On the correlation of shifted values of the Riemann zeta function*, Quarterly journal of mathematics, 62 (2011), pp. 545–572.

[31] V. Chandee and X. Li, *The eighth moment of Dirichlet L-functions*, Advances in Mathematics, 259 (2014), pp. 339–375.

[32] J. Clancy, N. Kaplan, T. Leake, S. Payne, and M. M. Wood, *On a Cohen–Lenstra heuristic for Jacobians of random graphs*, Journal of Algebraic Combinatorics, 42 (2015), pp. 701–723.

[33] B. Conrey and J. P. Keating, *Moments of zeta and correlations of divisor-sums: I*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373 (2015), p. 20140313.

[34] ——, *Moments of zeta and correlations of divisor-sums: II*, in Advances in the Theory of Numbers, Springer, 2015, pp. 75–85.

[35] ——, *Moments of zeta and correlations of divisor-sums: III*, Indagationes Mathematicae, 26 (2015), pp. 736–747.

[36] ——, *Moments of zeta and correlations of divisor-sums: IV*, Research in Number Theory, 2 (2016), pp. 1–24.

[37] ——, *Moments of zeta and correlations of divisor-sums: V*, Proceedings of the London Mathematical Society, 118 (2019), pp. 729–752.

[38] J. B. Conrey, *A note on the fourth power moment of the Riemann zeta-function*, in Analytic Number Theory, Springer, 1996, pp. 225–230.

[39] J. B. Conrey, D. W. Farmer, J. P. Keating, M. O. Rubinstein, and N. C. Snaith, *Integral moments of L-functions*, Proceedings of the London Mathematical Society, 91 (2005), pp. 33–104.

[40] ——, *Lower order terms in the full moment conjecture for the Riemann zeta function*, Journal of Number Theory, 128 (2008), pp. 1516–1554.

[41] J. B. Conrey and A. Ghosh, *A conjecture for the sixth power moment of the Riemann zeta-function*, arXiv preprint math/9807187, (1998).

[42] J. B. Conrey and S. M. Gonek, *High moments of the Riemann zeta-function*, Duke Mathematical Journal, 107 (2001), pp. 577–604.

[43] J. B. Conrey, H. Iwaniec, and K. Soundararajan, *The sixth power moment of Dirichlet L-functions*, Geometric and Functional Analysis, 22 (2012), pp. 1257–1288.

[44] K. Courteaut, K. Johansson, and G. Lambert, *From Berry-Esseen to super-exponential*, arXiv preprint arXiv:2204.03282, (2022).

[45] P. Deligne, *Formes modulaires et representations e-adiques*, in Séminaire Bourbaki vol. 1968/69 Exposés 347-363, Springer, 1971, pp. 139–172.

[46] ——, *La conjecture de Weil. I*, Publications Mathématiques de l'Institut des Hautes Études Scientifiques, 43 (1974), pp. 273–307.

[47] J.-M. Deshouillers and H. Iwaniec, *Kloosterman sums and Fourier coefficients of cusp forms*, Inventiones mathematicae, 70 (1982), pp. 219–288.

[48] K. Destagnol and E. Sofos, *Rational points and prime values of polynomials in moderately many variables*, Bulletin des Sciences Mathématiques, 156 (2019), p. 102794.

[49] P. Diaconis and M. Shahshahani, *On the eigenvalues of random matrices*, Journal of Applied Probability, 31 (1994), pp. 49–62.

[50] F. J. Dyson, *Statistical theory of the energy levels of complex systems. I*, Journal of Mathematical Physics, 3 (1962), pp. 140–156.

[51] T. Ekedahl, *An infinite version of the Chinese remainder theorem*, Rikkyo Daigaku sugaku zasshi, 40 (1991), pp. 53–59.

[52] J. S. Ellenberg, S. Jain, and A. Venkatesh, *Modeling λ-invariants by p-adic random matrices*, Communications on Pure and Applied Mathematics, 64 (2011), pp. 1243–1262.

[53] S. N. Evans, *Local fields, Gaussian measures, and Brownian motions*, Topics in probability and Lie groups: boundary theory, 28 (2001), pp. 11–50.

[54] ——, *Elementary divisors and determinants of random matrices over a local field*, Stochastic processes and their applications, 102 (2002), pp. 89–102.

[55] B. Fisher, *A note on Hensel's lemma in several variables*, Proceedings of the American Mathematical Society, 125 (1997), pp. 3185–3189.

[56] E. Friedman and L. C. Washington, *On the distribution of divisor class groups of curves over a finite field*, Théorie des Nombres/Number Theory Laval, (1987).

[57] J. Fulman, *Random matrix theory over finite fields*, Bulletin of the American Mathematical Society, 39 (2002), pp. 51–85.

[58] J. Fulman and L. Goldstein, *Stein's method and the rank distribution of random matrices over finite fields*, The Annals of Probability, 43 (2015), pp. 1274–1314.

[59] J. Fulman and N. Kaplan, *Random Partitions and Cohen–Lenstra Heuristics*, Annals of Combinatorics, 23 (2019), pp. 295–315.

[60] Y. V. Fyodorov, G. A. Hiary, and J. P. Keating, *Freezing transition, characteristic polynomials of random matrices, and the Riemann zeta function*, Physical review letters, 108 (2012), p. 170601.

[61] Y. V. Fyodorov and J. P. Keating, *Freezing transitions and extreme values: random matrix theory, and disordered landscapes*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372 (2014), p. 20120503.

[62] A. Gamburd, *Some applications of symmetric functions theory in random matrix theory*, London Mathematical Society Lecture Note Series, 341 (2007), p. 143.

[63] M. Gerstenhaber, *On the number of nilpotent matrices with coefficients in a finite field*, Illinois Journal of Mathematics, 5 (1961), pp. 330–333.

[64] O. Gorodetsky and B. Rodgers, *Traces of powers of matrices over finite fields*, Transactions of the American Mathematical Society, 374 (2021), pp. 4579–4638.

[65] A. Granville and K. Soundararajan, *Large character sums*, Journal of the American Mathematical Society, 14 (2001), pp. 365–397.

[66] G. H. Hardy and J. E. Littlewood, *Contributions to the theory of the Riemann zeta-function and the theory of the distribution of primes*, Acta Mathematica, 41 (1916), pp. 119–196.

[67] A. J. Harper, *A note on the maximum of the Riemann zeta function, and log-correlated random variables*, arXiv preprint arXiv:1304.0677, (2013).

[68] ——, *Sharp conditional bounds for moments of the Riemann zeta function*, arXiv preprint arXiv:1305.4618, (2013).

[69] ——, *On the partition function of the Riemann zeta function, and the Fyodorov–Hiary–Keating conjecture*, arXiv preprint arXiv:1906.05783, (2019).

[70] ——, *Almost sure large fluctuations of random multiplicative functions*, arXiv preprint arXiv:2012.15809, (2020).

[71] D. R. Hayes, *The distribution of irreducibles in $GF[q, x]$*, Transactions of the American Mathematical Society, 117 (1965), pp. 101–127.

[72] W. Heap and K. Soundararajan, *Lower bounds for moments of zeta and L-functions revisited*, arXiv preprint arXiv:2007.13154, (2020).

[73] D. Heath-Brown, *The fourth power moment of the Riemann zeta function*, Proceedings of the London Mathematical Society, 3 (1979), pp. 385–422.

[74] ——, *Fractional Moments of the Riemann Zeta-Function*, Journal of the London Mathematical Society, 2 (1981), pp. 65–78.

[75] A. Hildebrand, *On the number of positive integers $\leq x$ and free of prime factors $> y$*, Journal of Number Theory, 22 (1986), pp. 289–307.

[76] C. Hooley, *On the power free values of polynomials*, Mathematika, 14 (1967), pp. 21–26.

[77] B. Hough, *The angle of large values of L-functions*, Journal of Number Theory, 167 (2016), pp. 353–393.

[78] C.-N. Hsu, *Estimates for coefficients of L-functions for function fields*, Finite Fields and their Applications, 5 (1999), pp. 76–88.

[79] C. P. Hughes and M. P. Young, *The twisted fourth moment of the Riemann zeta function*, Journal für die reine undangewandte Mathematik, (2010).

[80] M. Huxley, *On stationary phase integrals*, Glasgow Mathematical Journal, 36 (1994), pp. 355–362.

[81] A. E. Ingham, *Mean-value theorems in the theory of the Riemann zeta-function*, Proceedings of the London Mathematical Society, 2 (1928), pp. 273–300.

[82] A. Ivić and Y. Motohashi, *The Mean Square of the Error Term for the Fourth Power Moment of the Zeta-Function*, Proceedings of the London Mathematical Society, 3 (1994), pp. 309–329.

[83] ——, *On some estimates involving the binary additive divisor problem*, The Quarterly Journal of Mathematics, 46 (1995), pp. 471–483.

[84] H. Iwaniec, *Spectral methods of automorphic forms*, vol. 53, American Mathematical Society, Revista Matemática Iberoamericana (RMI . . . , 2021.

[85] K. Johansson, *On random matrices from the compact classical groups*, Annals of mathematics, 145 (1997), pp. 519–545.

[86] K. Johansson and G. Lambert, *Multivariate normal approximation for traces of random unitary matrices*, The Annals of Probability, 49 (2021), pp. 2961–3010.

[87] M. Jutila, *The spectral mean square of Hecke L-functions on the critical line*, Publications de l'Institut Mathematique, 76 (2004), pp. 41–55.

[88] M. Jutila and Y. Motohashi, *Uniform bound for Hecke L-functions*, Acta Mathematica, 195 (2005), pp. 61–115.

[89] N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, vol. 45, American Mathematical Soc., 1999.

[90] J. Keating and Z. Rudnick, *Squarefree polynomials and Möbius values in short intervals and arithmetic progressions*, Algebra & Number Theory, 10 (2016), pp. 375–420.

[91] J. P. Keating and Z. Rudnick, *The variance of the number of prime polynomials in short intervals and in residue classes*, International Mathematics Research Notices, 2014 (2014), pp. 259–288.

[92] J. P. Keating and N. C. Snaith, *Random matrix theory and $\zeta(1/2 + it)$*, Communications in Mathematical Physics, 214 (2000), pp. 57–89.

[93] H. Kim, *Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2*, Journal of the American Mathematical Society, 16 (2003), pp. 139–183.

[94] O. Klurman, A. P. Mangerel, and J. Teräväinen, *Correlations of multiplicative functions in function fields*, arXiv preprint arXiv:2009.13497, (2020).

[95] V. Kovaleva, *A note on the distribution of equivalence classes of random symmetric p-adic matrices*, arXiv preprint arXiv:2008.10732, (2020).

[96] E. Kowalski, P. Michel, and J. VanderKam, *Mollification of the fourth moment of automorphic L-functions and arithmetic applications*, Inventiones mathematicae, 142 (2000), pp. 95–151.

[97] I. G. Macdonald, *Symmetric functions and Hall polynomials*, Oxford university press, 1998.

[98] K. Maples, *A Brun-Titchmarsh inequality for discrete random matrices*, Submitted, November, (2010).

[99] ——, *Cokernels of random matrices satisfy the Cohen-Lenstra heuristics*, arXiv preprint arXiv:1301.1239, (2013).

[100] K. McCurley, *Relatively prime values of polynomials*, A Tribute to Emil Grosswald: Number Theory and Related Analysis: Number Theory and Related Analysis, 143 (1993), p. 437.

[101] H. L. Montgomery, *The pair correlation of zeros of the zeta function*, in Proceedings of Symposia in Pure Mathematics, vol. 24, 1973, pp. 181–193.

[102] H. L. Montgomery and R. C. Vaughan, *Exponential sums with multiplicative coefficients*, Inventiones Mathematicae, 43 (1977), pp. 69–82.

[103] C. Moreno and F. Shahidi, *The L-functions $L(s, Sym^m(r), \pi)$*, Canadian Mathematical Bulletin, 28 (1985), pp. 405–410.

[104] Y. Motohashi, *Spectral mean values of Maass waveform L-functions*, Journal of Number theory, 42 (1992), pp. 258–284.

[105] ——, *An explicit formula for the fourth power mean of the Riemann zeta-function*, Acta Mathematica, 170 (1993), pp. 181–220.

[106] ——, *The binary additive divisor problem*, in Annales scientifiques de l'Ecole normale supérieure, vol. 27, 1994, pp. 529–572.

[107] M. R. Murty, *On the estimation of eigenvalues of Hecke operators*, The Rocky Mountain Journal of Mathematics, 15 (1985), pp. 521–533.

[108] J. Najnudel, *On the extreme values of the Riemann zeta function on random intervals of the critical line*, Probability Theory and Related Fields, 172 (2018), pp. 387–452.

[109] N. Ng, *The sixth moment of the Riemann zeta function and ternary additive divisor sums*, arXiv preprint arXiv:1610.04977, (2016).

[110] N. Ng, Q. Shen, and P.-J. Wong, *The eighth moment of the Riemann zeta function*, arXiv preprint arXiv:2204.13891, (2022).

[111] ——, *Shifted moments of the Riemann zeta function*, arXiv preprint arXiv:2206.03350, (2022).

[112] H. H. Nguyen and M. M. Wood, *Random integral matrices: universality of surjectivity and the cokernel*, arXiv preprint arXiv:1806.00596, (2018).

[113] J. Park, B. Poonen, J. Voight, and M. M. Wood, *A heuristic for boundedness of ranks of elliptic curves*, Journal of the European Mathematical Society, 21 (2019), pp. 2859–2903.

[114] B. Poonen, *Squarefree values of multivariable polynomials*, Duke Mathematical Journal, 118 (2003), pp. 353–373.

[115] ——, *Heuristics for the arithmetic of elliptic curves*, in Proceedings of the International Congress of Mathematicians-Rio de, vol. 2, World Scientific, 2018, pp. 399–414.

[116] B. Poonen and E. Rains, *Random maximal isotropic subspaces and Selmer groups*, Journal of the American Mathematical Society, 25 (2012), pp. 245–269.

[117] B. Poonen and M. Stoll, *The Cassels-Tate pairing on polarized abelian varieties*, Annals of Mathematics, (1999), pp. 1109–1149.

[118] M. Radziwiłł and K. Soundararajan, *Continuous lower bounds for moments of zeta and L-functions*, Mathematika, 59 (2013), pp. 119–128.

[119] ——, *Selberg's central limit theorem for* $\log|\zeta(1/2+it)|$, L'enseignement Mathematique, 63 (2018), pp. 1–19.

[120] E. M. Rains, *High powers of random elements of compact Lie groups*, Probability Theory and Related Fields, 107 (1997), pp. 219–241.

[121] I. Reiner, *On the number of matrices with given characteristic polynomial*, Illinois Journal of Mathematics, 5 (1961), pp. 324–329.

[122] G. Rhin, *Répartition modulo 1 dans un corps de séries formelles sur un corps fini*, Instytut Matematyczny Polskiej Akademi Nauk, 1972.

[123] M. Rosen, *Number theory in function fields*, vol. 210 of Graduate Texts in Mathematics, Springer-Verlag, New York, 2002.

[124] A. Selberg, *Contribution to the theory of the Riemann zeta-function*, Arch. for Math. og Naturv. B, (1946).

[125] ——, *Old and new conjectures and results about a class of Dirichlet series*, in Proceedings of the Amalfi Conference on Analytic Number Theory (Maiori, 1989), vol. 2, 1992, pp. 47–63.

[126] K. Soundararajan, *Moments of the Riemann zeta function*, Annals of Mathematics, (2009), pp. 981–993.

[127] T. Tao, *Lecture notes 8 for 247B*, URL https://www.math.ucla.edu/ tao/247b.1.07w/notes8.pdf, 1 (2007).

[128] E. C. Titchmarsh and D. R. Heath-Brown, *The theory of the Riemann zeta-function*, Oxford university press, 1986.

[129] R. Van Peski, *Limits and fluctuations of p-adic random matrix products*, Selecta Mathematica, 27 (2021), pp. 1–71.

[130] Y. Wang and R. P. Stanley, *The Smith normal form distribution of a random integer matrix*, SIAM Journal on Discrete Mathematics, 31 (2017), pp. 2247–2268.

[131] N. Watt, *Kloosterman sums and a mean value for Dirichlet polynomials*, Journal of Number Theory, 53 (1995), pp. 179–210.

[132] M. Wood, *The distribution of sandpile groups of random graphs*, Journal of the American Mathematical Society, 30 (2017), pp. 915–958.

[133] M. M. Wood, *Random integral matrices and the Cohen-Lenstra heuristics*, American Journal of Mathematics, 141 (2019), pp. 383–398.

[134] R. Zacharias, *Mollification of the fourth moment of Dirichlet L-functions*, arXiv preprint arXiv:1611.09582, (2016).