

BOUNDARY EFFECT ON THE NODAL LENGTH FOR ARITHMETIC RANDOM WAVES, AND SPECTRAL SEMI-CORRELATIONS

VALENTINA CAMMAROTA¹, OLEKSIY KLURMAN², AND IGOR WIGMAN³

Dedicated to the memory of Jean Bourgain

ABSTRACT. We test M. Berry’s ansatz on nodal deficiency in presence of boundary. The square billiard is studied, where the high spectral degeneracies allow for the introduction of a Gaussian ensemble of random Laplace eigenfunctions (“boundary-adapted arithmetic random waves”). As a result of a precise asymptotic analysis, two terms in the asymptotic expansion of the expected nodal length are derived, in the high energy limit along a generic sequence of energy levels. It is found that the precise nodal deficiency or surplus of the nodal length depends on arithmetic properties of the energy levels, in an explicit way.

To obtain the said results we apply the Kac-Rice method for computing the expected nodal length of a Gaussian random field. Such an application uncovers major obstacles, e.g. the occurrence of “bad” subdomains, that, one hopes, contribute insignificantly to the nodal length. Fortunately, we were able to reduce this contribution to a number theoretic question of counting the “spectral semi-correlations”, a concept joining the likes of “spectral correlations” and “spectral quasi-correlations” in having impact on the nodal length for arithmetic dynamical systems.

This work rests on several breakthrough techniques of J. Bourgain, whose interest in the subject helped shaping it to high extent, and whose fundamental work on spectral correlations, joint with E. Bombieri, has had a crucial impact on the field.

1. INTRODUCTION

1.1. Nodal length of Laplace eigenfunctions. Let $f : \mathcal{M} \rightarrow \mathbb{R}$ be a smooth function on a compact smooth Riemannian surface (\mathcal{M}, g) , with or without boundary, with no critical zeros. The zero set of f , called the *nodal line* is a *smooth* curve with no self-intersections; it is an important qualitative descriptor of f . We are interested in the geometry of the nodal lines of Laplace eigenfunctions on \mathcal{M} , in the high energy limit, i.e. the solutions $\{(\varphi_j, \lambda_j)\}_{j \geq 1}$ of the Schrödinger equation

$$(1.1) \quad \Delta \varphi_j + \lambda_j \varphi_j = 0,$$

satisfying either Dirichlet or Neumann boundary conditions, where $\Delta = \operatorname{div} \circ \nabla$ is the Laplace-Beltrami (Laplacian) operator on \mathcal{M} , and $\lambda_j \geq 0$ are the energy levels (or simply the energies). It is well-known that the spectrum of Δ is purely discrete, i.e. there exists a complete orthonormal system $\{\varphi_j\}_{j \geq 1}$ (orthonormal basis), spanning the whole of $L^2(\mathcal{M})$, so that all the spectral multiplicities are finite, and $\lambda_j \rightarrow \infty$ as $j \rightarrow \infty$ being the high energy limit.

Much of the focus in the study of the nodal lines of Laplace eigenfunctions has been turned to the study of the *nodal length*, i.e. the total length $\mathcal{L}(\varphi_j)$ of the curve $\varphi_j^{-1}(0)$ on \mathcal{M} , as $j \rightarrow \infty$. Yau’s conjecture [27] asserts that the nodal length is commensurable with $\sqrt{\lambda_j}$, i.e.

$$(1.2) \quad c_{\mathcal{M}} \cdot \sqrt{\lambda_j} \leq \mathcal{L}(\varphi_j) \leq C_{\mathcal{M}} \cdot \sqrt{\lambda_j}$$

for some positive constants $c_{\mathcal{M}}, C_{\mathcal{M}} > 0$. Yau’s conjecture was proven [6, 7, 11] for \mathcal{M} analytic, and more recently the optimal lower bound [18] and polynomial upper bound [19] were established for the more general, smooth, case (see also [17]).

¹DEPARTMENT OF STATISTICS, SAPIENZA UNIVERSITY OF ROME

²DEPARTMENT OF MATHEMATICS, KTH ROYAL INSTITUTE OF TECHNOLOGY, STOCKHOLM

³DEPARTMENT OF MATHEMATICS, KING’S COLLEGE LONDON

E-mail addresses: valentina.cammarota@uniroma1.it, lklurman@gmail.com,

igor.wigman@kcl.ac.uk.

Date: May 24, 2019.

1.2. Nodal length for random fields. One way to obtain stronger (or more precise) results than (1.2) is to study the nodal length $\mathcal{L}(f)$ of *random* functions f , an approach that has been actively pursued, in particular in the recent few years. As a concrete direction of research within the indicated scope, one may take a Gaussian random field $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ (or $f : \mathbb{R}^d \rightarrow \mathbb{R}$, $d \geq 2$) and study the distribution of the nodal length $\mathcal{L}(f; R)$ of f restricted to $B(R) \subseteq \mathbb{R}^2$, the radius- R centred ball, $R \rightarrow \infty$. For¹ $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ *stationary* a straightforward application of the standard Kac-Rice formula yields a precise expression

$$(1.3) \quad \mathbb{E}[\mathcal{L}(f; R)] = c \cdot \text{Vol}(B(R)),$$

whereas a significantly heavier machinery involving perturbation theory (and asymptotic analysis of the 2-point correlation function) yields an asymptotic expression for the variance

$$\text{Var}(\mathcal{L}(f; R)),$$

as $R \rightarrow \infty$. One may go further by applying the Wiener chaos decomposition on $\mathcal{L}(f; R)$ to obtain [15] a limit law for the distribution of

$$\frac{\mathcal{L}(f; R)}{\sqrt{\text{Var}(\mathcal{L}(f; R))}}.$$

Alternatively to working with a fixed random field restricted to expanding balls, one may fix a compact surface \mathcal{M} , consider a Gaussian ensemble of random functions on \mathcal{M} , i.e. a sequence $f_n : \mathcal{M} \rightarrow \mathbb{R}$ of Gaussian random fields indexed by \mathcal{M} , and study the asymptotic distribution of the nodal length of f_n , that is the total length $\mathcal{L}(f_n)$ of $f_n^{-1}(0)$, as $n \rightarrow \infty$; in some natural examples (to be discussed below) f_n possesses a natural scaling with n .

Berry [2] suggested that for \mathcal{M} generic *chaotic*, there exists a (non-rigorous) *link* between the (deterministic) eigenfunctions φ_j as in (1.1), and the restriction of monochromatic isotropic random wave g (a particular random field on \mathbb{R}^2 to be defined immediately below), to $B(R)$ with $R \approx \sqrt{\lambda_j}$; this vague relation, usually referred to as ‘‘Berry’s Random Wave Model’’ (RWM), agreed in a wide community, is subject to many numerical tests with overwhelmingly positive outcomes. In particular, the study of the nodal structures of g restricted to $B(R)$ as $R \rightarrow \infty$ facilitates our understanding of the nodal structures of φ_j in the high energy limit. Berry’s monochromatic isotropic random wave g is uniquely defined as the centred Gaussian random field on \mathbb{R}^2 with covariance function

$$r_g(x - y) = r_g(x, y) := \mathbb{E}[g(x) \cdot g(y)] = J_0(|x - y|),$$

$x, y \in \mathbb{R}^2$, whose Fourier transform on \mathbb{R}^2 is the arc length of the unit circle (meaning that the monochromatic waves are propagating uniformly in all directions). Since r_g depends only on the Euclidean distance $|x - y|$, the law of g is invariant under all translations $g(\cdot) \mapsto g(\cdot + z)$, $z \in \mathbb{R}^2$, and rotations $g(\cdot) \mapsto g(o \cdot)$, $o \in O(2)$ (i.e. g is stationary isotropic); it has applications in the study of ocean waves propagating [20, 21].

Consistent to the above (1.3), the expected nodal length for this stationary model is easily found to be

$$(1.4) \quad \mathbb{E}[\mathcal{L}(g; R)] = c_0 \cdot \text{Vol}(B(R)),$$

with $c_0 > 0$ explicitly evaluated, via a straightforward application of the Kac-Rice formula. Berry [3, Formula (28)] further found that the variance is *logarithmic*, i.e. satisfying the asymptotic law

$$(1.5) \quad \text{Var}(\mathcal{L}(g; R)) = c_1 \cdot R^2 \log R + O_{R \rightarrow \infty}(R^2),$$

much smaller than one would expect, due to an ‘‘obscure cancellation’’ (see also [26]).

The central objective of this manuscript is investigating the effect of nontrivial boundary on the nodal structures of Laplace eigenfunctions, first and foremost on the nodal length, either in the vicinity of the boundary, or globally. Berry argued that, since the nodal line is perpendicular to

¹From this point we will tacitly assume that all the involved random fields are sufficiently smooth and are satisfying some non-degeneracy assumptions.

the boundary [9] (except for intersection points with higher degree vanishing), its presence should impact its length negatively compared to (1.4), he referred to as “nodal deficiency”. He backed this ansatz by a precise evaluation of the secondary term around the boundary for the “boundary-adapted random waves”, a Gaussian random field constrained to satisfy the boundary conditions, Dirichlet or Neumann, on an infinite straight line.

It was concluded that, bearing in mind that the primary term in his asymptotic expansion of nodal length for this boundary-adapted case is consistent to (1.4), whereas the secondary term was, a large number of wavelengths away from the boundary, negative (identical between Dirichlet and Neumann boundary conditions), with total contribution in absolute value larger than the length fluctuations in (1.5) (possibly extending to the boundary-adapted case), based on one sample only, one should be able to detect the deficiency of the total nodal length of Laplace eigenfunctions for surfaces with boundary compared to the boundary-less case. Gnuzmann and Lois [12] supported Berry’s deficiency ansatz by performing a mean nodal volume calculation for \mathcal{M} cuboid of arbitrarily high dimension, a dynamical system with separation of variables, while averaging w.r.t. energy levels (rather than w.r.t. a Gaussian ensemble).

1.3. Arithmetic Random Waves. The “usual” Arithmetic Random Waves are random toral Laplace eigenfunctions. Let

$$(1.6) \quad S = \{a^2 + b^2 : a, b \in \mathbb{Z}\}$$

be the set of all integers expressible as sum of two squares, $n \in S$, and

$$(1.7) \quad \mathcal{E}_n = \{\mu = (\mu_1, \mu_2) \in \mathbb{Z}^2 : \mu_1^2 + \mu_2^2 = n\}$$

the lattice points set lying on the centred radius- \sqrt{n} circle. It is well-known that $n \in S$, if and only if the prime decomposition of n is of the form

$$(1.8) \quad n = 2^a \cdot \prod_{j=1}^s p_j^{e_j} \prod_{k=1}^r q_k^{2h_k},$$

for some nonnegative integers a , $\{e_j\}_{j \leq s}$, $\{h_k\}_{k \leq r}$, and $p_j \equiv 1 \pmod{4}$ and $q \equiv 3 \pmod{4}$ are primes.

By a classical result due to E. Landau [16] the sequence $S \subseteq \mathbb{Z}$ is thin, i.e. of asymptotic density 0, and, what is stronger,

$$\frac{|\{n \leq X : n \in S\}|}{X} \underset{X \rightarrow \infty}{\sim} c_0 \cdot \frac{X}{\sqrt{\log X}},$$

with some semi-explicit constant $c_0 > 0$. Any function $g_n : \mathbb{T}^2 \rightarrow \mathbb{R}$ on the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ of the form

$$(1.9) \quad g_n(x) = \frac{1}{\sqrt{N_n}} \sum_{\mu \in \mathcal{E}_n} a_\mu e(\langle x, \mu \rangle),$$

where $a_\mu \in \mathbb{C}$ are some constants satisfying the condition

$$(1.10) \quad a_{-\mu} = \overline{a_\mu}$$

and $N_n = |\mathcal{E}_n|$ is the size of the lattice points set \mathcal{E}_n (equivalently, $N_n = r_2(n)$, the number of ways to express n as a sum of two squares), is a real-valued Laplace eigenfunction with eigenvalue

$$(1.11) \quad \lambda'_n = 4\pi^2 n.$$

The convenience pre-factor $\frac{1}{\sqrt{N_n}}$ on the r.h.s. of (1.9) has no bearing on the nodal set of g_n , and will be understood below. Conversely, any real-valued Laplace eigenfunction on \mathbb{T}^2 is necessarily of the form (1.9) for some $n \in S$.

The linear space of functions (1.9) may be endowed with a probability measure by making the coefficients $\{a_\mu\}_{\mu \in \mathcal{E}_n}$ i.i.d. standard complex valued Gaussian random variables, save for the condition

(1.10) to ensure the g_n are real-valued; this model is called ‘‘Arithmetic Random Waves’’. Alternatively and equivalently, Arithmetic Random Waves is the Gaussian ensemble of centred stationary random fields with the covariance functions

$$(1.12) \quad p_n(x) = \mathbb{E}[g_n(x) \cdot g_n(0)] = \frac{1}{N_n} \sum_{\mu \in \mathcal{E}_n} \cos(2\pi \langle x, \mu \rangle);$$

its (random) nodal length $\mathcal{L}_n = \mathcal{L}(g_n)$ on \mathbb{T}^2 is our etalon, representing the boundary-less cases for comparison against the appearance of nontrivial boundary. As it is the case with stationary random fields, it is easy to evaluate its expected nodal length to be

$$(1.13) \quad \mathbb{E}[\mathcal{L}_n] = \frac{\sqrt{\lambda'_n}}{2\sqrt{2}\sqrt{n}} = \frac{\pi}{\sqrt{2}} \cdot \sqrt{n}.$$

Rudnick and Wigman [23] gave the useful upper bound

$$(1.14) \quad \text{Var}(\mathcal{L}_n) \ll_{N_n \rightarrow \infty} \frac{n}{\sqrt{N_n}},$$

showing, in particular, that the distribution of $\frac{\mathcal{L}_n}{\sqrt{n}}$ concentrates around the constant $\frac{\pi}{\sqrt{2}}$.

Krishnapur-Kurlberg-Wigman [13] further resolved the question of the true asymptotic behaviour of the variance on the l.h.s. of (1.14), requiring the following background in the two squares problem. For every n we define the atomic probability measure

$$(1.15) \quad \nu_n = \frac{1}{N_n} \sum_{\mu \in \mathcal{E}_n} \delta_{\mu/\sqrt{n}}$$

on the unit circle \mathcal{S}^1 , supported on the angles of \mathcal{S}^1 corresponding to points of \mathcal{E}_n . It is known that for a ‘‘generic’’ sequence $\{n\} \subseteq S$ the angles $\{\mu/\sqrt{n}\}_{\mu \in \mathcal{E}_n}$ equidistribute on \mathcal{S}^1 , i.e. there exists a relative density 1 sequence $\{n\} \subseteq S$, so that

$$(1.16) \quad \nu_n \Rightarrow \frac{d\theta}{2\pi},$$

with ‘ \Rightarrow ’ standing for the weak- $*$ convergence of probability measures on \mathcal{S}^1 , and in particular $N_n \rightarrow \infty$. However, there exist [10, 13, 14] other *attainable* measures, i.e. weak- $*$ partial limits of the sequence $\{\nu_n\}_{n \in S}$, and even under the (generic) constraint $N_n \rightarrow \infty$, the accumulation set of the sequence $\{\widehat{\nu}_n(4)\}$ of the 4th Fourier coefficients of ν_n (being the first nontrivial Fourier coefficient) is the whole of $[-1, 1]$. The said work [13] established the *precise* asymptotic relation

$$(1.17) \quad \text{Var}(\mathcal{L}_n) \underset{N_n \rightarrow \infty}{\sim} \frac{\pi^2}{128} (1 + \widehat{\nu}_n(4)^2) \cdot \frac{n}{N_n^2},$$

and, bearing in mind that $1 + \widehat{\nu}_n(4)^2$ is bounded away from both 0 and infinity, in particular, it shows that the fluctuations around the mean of \mathcal{L}_n are of the order of magnitude

$$(1.18) \quad \mathcal{L}_n - \mathbb{E}[\mathcal{L}_n] \approx \frac{\sqrt{n}}{N_n},$$

important below, just like (1.5), due to an unexpected cancellation (‘‘arithmetic Berry’s cancellation’’). Later a non-universal limit theorem for \mathcal{L}_n was obtained [22].

1.4. Boundary-adapted Arithmetic Random Waves. The boundary-adapted Arithmetic Random Waves are random Laplace eigenfunctions on the unit square $\mathcal{Q} = [0, 1]^2$, subject to Dirichlet boundary conditions. Let S be the as above (1.6), $n \in S$, and $\mu = (\mu_1, \mu_2) \in \mathcal{E}_n$ a lattice point with \mathcal{E}_n given by (1.7). Any function of the form $\mathcal{Q} \rightarrow \mathbb{R}$

$$(1.19) \quad x = (x_1, x_2) \mapsto \sin(\pi\mu_1 x_1) \cdot \sin(\pi\mu_2 x_2)$$

is a Laplace eigenfunction with eigenvalue

$$(1.20) \quad \lambda_n = \pi^2 n,$$

satisfying the Dirichlet boundary conditions on \mathcal{Q} (cf. (1.11)). However, given $\mu = (\mu_1, \mu_2) \in \mathcal{E}_n$ and $\mu' = (\pm\mu_1, \pm\mu_2) \in \mathcal{E}_n$, the resulting maps as in (1.19) differ at most by a sign. Therefore, to avoid redundancies, we introduce the equivalence relation on \mathcal{E} : $(\mu_1, \mu_2) \sim (\mu'_1, \mu'_2)$ if $\mu_1 = \pm\mu'_1$ and $\mu_2 = \pm\mu'_2$.

The general form of a Laplace eigenfunction on \mathcal{Q} satisfying Dirichlet boundary conditions assumes the form

$$(1.21) \quad f_n(x) = \frac{4}{\sqrt{N_n}} \sum_{\mu \in \mathcal{E}_n / \sim} a_\mu \sin(\pi\mu_1 x_1) \sin(\pi\mu_2 x_2),$$

for some $n \in S$, and we endow this linear space with a Gaussian probability measure by making the $\{a_\mu\}_{\mu \in \mathcal{E}_n / \sim}$ i.i.d. standard (real) Gaussian random variables. If either $\mu_1 = 0$ or $\mu_2 = 0$, then the corresponding summand in (1.21) vanishes, so that we are allowed to assume that both $\mu_1 \neq 0$ and $\mu_2 \neq 0$. We then call the random function (1.21) equipped with the said Gaussian probability measure ‘‘boundary-adapted Arithmetic Random Waves’’, much like the Arithmetic Random Waves (1.9). Alternatively (and equivalently), the boundary-adapted Arithmetic Random Waves is the ensemble of Gaussian centred random fields indexed by $x \in \mathcal{Q}$, with covariance functions

$$(1.22) \quad r_n(x, y) = \mathbb{E}[f_n(x) \cdot f_n(y)] = \frac{16}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \sin(\pi\mu_1 x_1) \sin(\pi\mu_2 x_2) \sin(\pi\mu_1 y_1) \sin(\pi\mu_2 y_2),$$

$n \in S$, $x = (x_1, x_2), y = (y_1, y_2) \in \mathcal{Q}$. The f_n are not stationary, even though, around generic points in \mathcal{Q} , far away from the boundary, f_n asymptotically tends to stationarity [8], understood in suitable regime, after suitable re-scaling. The main interest of this manuscript is the expected nodal length of f_n , and its comparison to (1.13), from this point on tacitly assuming $N_n \rightarrow \infty$.

Theorem 1.1. *Let*

$$\mathcal{L}_n = \mathcal{L}(f_n) = \text{len}(f_n^{-1}(0))$$

be the total nodal length of f_n on \mathcal{Q} , and recall the notation (1.15) and (1.20). There exists a subsequence of energy levels $S' \subseteq S$ satisfying the following properties:

- a. The sequence S' is of relative asymptotic density 1 within S .*
- b. The set of accumulation points of the sequence of numbers $\{\widehat{\nu}_n(4)\}_{n \in S'}$ is $[-1, 1]$.*
- c. Along $n \in S'$ we have $N_n \rightarrow \infty$, and*

$$(1.23) \quad \mathbb{E}[\mathcal{L}_n] = \frac{\sqrt{\lambda_n}}{2\sqrt{2}} \cdot \left(1 - \frac{1 + 4\widehat{\nu}_n(4)}{16} \cdot \frac{1}{N_n} + o_{N_n \rightarrow \infty} \left(\frac{1}{N_n} \right) \right).$$

The asymptotics (1.23) is expressed in terms of λ_n rather than in terms of n in a way that the leading term on the r.h.s. of (1.23) agrees with (1.13) explicitly, for there is a discrepancy factor of 2 otherwise, due to the discrepancy between (1.11) and (1.20). The boundary effect is then encapsulated within the second, *correction*, term

$$(1.24) \quad \mathcal{C}_n := -\frac{1 + 4\widehat{\nu}_n(4)}{16} \cdot \frac{1}{N_n}.$$

On one hand the asymptotics (1.23) shows that, since, outside a thin set of $n \in S$, we have the convergence (1.16) of ν_n to the uniform measure on \mathcal{S}^1 , for such a sequence of n the correction term is asymptotic to

$$\mathcal{C}_n \sim -\frac{1}{16N_n},$$

confirming Berry’s ansatz on the nodal deficiency. On the other hand, bearing in mind that, by property (b) of the sequence in Theorem 1.1, the accumulation set of $\{\widehat{\nu}_n(4)\}_{n \in S'}$ is the whole of $[-1, 1]$, without exclusions from S' ,

$$\mathcal{C}_n \cdot N_n$$

fluctuates infinitely in the asymmetric interval

$$\mathcal{C}_n \cdot N_n \in \left[-\frac{5}{16}, \frac{3}{16} \right].$$

The maximal nodal deficiency (resp. maximal nodal surplus) in (1.24) is uniquely attained by the Cilleruelo measure

$$\frac{1}{4}(\delta_{\pm 1} + \delta_{\pm i})$$

(resp. its tilt by $\pi/4$), consistent with our interpretation of the amplified horizontal and vertical wave propagation for Cilleruelo sequences (resp. their $\pi/4$ -tilt for the tilted Cilleruelo), in light of Berry's rationale of nodal deficiency occurring as a result of nodal lines perpendicular to the boundary. Finally, we notice that, judging by the analogous quantity for the Arithmetic Random Waves (1.18), and applying M. Berry's reasoning, explained in §1.2, we expect the fluctuations of \mathcal{L}_n to be of the same order of magnitude $\approx \frac{\sqrt{n}}{N_n}$ as \mathcal{C}_n (a by-product of the aforementioned "miraculous" cancellation). That means that, unlike the situation in [3], one cannot detect the nodal surplus or deficiency judging by the total nodal length based on one sample only. However, this could be mended by taking more samples, or, likely, by restricting the sample to the vicinity of the boundary.

The main conclusion (1.23) of Theorem 1.1 is valid for "generic" $n \in S' \subseteq S$ only, rather than for the whole sequence $n \in S$, though, importantly, this generic family is sufficiently rich so that to exhibit a variety of different asymptotic biases of the correction term (1.24). Below Theorem 1.4 will be stated, a version of Theorem 1.1 with an explicit control over the error term in (1.23), valid for the whole sequence $n \in S$ of energy level, expressed in terms of the so-called "spectral semi-correlations", defined in §1.5 (see Definition 1.2). Our failure to unrestrict the statement of Theorem 1.1 for the whole sequence $n \in S$ is then a by-product of Theorem 1.3 below asserting a bound for the semi-correlations for a generic sequence of energy levels. **We do believe that (1.23) holds for $n \in S$, with no further restriction.**

1.5. Spectral semi-correlations. Let

$$(1.25) \quad l = 2k$$

be an even number, with $k \geq 1$ an integer. The length- l spectral correlation set [13] is the set

$$(1.26) \quad \mathcal{R}_l(n) = \left\{ (\mu^1, \dots, \mu^l) \in \mathcal{E}_n^l : \sum_{j=1}^l \mu^j = 0 \right\}$$

of all l -tuples of lattice points in \mathcal{E}_n whose sum vanishes; by an elementary congruence obstruction modulo 2, for l odd the corresponding correlation sets are all empty. The size of $\mathcal{R}_l(n)$ is directly related to the l -th moment of the covariance function (1.12) corresponding to the Arithmetic Random Waves:

$$\int_{\mathbb{T}^2} p_n(x)^l dx = \frac{1}{N_n^l} |\mathcal{R}_l(n)|,$$

and bounding $|\mathcal{R}_6(n)|$ was a key ingredient for bounding the remainder while proving (1.17) in [13]. Since, for $k \geq 2$, fixing μ^1, \dots, μ^{l-2} so that

$$\sum_{j=1}^{l-2} \mu^j \neq 0,$$

the relation

$$\sum_{j=1}^l \mu^j = 0$$

determines the remaining two lattice points μ^{l-1} and μ^l up to permutation, it is readily seen that for every $l \geq 4$,

$$|\mathcal{R}_l(n)| = O(N_n^{l-2}).$$

Let

$$\mathcal{D}_l(n) = \{ \pi(\mu^1, -\mu^1, \dots, \mu^k, -\mu^k) : \pi \in S_l \}$$

be the set of all l -tuples cancelling out in pairs, where S_l is the symmetric group permuting the l -tuples, of size

$$|\mathcal{D}_l(n)| \sim c_k \cdot N_n^k.$$

Evidently, for every l and $n \in S$, we have the inclusion

$$\mathcal{D}_l(n) \subseteq \mathcal{R}_l(n).$$

Hence, in particular

$$(1.27) \quad |\mathcal{R}_l(n)| \gg N_n^k,$$

recalling (1.25). Bombieri and Bourgain [4] showed that

$$(1.28) \quad |\mathcal{R}_6(n)| \ll N_n^{7/2},$$

and established the striking inequality

$$|\mathcal{R}_6(n) \setminus \mathcal{D}_6(n)| \ll N_n^{l/2-\gamma},$$

for some $\gamma > 0$, valid for density-1 sequence of $n \in S$, or, alternatively, conditionally for the full sequence S , so that, in particular, for these n , the optimal inequality

$$(1.29) \quad |\mathcal{R}_l(n)| \ll N_n^k$$

holds (recall (1.25)).

In [5], the notion of spectral *quasi-correlations*, was instrumental for studying the analogue of \mathcal{E}_n for the Arithmetic Random Waves (1.9), restricted to domains decreasing with n above Planck scale, e.g discs with radius $n^{-1/2+\delta}$ (“Shrinking balls”). For l as above and $\epsilon > 0$, a length- l quasi correlation is an l -tuple (μ^1, \dots, μ^l) of points in \mathcal{E}_n so that

$$0 < \left\| \sum_{j=1}^l \mu^j \right\| < n^{1/2-\epsilon}.$$

It was shown [5, Theorem 1.4] that for n generic and l arbitrary even number, the quasi-correlation set is *empty*.

In this manuscript we introduce a new concept, of *semi-correlations*, instrumental within the proof of Theorem 1.1, as it will allow us to control the problematic *singular* set in \mathcal{Q} , see Corollary 2.6, and, we believe, of independent interest on its own right.

Definition 1.2 (Semi-correlations). For $l = 2k$, $n \in S$, the length- l semi-correlation set is the collection

$$(1.30) \quad \mathcal{M}_l(n) = \left\{ (\mu^1, \dots, \mu^l) \in \mathcal{E}_n^l : \sum_{j=1}^l \mu_1^j = 0 \right\}$$

of all l -tuples of lattice points in \mathcal{E}_n with the first coordinate summing up to 0.

From the above definition, it is evident that

$$\mathcal{D}_l(n) \subseteq \mathcal{R}_l(n) \subseteq \mathcal{M}_l(n),$$

so that, in particular,

$$|\mathcal{M}_l(n)| \gg N_n^k,$$

cf. (1.27). Quite remarkably, the following *optimal* upper bound holds for the semi-correlation set size, albeit for a generic sequence only.

Theorem 1.3 (Bound for the number of semi-correlations.). *For every $l = 2k \geq 4$ even integer, there exists a sequence $S' = S'(l) \subseteq S$ satisfying the following properties.*

- a. *The sequence $S' \subseteq S$ is of relative asymptotic density 1.*
- b. *The set of accumulation points of the sequence of numbers $\{\widehat{\nu}_n(4)\}_{n \in S'}$ is the whole of $[-1, 1]$.*

c. Along $n \in S'$ we have $N_n \rightarrow \infty$ and

$$(1.31) \quad |\mathcal{M}_l(n)| = O(N_n^k).$$

By using a standard diagonal argument it is possible to choose a density 1 sequence $S' \subseteq S$, satisfying (1.31) for *all* $l \geq 4$ even (with constant involved in the ‘O’-notation depending on l). Theorem 1.3 is stronger compared to the upper bound (1.29) for the spectral correlations, due to Bombieri-Bourgain, also valid for a density one sequence of $n \in S$. In addition to claiming the upper bound for the semi-correlations rather, significantly weaker than correlation, it also asserts the richness of the postulated sequence in terms of the angular distribution of \mathcal{E}_n , expressed in terms of the Fourier coefficients $\{\widehat{\nu}_n(4)\}_{n \in S'}$. The following result is a version of Theorem 1.1, with an explicit control over the error term in (1.23), expressed in terms of the spectral semi-correlations. After a significant amount of effort put into, **we still do not know whether (1.31) holds for all l even, along $n \in S$ with no further restriction**, and believe this question to be of sufficiently high interest, both for applications of the type of Theorem 1.1, and intrinsic, to be addressed in the future.

Theorem 1.4 (Explicit unrestricted version of Theorem 1.1). *For every $l \geq 4$ even we have*

$$(1.32) \quad \mathbb{E}[\mathcal{L}_n] = \frac{\sqrt{\lambda_n}}{2\sqrt{2}} \cdot \left(1 - \frac{1 + 4\widehat{\nu}_n(4)}{16} \cdot \frac{1}{N_n} + O_{N_n \rightarrow \infty} \left(\frac{1}{N_n^2} + N_n^{2-l} |\mathcal{M}_l(n)| \right) \right).$$

Theorem 1.4 implies Theorem 1.1 at once, by working with the sequence resulting from an application of Theorem 1.3 on $l = 8$ (say), and from this point on we will only care to prove Theorem 1.4 (and Theorem 1.3).

Acknowledgements. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013), ERC grant agreement n° 335141 (V.C and I.W.). We are grateful to Sven Gnutzmann for raising the question of nodal deficiency in arithmetic context as ours, and to Andrew Granville, Pär Kurlberg, Domenico Marinucci, Zeév Rudnick and Mikhail Sodin for the very inspiring and fruitful conversations concerning subjects relevant to this manuscript.

2. OUTLINE OF THE PAPER

2.1. Outline of the proof of Theorem 1.4.

2.1.1. *The Kac-Rice formula.* The Kac-Rice formula is a meta-theorem allowing one to evaluate the $(d-1)$ -volume of the zero set of a random field $F : \mathbb{R}^d \rightarrow \mathbb{R}$, for F satisfying some smoothness and non-degeneracy conditions. For $F : \mathbb{R}^d \rightarrow \mathbb{R}$, a sufficiently smooth centred Gaussian random field, we define

$$K_1(x) := \frac{1}{(2\pi)^{1/2} \sqrt{\text{Var}(F(x))}} \cdot \mathbb{E}[|\nabla F(x)| | F(x) = 0]$$

the zero density (first intensity) of F . Then the Kac-Rice formula asserts that for some suitable class of random fields F and $\overline{\mathcal{D}} \subseteq \mathbb{R}^d$ a compact closed subdomain of \mathbb{R}^d , one has the equality

$$(2.1) \quad \mathbb{E}[\text{Vol}_{d-1}(F^{-1}(0) \cap \overline{\mathcal{D}})] = \int_{\overline{\mathcal{D}}} K_1(x) dx.$$

We would like to apply (2.1) on the random fields f_n in (1.21) to evaluate the expectation on the l.h.s. of (1.23). Unfortunately, for some n , the aforementioned non-degeneracy conditions fail decisively for some points of \mathcal{Q} . For these cases, an *approximate* version of Kac-Rice was developed [24, Proposition 1.3] (in a slightly different context of evaluating the variance), so that rather than holding precisely, (2.1) would hold *approximately*, still yielding the asymptotic law for the evaluated expectation. Nevertheless, for this particular case, by excising some neighbourhoods of the problematic *degenerate* set, consisting of a union of a *grid* and finitely many *isolated* points, and by applying the Monotone Convergence Theorem, we will be able to deduce that (2.1) holds *precisely*, save for the

length of the said *deterministic grid* contained in the nodal set of f_n for some $n \in S$ of a particular form.

Let $n \in S$ be of the form (1.8), and denote the associated number

$$(2.2) \quad Q_n = 2^{\lfloor a/2 \rfloor} \prod_{k=1}^r q_k^{h_k},$$

so that, in particular, $Q_n^2 | n$. We will establish later (cf. Lemma 3.1 below) that in such a scenario, all of the μ_1 and μ_2 on the r.h.s. of (1.21) are divisible by Q_n , so that if $Q_n > 1$, then necessarily the (deterministic) grid

$$(2.3) \quad \mathcal{G}_n = \mathcal{G}(Q_n) = \bigcup_{k=1}^{Q_n-1} \left\{ (x_1, x_2) \in \mathcal{Q} : x_1 = \frac{k}{Q_n} \right\} \cup \bigcup_{k=1}^{Q_n-1} \left\{ (x_1, x_2) \in \mathcal{Q} : x_2 = \frac{k}{Q_n} \right\} \subseteq \mathcal{Q}$$

is a.s. contained inside the nodal line $f_n^{-1}(0)$ (see figures 1-2); \mathcal{G}_n is of length

$$(2.4) \quad \text{len}(\mathcal{G}_n) = 2 \cdot (Q_n - 1).$$

Lemma 3.1 will also assert that such a situation is only possible in this scenario, i.e. all the components $\{\mu_1\}_{\mu \in \mathcal{E}_n/\sim}$ are divisible by a maximal number $d > 1$, if and only if $d = Q_n$ in (2.2). The following proposition is the announced Kac-Rice formula, with the said caveat (namely, the length of \mathcal{G}_n , manifested on the r.h.s. of (2.6)).

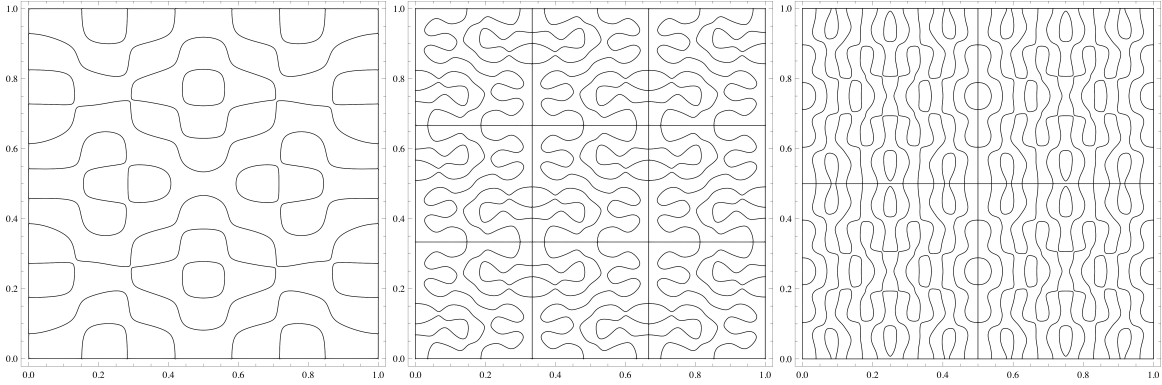


FIGURE 1. Left: Nodal line for some f_n , $n = 170$, not containing a grid ($Q_n = 1$). Middle: Same for some f_n , $n = 765$. Here $Q_n = 3$, and thereupon its nodal line contains the grid $\mathcal{G}(1/3)$. Right: Same for $n = 1000$, $Q_n = 2$.

Proposition 2.1. *Let f_n be as in (1.21), \mathcal{L}_n the nodal length of f_n , and*

$$(2.5) \quad K_1(x) = K_{1;n}(x) = \frac{1}{(2\pi)^{1/2} \sqrt{\text{Var}(f_n(x))}} \cdot \mathbb{E}[|\nabla f_n(x)| | f_n(x) = 0]$$

be the zero density of f_n . Then, for every $n \in S$, we have $K_1 \in L^1(\mathcal{Q})$, and moreover, we have

$$(2.6) \quad \mathbb{E}[\mathcal{L}_n] = \int_{\mathcal{Q}} K_1(x) dx + 2(Q_n - 1),$$

where Q_n is as in (2.2).

For example, by comparing (2.6) to (2.1), we may deduce as a particular by-product of Proposition 2.1, that (2.1) holds precisely in our case, if and only if $Q_n = 1$, i.e. the grid is empty. Below it will be demonstrated that Q_n on the r.h.s. of (2.6) does not contribute to the Kac-Rice integral (nor to the correction term $\sqrt{n}\mathcal{C}_n$ in (1.23), with \mathcal{C}_n given by (1.24)), e.g., it is easily dominated by $\frac{\sqrt{n}}{N^A}$, for every $A > 0$ (see the proof of Theorem 1.4 in §2.1.5 below).

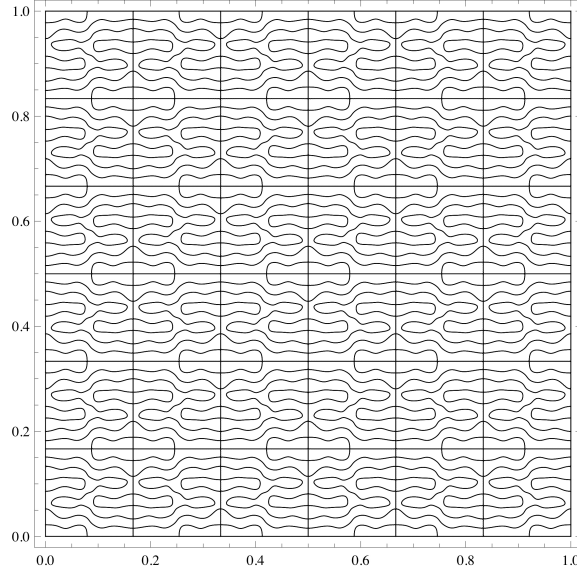


FIGURE 2. Nodal line for some f_n , $n = 3060$. Here $Q_n = 6$, and the nodal line of f_n contains $\mathcal{G}(1/6)$.

2.1.2. *The joint distribution of $(f_n(x), \nabla f_n(x))$.* By the definition (2.5) of the zero density function of f_n , to investigate $K_1(x)$ we naturally encounter the value distribution of both $f_n(x)$, determined by $\text{Var}(f_n(x))$, and $\nabla f_n(x)$ conditioned on $f_n(x) = 0$, determined by its 2×2 (conditional) covariance matrix. On recalling that the covariance function r_n of f_n is given by (1.22), so that on the diagonal

$$v_n(x) := r_n(x, x) = \text{Var}(f_n(x)),$$

with an elementary manipulation and well-known trigonometric identities yielding

$$(2.7) \quad v_n(x) = \text{Var}(f_n(x)) = 1 - s_n(x),$$

where

$$(2.8) \quad s_n(x) = \frac{4}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} (\cos(2\mu_1 \pi x_1) + \cos(2\mu_2 \pi x_2) - \cos(2\mu_1 \pi x_1) \cdot \cos(2\mu_2 \pi x_2)).$$

Further, we need to evaluate the covariance matrix of $\nabla f_n(x)$ conditioned on $f_n(x) = 0$. A scrupulous direct computation, carried out in Appendix A shows that the corresponding (normalised) covariance matrix is given by the following:

Lemma 2.2. *The 2×2 covariance matrix of $\nabla f_n(x)$, conditioned on $f_n(x) = 0$, and appropriately normalised, is the following 2×2 real symmetric matrix:*

$$(2.9) \quad \Omega_n(x) := \frac{2}{\pi^2 n} \cdot \mathbb{E} [\nabla f_n(x) \cdot \nabla^t f_n(x) | f_n(x) = 0] =: I_2 + \Gamma_n(x),$$

where

$$(2.10) \quad \Gamma_n(x) = \frac{8}{nN_n} \begin{pmatrix} b_{11;n}(x) & b_{12;n}(x) \\ b_{21;n}(x) & b_{22;n}(x) \end{pmatrix} - \frac{128}{nN_n^2 v_n(x)} \begin{pmatrix} d_{1;n}(x)^2 & d_{1;n}(x) \cdot d_{2;n}(x) \\ d_{1;n}(x) \cdot d_{2;n}(x) & d_{2;n}(x)^2 \end{pmatrix},$$

where v_n is given by (2.7) and (2.8);

$$(2.11) \quad b_{11;n}(x) = \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 (\cos(2\mu_1 \pi x_1) - \cos(2\mu_2 \pi x_2) - \cos(2\mu_1 \pi x_1) \cdot \cos(2\mu_2 \pi x_2)),$$

$$b_{12;n}(x) = b_{21;n}(x) = \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1 \mu_2 \sin(2\mu_1 \pi x_1) \cdot \sin(2\mu_2 \pi x_2),$$

$$b_{22;n}(x) = \sum_{\mu \in \mathcal{E}_n / \sim} \mu_2^2 (\cos(2\mu_2 \pi x_2) - \cos(2\mu_1 \pi x_1) - \cos(2\mu_1 \pi x_1) \cdot \cos(2\mu_2 \pi x_2)),$$

$$d_{1;n}(x) = \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1 \sin(2\mu_1 \pi x_1) \cdot \sin(\mu_2 \pi x_2)^2$$

and

$$(2.12) \quad d_{2;n}(x) = \sum_{\mu \in \mathcal{E}_n / \sim} \mu_2 \sin(2\mu_2 \pi x_2) \cdot \sin(\mu_1 \pi x_1)^2.$$

2.1.3. *Singular set.* Next, we aim at analysing the asymptotic behaviour of the r.h.s. of (2.6). Towards this goal we will separate the domain \mathcal{Q} of the integration on the r.h.s. of (2.6) into the “good” or nonsingular set, where K_1 is “tame”, i.e. admits precise asymptotic (Proposition 2.7), and the “bad” or singular sets, which itself consists of small singular squares so that to be able to control the integral of K_1 . First, we will bound the total contribution of the singular set (Corollary 2.6) from above, by separately bounding the number of small singular squares (Proposition 2.4), appealing to the bound for spectral semi-correlations in Theorem 1.3, and the contribution of a singular small square (Proposition 2.5).

Below it will be asserted that for *most* of the points $x \in \mathcal{Q}$, both the value of $v_n(x)$ is close to unit (equivalently, $s_n(x)$ is *small*), and Ω_n in (2.9) is close to the unit matrix (equivalently, Γ_n is small); we will designate the *other* points as “singular”, and excise them while performing an asymptotic analysis on K_1 . To quantify it, we take $\epsilon_0 > 0$ and $c_0 > 0$, and keep them fixed but sufficiently small throughout. We will endow the singular set with a structure of a union of small squares (cf. [23, 24]).

Definition 2.3 (Singular set). Let $\epsilon_0 > 0$ and $c_0 > 0$ be two parameters. Take

$$\delta_0 = \delta_0(n) = \frac{c_0}{\sqrt{n}},$$

and $K := \left\lfloor \frac{1}{\delta_0} \right\rfloor + 1$. For $1 \leq i \leq K$ define the interval

$$I_i = \left[(i-1) \cdot \frac{1}{K}, i \cdot \frac{1}{K} \right] \subseteq [0, 1],$$

and for $1 \leq i, j \leq K$ denote the small square

$$(2.13) \quad Q_{ij} := I_i \times I_j \subseteq \mathcal{Q}.$$

We have the partition

$$\mathcal{Q} = \bigcup_{1 \leq i, j \leq K} Q_{ij}$$

of the square into a union of small squares, disjoint save for boundary overlaps.

a. Recall the notation in (2.8) and (2.9). A small square Q_{ij} in (2.13) is “singular” if it contains a point $x_0 \in Q_{ij}$ satisfying either of the three inequalities:

$$|s_n(x_0)| > \epsilon_0,$$

or

$$|\operatorname{tr}(\Gamma_n(x_0))| > \epsilon_0,$$

or

$$|\det(\Gamma_n(x_0))| > \epsilon_0.$$

b. The singular set is the union

$$\mathcal{Q}_s := \bigcup_{Q_{ij} \text{ singular}} Q_{ij}.$$

of all small singular squares.

c. The complement $\mathcal{Q} \setminus \mathcal{Q}_s$ of the singular set is called “nonsingular set”.

The following couple of propositions assert a bound for the total measure of the singular set, and for the contribution of a single singular square respectively. Combining these two will yield an upper bound for the total contribution of the singular set \mathcal{Q}_s to the integral on the r.h.s. of (2.6).

Proposition 2.4 (Bound for the measure of the singular set). *For every $l \geq 4$ even integer we have the following bound for the measure of the singular set in terms of the length- l spectral correlation set (1.30):*

$$(2.14) \quad \text{meas}(\mathcal{Q}_s) = O\left(N_n^{-l} |\mathcal{M}_l(n)|\right),$$

where the constant involved in the ‘ O ’-notation depends only on l .

Proposition 2.5 (Bound for a single small square). *Let $Q \subseteq \mathcal{Q}$ be an arbitrary square of side length $\frac{c_0}{\sqrt{n}}$ with $c_0 > 0$ sufficiently small. Then*

$$(2.15) \quad \int_Q K_1(x) dx = O\left(\frac{N_n^2}{\sqrt{n}}\right),$$

with the constant involved in the ‘ O ’-notation in (2.15) absolute.

It is worthy of a mention that the doubling exponent method due to Donnelly-Fefferman [11], with relation to Yau’s conjecture (1.2), yields the deterministic bound of $O(1)$ for the nodal length of f_n restricted to Q as in Proposition 2.5, that, being better than the global bound of \sqrt{n} , falls short from being sufficient for our needs², by a significant margin. We believe that the optimal upper bound on the r.h.s. of (2.15) should be $O\left(\frac{1}{\sqrt{n}}\right)$, however, after some effort, we were not able to prove that. Instead, we sacrifice a power of N_n by virtually not exploiting the summation in (1.22) (except the invariance of \mathcal{E}_n w.r.t. $\mu = (\mu_1, \mu_2) \mapsto (\mu_2, \mu_1)$), in the hope to gain the lost power of N_n while bounding the number of singular squares (equivalently, the measure of \mathcal{Q}_s), which is precisely what is achieved in Proposition 2.4, with the help of Theorem 1.3. In particular, Proposition 2.5 applies to all singular squares $Q_{ij} \subseteq \mathcal{Q}_s$, leading to the following, possibly sub-optimal, result.

Corollary 2.6. *For every $l \geq 4$ even integer we have the following bound for the contribution of the singular set to the integral on the r.h.s. of (2.6):*

$$(2.16) \quad \int_{\mathcal{Q}_s} K_1(x) dx = O\left(N_n^{2-l} \sqrt{n} \cdot |\mathcal{M}_l(n)|\right).$$

The upshot of Corollary 2.6 is that, thanks to Theorem 1.3, by choosing l sufficiently big, we can make the r.h.s. of (2.16) smaller than $\sqrt{n} \cdot N_n^{-A}$, with $A > 0$ arbitrarily large. That is crucial if we are to majorise it by the second term in the claimed asymptotic expansion (1.23), of order of magnitude $\approx \frac{\sqrt{n}}{N_n}$. The proof of Corollary 2.6 is immediate given Propositions 2.4 and 2.5, and, thereupon, conveniently omitted.

2.1.4. *Perturbative analysis on the non-singular set.* Outside the singular set, the precise analysis for the density function is feasible.

Proposition 2.7 (Asymptotics for K_1 outside \mathcal{Q}_s). *Let $\epsilon_0 > 0$ be a sufficiently small number, and recall that $s_n(\cdot)$ and $\Gamma_n(\cdot)$ are given by (2.8) and (2.10) respectively. Then K_1 admits the following asymptotics, uniformly for $x \in \mathcal{Q} \setminus \mathcal{Q}_s$,*

$$K_1(x) = \frac{\sqrt{\lambda_n}}{2\sqrt{2}} + L_n(x) + \Upsilon_n(x),$$

where the leading term is given by

$$(2.17) \quad L_n(x) = \frac{\sqrt{n}\pi}{4\sqrt{2}} \left(s_n(x) + \frac{\text{tr } \Gamma_n(x)}{2} + \frac{3}{4} s_n(x)^2 + \frac{1}{4} s_n(x) \cdot \text{tr } \Gamma_n(x) - \frac{\text{tr}(\Gamma_n(x)^2)}{16} - \frac{(\text{tr}(\Gamma_n(x)))^2}{32} \right),$$

²Any bound of the form $O\left(\frac{N_n^A}{\sqrt{n}}\right)$, $A > 0$, would be sufficient for our purposes, by choosing l sufficiently high (see how Theorem 1.1 is inferred from Theorem 1.4, as explained immediately after the latter theorem).

and the error term is bounded by

$$(2.18) \quad |\Upsilon_n(x)| = O\left(\sqrt{n} \cdot (|s_n(x)|^3 + |\Gamma_n(x)|^3)\right),$$

with constant involved in the ‘ O ’-notation absolute.

We observe that, by the definition (2.10) with (2.11)-(2.12), the diagonal entries of $\Gamma_n(x)$ are bounded by an absolute constant (using the nonnegativity of the $d_{:,n}(x)^2$), and therefore, so are the diagonal entries of $\Omega_n(x)$ in (2.9), and, further, *all* the entries of Ω_n are bounded by an absolute constant, by the Cauchy-Schwarz inequality³. Taking also into account (2.8), and the definition (2.17) of $L(x) = L_n(x)$, we conclude that $L(x)$ is uniformly bounded

$$(2.19) \quad |L(x)| = O(\sqrt{n}),$$

with the involved constant absolute; this will prove useful later, while restricting (or, rather, unrestricting) the range of the integration in (2.6) to $\mathcal{Q} \setminus \mathcal{Q}_s$.

2.1.5. *Proof of Theorem 1.4.* The following lemma will be proved in §7 below.

Lemma 2.8. *a. Let $L(x) = L_n(x)$ be as in (2.17), and $l \geq 4$ an even number. Then*

$$\int_{\mathcal{Q}} L(x) dx = -\frac{\pi(1 + 4\hat{\nu}_n)}{32\sqrt{2}} \cdot \frac{\sqrt{n}}{N_n} + O\left(\frac{\sqrt{n}}{N_n^2}\right) + O\left(n^{1/2} N_n^{-l-1} |\mathcal{M}_l(n)|\right).$$

b. Let $\Upsilon(x) = \Upsilon_n(x)$ be a function defined on \mathcal{Q} , satisfying (2.18). Then

$$\int_{\mathcal{Q}} |\Upsilon(x)| dx = O\left(\frac{\sqrt{n}}{N_n^2}\right).$$

Given the above results, the proof of Theorem 1.4 is rather straightforward.

Proof of Theorem 1.4 assuming Corollary 2.6, Proposition 2.7 and Lemma 2.8. Let l be given. We invoke Proposition 2.1, and separate the contribution of the nonsingular and the singular sets in the integral on the r.h.s. of (2.6) to write, with the help of Corollary 2.6,

$$(2.20) \quad \mathbb{E}[\mathcal{L}_n] = \int_{\mathcal{Q} \setminus \mathcal{Q}_s} K_1(x) dx + \int_{\mathcal{Q}_s} K_1(x) dx + 2(Q_n - 1) = \int_{\mathcal{Q} \setminus \mathcal{Q}_s} K_1(x) dx + O(Q_n + N_n^{2-l} \sqrt{n} \cdot |\mathcal{M}_l(n)|),$$

with Q_n given by (2.2). First, we claim that for every $A > 0$,

$$(2.21) \quad Q_n = O\left(\frac{\sqrt{n}}{N_n^A}\right),$$

so that the contribution of the grid length is majorised by the error term on the r.h.s. of (1.32). To this end, we recall the prime decomposition (1.8) of n , and write

$$P_n = \prod_{j=1}^s p_j^{e_j},$$

so that

$$n = \begin{cases} P_n Q_n^2 \\ 2P_n Q_n^2 \end{cases}.$$

Since, as it is well-known,

$$N_n = N_{P_n} = 4 \prod_{j=1}^s (e_j + 1),$$

and $N_n = O(n^\epsilon)$ for every $\epsilon > 0$, we may easily write (taking $A := 1/(2\epsilon)$)

$$\sqrt{n} \geq Q_n \cdot \sqrt{P_n} \gg Q_n \cdot N_{P_n}^A = Q_n \cdot N_n^A,$$

³This argument recovers the Gaussian Correlation Inequality in this particular case, see e.g. [25].

which is (2.21).

Next, we use the asymptotics of K_1 on the nonsingular set claimed in Proposition 2.7 to write

$$(2.22) \quad \begin{aligned} \int_{\mathcal{Q} \setminus \mathcal{Q}_s} K_1(x) dx &= \frac{\sqrt{\lambda_n}}{2\sqrt{2}} \cdot \text{meas}(\mathcal{Q} \setminus \mathcal{Q}_s) + \int_{\mathcal{Q} \setminus \mathcal{Q}_s} L_n(x) dx + \int_{\mathcal{Q} \setminus \mathcal{Q}_s} \Upsilon_n(x) dx \\ &= \frac{\sqrt{\lambda_n}}{2\sqrt{2}} + \int_{\mathcal{Q}} L_n(x) dx + O \left(\int_{\mathcal{Q} \setminus \mathcal{Q}_s} |\Upsilon_n(x)| dx + \sqrt{n} \cdot N_n^{-l} |\mathcal{M}_l(n)| \right) \end{aligned}$$

with a bound (2.18) for the error term, thanks to the uniform bound (2.19) on $L_n(x)$ together with Proposition 2.4.

Upon substituting (2.22) into (2.20), and exploiting (2.21), we may deduce (with the error term of $\sqrt{n} \cdot N_n^{-l} |\mathcal{M}_l(n)|$ being majorized by $N_n^{2-l} \sqrt{n} \cdot |\mathcal{M}_l(n)|$):

$$(2.23) \quad \mathbb{E}[\mathcal{L}_n] = \frac{\sqrt{\lambda_n}}{2\sqrt{2}} + \int_{\mathcal{Q}} L_n(x) dx + O \left(\int_{\mathcal{Q} \setminus \mathcal{Q}_s} |\Upsilon_n(x)| dx + N_n^{2-l} \sqrt{n} \cdot |\mathcal{M}_l(n)| \right).$$

The statement (1.32) of Theorem 1.4 now follows upon employing Lemma 2.8a for evaluating the integral $\int_{\mathcal{Q}} L_n(x) dx$ on the r.h.s. of (2.23), and Lemma 2.8b for bounding the error term $\int_{\mathcal{Q} \setminus \mathcal{Q}_s} |\Upsilon_n(x)| dx$. \square

2.2. Outline of the proof of Theorem 1.3. To prove Theorem 1.3, our first step as in [4], is to restrict ourselves to the set of integers $n \in S$ with “typical” factorization type. This is accomplished (for square-free n) with the help of Lemma 8.1. Fixing “typical” $n = p_1 p_2 \dots p_r$ with $p_1 < p_2 < \dots < p_r$, the key observation then is that any non-trivial relation of the form

$$(\mu_1, \dots, \mu_l) \in \mathcal{E}_n^l : \sum_{j=1}^l \mu_1^j = 0,$$

can be rewritten as a non-degenerate *quasi-linear* equation with respect to the Gaussian primes $p_j = \pi_j \bar{\pi}_j$,

$$(2.24) \quad \sum_{s=1}^l \text{Re} \left(i^{\alpha_s} \prod_{j=1}^r \pi_{j,s}^* \right) = 0,$$

where each $\pi_{j,s}^* = \{\pi_j, \bar{\pi}_j\}$ with rotation factor $\alpha_s \in \mathbb{Z}$. Thus, having primes $p_1, p_2 \dots p_{r-1}$ fixed, the equation (2.24) determines $\arg \pi_r$, and therefore prime p_r in a unique fashion (see the proof of Proposition 8.2 for the details). After conditioning on this value of p_r and taking into account Lemma 8.1, we deduce that equality (2.24) can occur only for small proportion of numbers $n \in S$. This is accomplished in Proposition 8.2 for those $n \in S$ which are free of small prime factors and in Proposition 8.3 for general $n \in S$.

In order to show that, the set of accumulation points of the sequence of numbers $\{\widehat{U}_n(4)\}_{n \in S'}$ is the whole of $[-1, 1]$, we choose $n \in S$ of the form $n = p_n^m p$, where p_n and p are appropriately chosen primes and $m \in S$ using classical result due to Kubilius (Lemma 8.4). This is the content of Proposition 8.5.

Outline of the paper. The rest of the paper is organised as follows. The proof of a version of the Kac-Rice formula in Proposition 2.1 will be given in §3, and the asymptotics for the Kac-Rice integral on the r.h.s. of (2.6) will be analysed throughout §4-§7, as follows. An upper bound of Proposition 2.4 for the singular set \mathcal{Q}_s as in Definition (2.3) in terms of the semi-correlations set will be established in §4, whereas a contribution of a single small square $Q_{ij} \subseteq \mathcal{Q}_s$ of Proposition 2.5 will be controlled in §5.

The perturbative analysis of Proposition 2.7 for the zero density on the nonsingular set will be carried out in §6, whose contribution to the expected nodal length of f_n will be evaluated in §7. A proof of Theorem 1.3, bounding the semi-correlation set, will be given in §8, whereas some more technically demanding computations, required as part of proofs for the said results, will be performed in the appendix.

3. PROOF OF PROPOSITION 2.1: KAC-RICE FORMULA FOR EXPECTED NODAL LENGTH

In view of [1, Theorem 6.3] (see also [1, Proposition 1.2]), the equality (2.6) holds provided that the Gaussian distribution of $f_n(x)$ is non-degenerate for every $x \in \mathcal{Q}$. It is easy to construct examples of numbers n , so that this non-degeneracy condition fails decisively for some points in \mathcal{Q} . Let

$$(3.1) \quad \mathcal{H}_n = \{x \in \mathcal{Q} : v_n(x) = 0\} = \{x \in \mathcal{Q} \setminus \partial\mathcal{Q} : \forall \mu \in \mathcal{E}_n / \sim, \mu_1 x_1 \in \mathbb{Z} \vee \mu_2 x_2 \in \mathbb{Z}\} \subseteq \mathcal{Q}$$

be the *degenerate* set.

Lemma 3.1. *Let n be of the form (1.8), recall that Q_n is given by (2.2), and the grid \mathcal{G}_n as in (2.3). Then we have the decomposition*

$$(3.2) \quad \mathcal{H}_n = \mathcal{G}_n \cup \mathcal{A}_n,$$

where $\mathcal{A}_n \subseteq \mathcal{Q}$ is a finite set of isolated points in \mathcal{Q} .

Proof. First we aim at proving the announced decomposition (3.2). Let $x = (x_1, x_2) \in \mathcal{H}_n$, whence, by the definition (3.1) of \mathcal{H}_n , for all $\mu \in \mathcal{E}_n / \sim$ either $\mu_1 x_1 \in \mathbb{Z}$ or $\mu_2 x_2 \in \mathbb{Z}$ holds, and recall that we assumed that $\mu_1, \mu_2 \neq 0$ for all $\mu \in \mathcal{E}_n / \sim$ (as otherwise the corresponding summand in (1.21) necessarily vanishes). Assume that for some $\mu \in \mathcal{E}_n$ we have $\mu_1 x_1 \in \mathbb{Z}$, and denote $l := \mu_1 x_1 \in \mathbb{Z}$. Then, since $x_1 \in [0, 1]$ and $\mu_1^2 + \mu_2^2 = n$, necessarily $|l| \leq \sqrt{n}$, and $x_1 = \frac{l}{\mu_1}$. Hence, if for some $\mu = (\mu_1, \mu_2) \in \mathcal{E}_n / \sim$ we have $\mu_1 x_1 \in \mathbb{Z}$, and for some $\tilde{\mu} = (\tilde{\mu}_1, \tilde{\mu}_2) \in \mathcal{E}_n / \sim$ we have $\tilde{\mu}_2 x_2 \in \mathbb{Z}$, then both coordinates (x_1, x_2) belong to the finite set

$$\mathcal{A}'_n := \left\{ \frac{l}{\mu_j} : 0 < |l| \leq \sqrt{n}, \mu \in \mathcal{E}_n / \sim \right\},$$

so that, by prescribing $\mathcal{A}_n \subseteq \mathcal{A}'_n$, the \mathcal{A}_n in the decomposition (3.2), is finite, provided that we prove that the rest of \mathcal{H}_n is indeed the grid \mathcal{G}_n .

By the above, we may assume that $x \in \mathcal{H}_n$ satisfies

$$(3.3) \quad \mu_1 x_1 \in \mathbb{Z} \text{ for all } \mu \in \mathcal{E}_n / \sim,$$

and claim that in this case necessarily x_1 is of the form

$$(3.4) \quad x_1 = \frac{k}{Q_n}$$

for some $1 \leq k \leq Q_n - 1$, taking care of the symmetric case ($\mu_2 x_2 \in \mathbb{Z}$ for all $\mu \in \mathcal{E}_n / \sim$) along identical lines. Once having (3.4) established, that would yield that $x \in \mathcal{G}_n$ on the grid (see (2.3)), and we would only have the burden of proving the converse inclusion $\mathcal{G}_n \subseteq \mathcal{H}_n$ (which is easy).

To the end of proving (3.4), we let

$$(3.5) \quad Q'_n := \gcd\{\mu_1 : \mu \in \mathcal{E}_n / \sim\} = \gcd\{\mu_2 : \mu \in \mathcal{E}_n / \sim\}$$

be the greatest common divisor of the abscissas of all the lattice points in \mathcal{E}_n . Then, since the set of integers d so that $d \cdot x_1 \in \mathbb{Z}$ is an ideal in \mathbb{Z} , we have $Q'_n x_1 \in \mathbb{Z}$ (equivalently, since we can express Q'_n as a linear combination of $\{\mu_1 : \mu \in \mathcal{E}_n / \sim\}$). The above shows that (3.3) is equivalent to the single condition $Q'_n x_1 \in \mathbb{Z}$. That is, $x_1 = \frac{k}{Q'_n}$, and, since $x_1 \in (0, 1)$, we also get $1 \leq k \leq Q'_n - 1$, yielding (3.4) (that, as mentioned above, in turn implies $x \in \mathcal{G}_n$), once we prove that $Q_n = Q'_n$, to be shown next.

To this end we recall the prime decomposition (1.8) of n , and work in the ring of Gaussian integers $\mathbb{Z}[i]$ (which is a unique factorization domain, or, simply, UFD), where we think of $\mathcal{E}_n \subseteq \mathbb{R}^2$

as embedded into \mathbb{C} , via the map $\mu = (\mu_1, \mu_2) \mapsto \mu_1 + i\mu_2$. For every prime p_j in the decomposition (1.8) we associate a prime element $\pi_j \in \mathbb{Z}[i]$ with norm $\|\pi_j\|^2 = p_j$, and take

$$a_0 = a - 2\lfloor a/2 \rfloor = \begin{cases} 0 & a \text{ even} \\ 1 & a \text{ odd} \end{cases}.$$

With this notation, and (2.2), one may express every $\mu \in \mathcal{E}_n / \sim$ (i.e., as an element of \mathbb{C} , up to a sign or complex conjugation) as

$$(3.6) \quad \mu = Q_n \cdot (1 + i)^{a_0} \prod_{j=1}^s \pi_j^{k_j} \overline{\pi_j}^{e_j - k_j},$$

for some $0 \leq k_j \leq e_j$, $j = 1, \dots, s$. This implies that $Q_n | Q'_n$ at once. To see that also $Q'_n | Q_n$, we further exploit the UFD property of $\mathbb{Z}[i]$, implying, in particular, that the gcd in $\mathbb{Z}[i]$ is well-defined.

By the definition (3.5) of Q'_n , we have that $Q'_n | \mu_1, \mu_2$ for all $\mu \in \mathcal{E}_n / \sim$, and so $Q'_n | \mu$ in $\mathbb{Z}[i]$, valid for all $\mu \in \mathcal{E}_n$, i.e.

$$Q'_n | Q''_n := \gcd\{\mu : \mu \in \mathcal{E}_n / \sim\} \in \mathbb{Z}[i].$$

However, by making the two choices $k_j := e_j$, and $k_j := 0$, having only $Q_n \cdot (1 + i)^{a_0}$ as a common factor in (3.6), it shows that $Q''_n = Q_n \cdot (1 + i)^{a_0}$, and recalling that either $a_0 = 0$ or $a_0 = 1$, the readily established $Q_n | Q'_n$, and $Q'_n | Q''_n$ (so that Q'_n could be either Q_n or $Q_n \cdot (1 + i)$, the latter not being an integer number), this readily implies $Q'_n = Q_n$, that, as it was mentioned above, yields that $x \in \mathcal{G}_n$. To finish the statement of Lemma 3.1 it is sufficient to observe that if x_1 is of the form (3.4), then, in light of (3.5), and $Q_n = Q'_n$ above, (3.3) is satisfied, so that the inclusion $\mathcal{G}_n \subseteq \mathcal{H}_n$ holds. \square

With the above preparatory result we are now in a position to conclude the proof of Proposition 2.1.

Proof of Proposition 2.1. Around each points in $x_i \in \mathcal{H}_n$ we excise a small ball $\mathcal{B}(x_i, \varepsilon)$, and denote

$$\mathcal{Q}_\varepsilon = \mathcal{Q} \setminus \bigcup_{x_i \in \mathcal{H}_n} \mathcal{B}(x_i, \varepsilon),$$

with the intention to apply the Kac-Rice method to evaluate the expected nodal length of the restriction $f_n|_{\mathcal{Q}_\varepsilon}$ of f_n to the remaining set. That is, we excised the radius- ε balls centred at each of the finitely many points \mathcal{A}_n , and, possibly, finitely many rectangles of the form $(x_1 - \varepsilon, x_1 + \varepsilon) \times (0, 1)$ and $(0, 1) \times (x_2 - \varepsilon, x_2 + \varepsilon)$, centred at the horizontal and vertical bars of the grid \mathcal{G}_n , in case it is non-empty. Since outside \mathcal{H}_n , the random field f_n satisfies the non-degeneracy hypothesis of [1, Theorem 6.3], the Kac-Rice formula (2.1) holds for f_n restricted to \mathcal{Q}_ε , asserting that the restricted expected nodal length is given by

$$\mathbb{E}[\mathcal{L}(f_n|_{\mathcal{Q}_\varepsilon})] = \int_{\mathcal{Q}_\varepsilon} K_1(x) dx,$$

with K_1 as in (2.5).

On one hand, since, on recalling (2.4), the restricted nodal length $\{\mathcal{L}(f_n|_{\mathcal{Q}_\varepsilon})\}_{\varepsilon > 0}$ is an increasing sequence of nonnegative random variables with the a.s. limit

$$\lim_{\varepsilon \rightarrow 0} \mathcal{L}(f_n|_{\mathcal{Q}_\varepsilon}) = \mathcal{L}_n - \text{len}(\mathcal{G}_n) = \mathcal{L}(f_n) - 2(Q_n - 1),$$

the Monotone Convergence Theorem applied as $\varepsilon \rightarrow 0$, yields

$$(3.7) \quad \lim_{\varepsilon \rightarrow 0} \mathbb{E}[\mathcal{L}(f_n|_{\mathcal{Q}_\varepsilon})] = \mathbb{E}[\mathcal{L}_n] - 2(Q_n - 1).$$

On the other hand, by the definition,

$$(3.8) \quad \lim_{\varepsilon \rightarrow 0} \int_{\mathcal{Q}_\varepsilon} K_1(x) dx = \int_{\mathcal{Q}} K_1(x) dx,$$

with the r.h.s. of (3.8) finite on infinite. The equality of the limits in (3.7) and (3.8) show that the main statement (2.6) of Proposition 2.1 holds, whether both the l.h.s. and r.h.s. of (2.6) are finite or infinite. That $K_1 \in L^1(\mathcal{Q})$ also follows, since the l.h.s. of (2.6) is finite by the deterministic bound (1.2) (alternatively, from the asymptotic analysis within Theorem 1.4 of the integral on the r.h.s. of (2.6), with no circular logic). \square

4. PROOF OF PROPOSITION 2.4: CONTROLLING THE MEASURE OF THE SINGULAR SET

4.1. Proof of Proposition 2.4. We will need the following auxiliary lemmas, whose proofs will be given in §4.2 below.

Lemma 4.1. *Let $Q_{ij} \subset \mathcal{Q}_s$ be a singular small square. Then necessarily one of the followings holds:*

a. For every $y \in Q_{ij}$,

$$|s_n(y)| > \epsilon_0/2.$$

b. For every $y \in Q_{ij}$,

$$|\operatorname{tr}(\Gamma_n(y))| > \epsilon_0/2.$$

c. For every $y \in Q_{ij}$,

$$|\det(\Gamma_n(y))| > \epsilon_0/2.$$

Lemma 4.1 allows for the following decomposition of \mathcal{Q}_s .

Definition 4.2 (Singular decomposition). a. The set $\mathcal{Q}_{s,1} \subset \mathcal{Q}_s$ is the union of all small squares $Q_{ij} \subset \mathcal{Q}_s$ so that for all $x \in Q_{ij}$ the inequality

$$(4.1) \quad |s_n(x)| > \epsilon_0/2$$

is satisfied.

b. The set $\mathcal{Q}_{s,2}$ is the union of all small squares $Q_{ij} \subseteq \mathcal{Q}_s \setminus \mathcal{Q}_{s,1}$, so that either for all $x \in Q_{ij}$ the inequality

$$|\operatorname{tr}(\Gamma_n(x))| > \epsilon_0/2$$

holds, or for all $x \in Q_{ij}$, the inequality

$$|\det(\Gamma_n(x))| > \epsilon_0/2$$

holds.

c. By Lemma 4.1,

$$(4.2) \quad \mathcal{Q}_s = \mathcal{Q}_{s,1} \cup \mathcal{Q}_{s,2},$$

(“singular decomposition”), with $\mathcal{Q}_{s,1}$ and $\mathcal{Q}_{s,2}$ disjoint save for boundary overlaps.

The respective measures of $\mathcal{Q}_{s,1}$ and $\mathcal{Q}_{s,2}$ will be bounded in the following lemma. Recall that $\mathcal{M}_l(n)$ is the length- l spectral semi-correlation set (1.30).

Lemma 4.3. *For every $l \geq 4$ even integer we have the following bounds for the measures of the singular sets $\mathcal{Q}_{s,1}$, $\mathcal{Q}_{s,2}$:*

$$(4.3) \quad \operatorname{meas}(\mathcal{Q}_{s,1}) = O(N_n^{-l} \cdot |\mathcal{M}_l(n)|),$$

$$(4.4) \quad \operatorname{meas}(\mathcal{Q}_{s,2}) = O(N_n^{-l} \cdot |\mathcal{M}_l(n)|),$$

with the constant involved in the ‘ O ’-notation depending only on l (also ϵ_0 and c_0).

Proof of Proposition 2.4 assuming lemmas 4.1-4.3. In light of the singular decomposition (4.2), the statement (2.14) of Proposition 2.4 follows at once from Lemma 4.3. \square

4.2. Proofs of the auxiliary lemmas 4.1-4.3.

Proof of Lemma 4.1. We assume that for some i, j , there exists $x_0 \in Q_{ij}$ with

$$(4.5) \quad |s_n(x_0)| > \epsilon_0,$$

and claim that for all $x \in Q_{ij}$, the inequality

$$(4.6) \quad |s_n(x)| > \epsilon_0/2$$

holds (assuming c_0 is sufficiently small), i.e. scenario (a) of Lemma 4.1 prevails. On recalling the definition (2.8) of s_n , and differentiating (2.8) explicitly, it is easy to obtain the uniform bound

$$\|\nabla s_n(x)\| \leq c_1 \cdot \sqrt{n},$$

with some *absolute* constant $c_1 > 0$. This readily implies that $s_n(\cdot/\sqrt{n})$ is a Lipschitz function with associated constant absolute, i.e.

$$(4.7) \quad |s_n(x) - s_n(y)| \leq c_1 \sqrt{n} \|x - y\|.$$

Hence, if $x \in Q_{ij}$, we have that $\|x - x_0\| \leq \sqrt{2}c_0 \cdot \sqrt{n}$, and together with (4.5) and (4.7) this implies (4.6), so long as we choose $c_0 > 0$ sufficiently small, depending on ϵ_0 (and c_1).

Essentially the same argument works for the other two scenarios (b) and (c) of Lemma 4.1, on recalling the definition (2.10) of $\Gamma_n(\cdot)$, and exploiting the Lipschitz property of both $\text{tr}(\Gamma_n(\cdot))$ and $\det(\Gamma_n(\cdot))$, in place of $s_n(\cdot)$. These are easy to establish to be with Lipschitz constant of order of magnitude at most \sqrt{n} , by differentiating the individual entries of $\Gamma_n(\cdot)$. \square

Proof of Lemma 4.3. We first aim to prove the first statement (4.3) of Lemma 4.3, the proof of the second statement (4.4) being quite similar, as explained in the last paragraph of this proof. By the defining inequality (4.1) of $\mathcal{Q}_{s,1}$, holding for all $x \in \mathcal{Q}_{s,1}$, we have that

$$\mathcal{Q}_{s,1} \subseteq \{x \in \mathcal{Q} : |s_n(x)| > \epsilon_0/2\},$$

so that we may apply the Chebyshev inequality to yield for every $l \geq 4$ even integer the bound

$$\text{meas}(\mathcal{Q}_{s,1}) \leq \text{meas} \left(\left\{ x \in \mathcal{Q}_{s,1} : |s_n(x)|^l > \frac{\epsilon_0^l}{2^l} \right\} \right) \leq \frac{2^l}{\epsilon_0^l} \int_{\mathcal{Q}_{s,1}} |s_n(x)|^l dx.$$

By the definition (2.8) of $s_n(x)$, we have $s_n(x) = A_n(x) + B_n(x) + C_n(x)$, where

$$\begin{aligned} A_n(x) &= \sum_{\mu \in \mathcal{E}_n/\sim} \cos(2\mu_1 \pi x_1), \\ B_n(x) &= \sum_{\mu \in \mathcal{E}_n/\sim} \cos(2\mu_2 \pi x_2), \\ C_n(x) &= - \sum_{\mu \in \mathcal{E}_n/\sim} \cos(2\mu_1 \pi x_1) \cos(2\mu_2 \pi x_2). \end{aligned}$$

Since l is even, we may bound

$$(4.8) \quad \begin{aligned} \int_{\mathcal{Q}_{s,1}} s_n(x)^l dx &\leq \frac{4^l}{N_n^l} \int_{\mathcal{Q}} [A_n(x) + B_n(x) + C_n(x)]^l dx \\ &\leq \frac{4^l \cdot 3^l}{N_n^l} \int_{\mathcal{Q}} [A_n(x)^l + B_n(x)^l + C_n(x)^l] dx. \end{aligned}$$

Recalling the definition of the semi-correlation set $\mathcal{M}_l(n)$ in (1.30), we observe that, for $i = 1, 2$, we easily evaluate:

$$(4.9) \quad \int_{\mathcal{Q}} A_n(x)^l dx = \int_0^1 A_n(x)^l dx_1 = \frac{2\pi}{2^l} |\mathcal{M}_l(n)|,$$

and the same

$$(4.10) \quad \int_{\mathcal{Q}} B_n(x)^l dx = \frac{2\pi}{2^l} |\mathcal{M}_l(n)|,$$

whereas for the other integral in (4.8), we recall the correlation set (1.26) to bound

$$(4.11) \quad \int_{\mathcal{Q}} C_n(x)^l dx \ll |\mathcal{R}_l(n)| \leq |\mathcal{M}_l(n)|,$$

as, obviously, $\mathcal{R}_l(n) \subseteq \mathcal{M}_l(n)$.

The first statement (4.3) of Lemma 4.3 follows directly from (4.8), (4.9), (4.10) and (4.11), via Chebyshev's inequality. Finally, the same argument as above also yields the second statement (4.4) of Lemma 4.3, upon observing that for every $y \in \mathcal{Q}_{s,2}$ we have $|s_n(y)| \leq \epsilon_0$ with ϵ_0 small; so on $\mathcal{Q}_{s,2}$ we can Taylor expand the function

$$\frac{1}{v_n(x)} = \frac{1}{1 - s_n(x)} = 1 + O(s_n(x))$$

that appear in the entries of the conditional covariance matrix $\Gamma_n(x)$. □

5. PROOF OF PROPOSITION 2.5: CONTROLLING THE CONTRIBUTION OF A SMALL SQUARE

5.1. Proof of Proposition 2.5. We will first state the following lemma, whose proof will be given in §5.2 immediately below.

Lemma 5.1. *Let $Q \subseteq \mathcal{Q}$ be an arbitrary square of side length $\frac{c_0}{\sqrt{n}}$ with $c_0 > 0$ sufficiently small. We have the following uniform bound, holding for all $\eta, \mu \in \mathcal{E}_n$ with $\eta_1 \neq 0$ and $\mu_1 \neq 0$:*

$$(5.1) \quad \int_Q \frac{|\sin(\eta_2 \pi x_2) \sin(\mu_2 \pi x_2)| \cdot |\mu_1 \sin(\eta_1 \pi x_1) \cos(\mu_1 \pi x_1) - \eta_1 \sin(\mu_1 \pi x_1) \cos(\eta_1 \pi x_1)|}{\{\sin^4(\eta_1 \pi x_1) \sin^4(\eta_2 \pi x_2) + \sin^4(\mu_1 \pi x_1) \sin^4(\mu_2 \pi x_2)\}^{1/2}} dx_1 dx_2 \\ = O\left(\frac{1}{\sqrt{n}}\right),$$

with constant involved in the 'O'-notation depending only on the constant $c_0 > 0$.

Proof of Proposition 2.5 assuming Lemma 5.1. Let X and Y be the conditional random variables

$$X = \partial_{x_1} f_n(x) | f_n(x) = 0, \quad Y = \partial_{x_2} f_n(x) | f_n(x) = 0,$$

and recall that the joint distribution of (X, Y) is centred Gaussian, with covariance matrix equal to $\Omega_n(x)$ in (2.9) (and (2.10)-(2.12)), up to the normalising constant read from (2.9). Then, with the aid of the Cauchy-Schwarz inequality, we obtain the inequality

$$\mathbb{E}[|\nabla f_n(x)| | f_n(x) = 0] \leq \mathbb{E}[|X|] + \mathbb{E}[|Y|] \leq \sqrt{(\mathbb{E}[|X|])^2} + \sqrt{(\mathbb{E}[|Y|])^2} \\ \leq \sqrt{\mathbb{E}[X^2]} + \sqrt{\mathbb{E}[Y^2]},$$

so that we may bound the zero density (2.5) as

$$K_1(x) \leq \tilde{K}_{1,1}(x) + \tilde{K}_{1,2}(x),$$

where

$$\tilde{K}_{1,1}(x) = \frac{1}{\sqrt{2\pi} \sqrt{v_n(x)}} \cdot \sqrt{\mathbb{E}[X^2]},$$

and

$$\tilde{K}_{1,2}(x) = \frac{1}{\sqrt{2\pi} \sqrt{v_n(x)}} \cdot \sqrt{\mathbb{E}[Y^2]}.$$

In what follows we are going to prove that

$$(5.2) \quad \int_Q \tilde{K}_{1,1}(x) dx = O\left(\frac{N_n^2}{\sqrt{n}}\right),$$

and the same proof (with all coordinates switched) yields the other estimate

$$(5.3) \quad \int_Q \tilde{K}_{1,2}(x) dx = O\left(\frac{N_n^2}{\sqrt{n}}\right);$$

together these imply (2.15). By (2.9), and on recalling that

$$\frac{n\pi^2}{2} + \frac{4\pi^2}{N_n} b_{11;n}(x) = \frac{16\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1^2 \sin^2(\mu_2 \pi x_2) \cos^2(\mu_1 \pi x_1),$$

we express the variance of the conditional derivative as

$$\mathbb{E}[X^2] = \frac{16\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1^2 \sin^2(\mu_2 \pi x_2) \cos^2(\mu_1 \pi x_1) - \frac{64\pi^2}{N_n^2 v_n(x)} d_{1;n}(x)^2$$

since

$$v_n(x) = \frac{16}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \sin^2(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2),$$

it follows that

$$\begin{aligned} \frac{\mathbb{E}[X^2]}{v_n(x)} &= \frac{16\pi^2}{N_n^2 v_n^2(x)} \left[v_n(x) N_n \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1^2 \sin^2(\mu_2 \pi x_2) \cos^2(\mu_1 \pi x_1) - 4d_{1;n}(x)^2 \right] \\ &= \frac{16\pi^2}{N_n^2 v_n^2(x)} \left[v_n(x) N_n \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1^2 \sin^2(\mu_2 \pi x_2) \cos^2(\mu_1 \pi x_1) \right. \\ &\quad \left. - 4^2 \left(\sum_{\mu \in \mathcal{E}_n/\sim} \mu_1 \sin(\mu_1 \pi x_1) \cos(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \right)^2 \right] \\ &= \frac{16^2 \pi^2}{N_n^2 v_n^2(x)} \left[v_n(x) \frac{N_n}{16} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1^2 \sin^2(\mu_2 \pi x_2) \cos^2(\mu_1 \pi x_1) \right. \\ &\quad \left. - \left(\sum_{\mu \in \mathcal{E}_n/\sim} \mu_1 \sin(\mu_1 \pi x_1) \cos(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \right)^2 \right] \end{aligned}$$

where

$$\begin{aligned} &v_n(x) \frac{N_n}{16} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1^2 \sin^2(\mu_2 \pi x_2) \cos^2(\mu_1 \pi x_1) - \left(\sum_{\mu \in \mathcal{E}_n/\sim} \mu_1 \sin(\mu_1 \pi x_1) \cos(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \right)^2 \\ &= \sum_{\eta \in \mathcal{E}_n/\sim} \sin^2(\eta_1 \pi x_1) \sin^2(\eta_2 \pi x_2) \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1^2 \sin^2(\mu_2 \pi x_2) \cos^2(\mu_1 \pi x_1) \\ &\quad - \sum_{\eta \in \mathcal{E}_n/\sim} \eta_1 \sin(\eta_1 \pi x_1) \cos(\eta_1 \pi x_1) \sin^2(\eta_2 \pi x_2) \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1 \sin(\mu_1 \pi x_1) \cos(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2). \end{aligned}$$

It follows that we have the following explicit expression:

$$(5.4) \quad \frac{\mathbb{E}[X^2]}{v_n(x)} = \frac{\pi^2}{\sum_{\eta, \mu \in \mathcal{E}_n/\sim} \sin^2(\eta_1 \pi x_1) \sin^2(\eta_2 \pi x_2) \sin^2(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2)} \\ \times \sum_{\eta, \mu \in \mathcal{E}_n/\sim} \sin(\eta_1 \pi x_1) \sin^2(\eta_2 \pi x_2) \cos(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \\ \times [\mu_1^2 \sin(\eta_1 \pi x_1) \cos(\mu_1 \pi x_1) - \eta_1 \mu_1 \sin(\mu_1 \pi x_1) \cos(\eta_1 \pi x_1)].$$

Next, by grouping (η, μ) together with (μ, η) on the r.h.s. of (5.4), this is equivalent to

$$\frac{\mathbb{E}[X^2]}{v_n(x)} = \frac{1}{2} \frac{\pi^2}{\sum_{\eta, \mu \in \mathcal{E}_n/\sim} \sin^2(\eta_1 \pi x_1) \sin^2(\eta_2 \pi x_2) \sin^2(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2)} \\ \times \sum_{\eta, \mu \in \mathcal{E}_n/\sim} \sin^2(\eta_2 \pi x_2) \sin^2(\mu_2 \pi x_2) \\ \times [\mu_1 \sin(\eta_1 \pi x_1) \cos(\mu_1 \pi x_1) - \eta_1 \sin(\mu_1 \pi x_1) \cos(\eta_1 \pi x_1)]^2.$$

Hence

$$(5.5) \quad \tilde{K}_{1,1}(x) \ll \frac{\left\{ \sum_{\eta, \mu \in \mathcal{E}_n/\sim} \sin^2(\eta_2 \pi x_2) \sin^2(\mu_2 \pi x_2) \cdot [\mu_1 \sin(\eta_1 \pi x_1) \cos(\mu_1 \pi x_1) - \eta_1 \sin(\mu_1 \pi x_1) \cos(\eta_1 \pi x_1)]^2 \right\}^{1/2}}{\left\{ \sum_{\eta, \mu \in \mathcal{E}_n/\sim} \sin^2(\eta_1 \pi x_1) \sin^2(\eta_2 \pi x_2) \sin^2(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \right\}^{1/2}} \\ \leq \sum_{\eta, \mu \in \mathcal{E}_n/\sim} \frac{|\sin(\eta_2 \pi x_2) \sin(\mu_2 \pi x_2)| \cdot |\mu_1 \sin(\eta_1 \pi x_1) \cos(\mu_1 \pi x_1) - \eta_1 \sin(\mu_1 \pi x_1) \cos(\eta_1 \pi x_1)|}{\left\{ \sum_{\eta, \mu \in \mathcal{E}_n/\sim} \sin^2(\eta_1 \pi x_1) \sin^2(\eta_2 \pi x_2) \sin^2(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \right\}^{1/2}} \\ \leq \sum_{\eta, \mu \in \mathcal{E}_n/\sim} \frac{|\sin(\eta_2 \pi x_2) \sin(\mu_2 \pi x_2)| \cdot |\mu_1 \sin(\eta_1 \pi x_1) \cos(\mu_1 \pi x_1) - \eta_1 \sin(\mu_1 \pi x_1) \cos(\eta_1 \pi x_1)|}{\left\{ \sin^2(\eta_1 \pi x_1) \sin^2(\eta_2 \pi x_2) \sin^2(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \right\}^{1/2}},$$

where we used the positivity of all the summands in the denominator, as well as the easy inequality

$$\sum_{k=1}^K a_k^2 \leq \left(\sum_{k=1}^K |a_k| \right)^2,$$

valid for any sequence of real numbers $\{a_k\}_{k=1}^K$. The bound (5.2) (and similarly (5.3)), implying, as it was mentioned above, the statement of Proposition 2.5, finally follows upon integrating the individual summands on the r.h.s. of (5.5), and using Lemma 5.1 to bound the contribution of each one of them (recall that we are allowed to assume that $\eta_1, \mu_1 \neq 0$).

□

5.2. Proof of Lemma 5.1.

Proof. Let $\eta, \mu \in \mathcal{E}_n/\sim$, and $Q \subseteq \mathcal{Q}$ be an arbitrary square of side length $\frac{c_0}{\sqrt{n}}$ with $c_0 > 0$ sufficiently small. Now we write

$$(5.6) \quad \eta_1 \pi x_1 = k_\eta \pi + h_\eta,$$

where $x_1 \in [0, 1]$, $k_\eta = \left[\frac{\eta_1 x_1}{\pi} \right] \in \mathbb{Z}$ is the integer value of $\eta_1 x_1 / \pi$, independent of x_1 by the above, and $h_\eta = h_\eta(x_1) \in (-\pi/2, \pi/2]$. We also denote

$$(5.7) \quad x_1^\eta = \frac{k_\eta \pi}{\eta_1} = \pi x_1 - \frac{h_\eta}{\eta_1};$$

the numbers k_μ , x_1^μ and $h_\mu = h_\mu(x_1)$ are defined analogously, with η_1 replaced by μ_1 . Note that, h_η and h_μ , both being linear functions of x_1 , satisfy the relation

$$(5.8) \quad \eta_1 h_\mu - \mu_1 h_\eta = \pi (\mu_1 k_\eta - \eta_1 k_\mu) = \eta_1 \mu_1 (x_1^\eta - x_1^\mu),$$

that will be exploited below. Finally, we denote

$$(5.9) \quad d(\eta, \mu) = d_{ij}(\eta, \mu) = |x_1^\eta - x_1^\mu|,$$

so that (5.8) reads

$$(5.10) \quad |\eta_1 h_\mu - \mu_1 h_\eta| = \eta_1 \mu_1 \cdot d(\eta, \mu)$$

crucially depending on η, μ and Q only (but independent of x_1).

First, we assume, that $x_1^\eta \neq x_1^\mu$, i.e.

$$(5.11) \quad d(\eta, \mu) > 0,$$

and by switching between η and μ if necessary, we may assume w.l.o.g., that

$$(5.12) \quad d(\eta, \mu) = x_1^\mu - x_1^\eta > 0.$$

We expand the numerator of the integrand on l.h.s. of (5.1) (first, without the absolute value):

$$\begin{aligned} & \mu_1 \sin(\eta_1 \pi x_1) \cos(\mu_1 \pi x_1) - \eta_1 \sin(\mu_1 \pi x_1) \cos(\eta_1 \pi x_1) \\ &= (-1)^{k_\eta + k_\mu} \cdot [\mu_1 \sin(h_\eta) \cos(h_\mu) - \eta_1 \sin(h_\mu) \cos(h_\eta)], \end{aligned}$$

so that

$$(5.13) \quad \begin{aligned} & |\mu_1 \sin(\eta_1 \pi x_1) \cos(\mu_1 \pi x_1) - \eta_1 \sin(\mu_1 \pi x_1) \cos(\eta_1 \pi x_1)| \\ &= |\mu_1 h_\eta - \eta_1 h_\mu| + E_{\eta, \mu}(x_1) = \eta_1 \mu_1 d(\eta, \mu) + E_{\eta, \mu}(x_1) \end{aligned}$$

by (5.10), where $d(\eta, \mu)$ is as in (5.9), and the error term is bounded by

$$(5.14) \quad |E_{\eta, \mu}(x_1)| = O(\mu_1 h_\eta^3 + \mu_1 h_\eta h_\mu^2 + \eta_1 h_\mu h_\eta^2 + \eta_1 h_\mu^3).$$

For the denominator of the integrand on l.h.s. of (5.1), we have

$$(5.15) \quad \begin{aligned} & \{\sin^4(\eta_1 \pi x_1) \sin^4(\eta_2 \pi x_2) + \sin^4(\mu_1 \pi x_1) \sin^4(\mu_2 \pi x_2)\}^{1/2} \\ & \gg \sin^2(\eta_1 \pi x_1) \sin^2(\eta_2 \pi x_2) + \sin^2(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \\ & \gg h_\eta^2 \sin(\eta_2 x_2)^2 + h_\mu^2 \sin(\mu_2 x_2)^2, \end{aligned}$$

and also

$$\begin{aligned} & \{\sin^4(\eta_1 \pi x_1) \sin^4(\eta_2 \pi x_2) + \sin^4(\mu_1 \pi x_1) \sin^4(\mu_2 \pi x_2)\}^{1/2} \\ & \gg |\sin(\eta_1 \pi x_1) \sin(\eta_2 \pi x_2) \sin(\mu_1 \pi x_1) \sin(\mu_2 \pi x_2)| \gg |h_\eta h_\mu \sin(\eta_2 \pi x_2) \sin(\mu_2 \pi x_2)|, \end{aligned}$$

and we plan to use the former inequality whenever we are going to bound the contribution of the main term of (5.13), and both inequalities for the error term in (5.13).

We denote $s_\eta = s_\eta(x) = \sin(\eta_2 \pi x_2)$ and $s_\mu = s_\mu(x) = \sin(\mu_2 \pi x_2)$. With the newly introduced notation, the denominator (5.15) is

$$(5.16) \quad \{\sin^4(\eta_1 \pi x_1) \sin^4(\eta_2 \pi x_2) + \sin^4(\mu_1 \pi x_1) \sin^4(\mu_2 \pi x_2)\}^{1/2} \gg s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2;$$

as we will keep x_2 fixed, and aim at first to integrate w.r.t. x_1 , we will treat s_η and s_μ as parameters. The estimates (5.13) and (5.16) imply that the integral in (5.1) is bounded by

$$(5.17) \quad \int_Q \frac{|\sin(\eta_2 \pi x_2) \sin(\mu_2 \pi x_2)| |\mu_1 \sin(\eta_1 \pi x_1) \cos(\mu_1 \pi x_1) - \eta_1 \sin(\mu_1 \pi x_1) \cos(\eta_1 \pi x_1)|}{\{\sin^4(\eta_1 \pi x_1) \sin^4(\eta_2 \pi x_2) + \sin^4(\mu_1 \pi x_1) \sin^4(\mu_2 \pi x_2)\}^{1/2}} dx_1 dx_2 \\ \ll \int_Q |s_\eta s_\mu| \frac{\eta_1 \mu_1 d(\eta, \mu) + E_{\eta, \mu}(x_1)}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2,$$

and, in what follows, we are going to bound the contribution of the main and the error terms by separate arguments.

First, we bound the contribution of the main term in the integrand on the r.h.s. of (5.17). To this end we notice that h_η and h_μ are both linear functions of x_1 , so we are going to exploit their inter-dependence (5.6) (and its analogue for h_μ) to write:

$$(5.18) \quad s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2 = s_\eta^2 \eta_1^2 (\pi x_1 - x_1^\eta)^2 + s_\mu^2 \mu_1^2 (\pi x_1 - x_1^\mu)^2 \\ = s_\eta^2 \eta_1^2 \Delta x_1^2 + s_\mu^2 \mu_1^2 (d(\eta, \mu) - \Delta x_1)^2,$$

where we denoted

$$(5.19) \quad \Delta x_1 := \pi x_1 - x_1^\eta,$$

a linear transformation of the variable x_1 , and used (5.12). We complete the expression on the r.h.s. of (5.18) to a square:

$$(5.20) \quad s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2 = s_\eta^2 \eta_1^2 \Delta x_1^2 + s_\mu^2 \mu_1^2 (d(\eta, \mu) - \Delta x_1)^2 \\ = \frac{1}{s_\eta^2 \eta_1^2 + s_\mu^2 \mu_1^2} \left\{ [(s_\eta^2 \eta_1^2 + s_\mu^2 \mu_1^2) \Delta x_1 - s_\mu^2 \mu_1^2 d(\eta, \mu)]^2 + s_\eta^2 s_\mu^2 \eta_1^2 \mu_1^2 d^2(\eta, \mu) \right\};$$

note that we may assume that $s_\mu \neq 0$ and $s_\eta \neq 0$ (holding outside a discrete set of x_2). Substituting the identity (5.20) into (5.18), we may bound the contribution of the main term of the integral on the r.h.s. of (5.17) as

$$(5.21) \quad \int_Q |s_\eta s_\mu| \frac{\eta_1 \mu_1 d(\eta, \mu)}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 = \eta_1 \mu_1 d(\eta, \mu) \int_{I_j} |s_\eta s_\mu| (s_\eta^2 \eta_1^2 + s_\mu^2 \mu_1^2) dx_2 \times \\ \times \int_{I_i} \frac{1}{[(s_\eta^2 \eta_1^2 + s_\mu^2 \mu_1^2) \Delta x_1 - s_\mu^2 \mu_1^2 d(\eta, \mu)]^2 + s_\eta^2 s_\mu^2 \eta_1^2 \mu_1^2 d^2(\eta, \mu)} dx_1 \\ = \frac{1}{\eta_1 \mu_1 d(\eta, \mu)} \int_{I_j} \frac{s_\eta^2 \eta_1^2 + s_\mu^2 \mu_1^2}{|s_\eta s_\mu|} dx_2 \int_{\tilde{I}_i} \frac{d\Delta x_1}{\left[\frac{s_\eta^2 \eta_1^2 + s_\mu^2 \mu_1^2}{\eta_1 \mu_1 s_\eta s_\mu d(\eta, \mu)} \Delta x_1 - \frac{s_\mu \mu_1}{s_\eta \eta_1} \right]^2 + 1},$$

where we have transformed the coordinates (5.19), and \tilde{I}_i is some shift of the interval I_i .

Another transformation of coordinates w.r.t. Δx_1 shows that, denoting \tilde{I}_i the new range of integration, (5.21) is equal to

$$(5.22) \quad \int_Q |s_\eta s_\mu| \frac{\eta_1 \mu_1 d(\eta, \mu)}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 = \int_{I_j} dx_2 \int_{\tilde{I}_i} \frac{dt}{1+t^2} \ll \frac{1}{\sqrt{n}},$$

since the integral w.r.t. t is bounded by an absolute constant.

Next, we turn to evaluating the contribution of the error term

$$\int_Q |s_\eta s_\mu| \cdot \frac{E_{\eta, \mu}(x_1)}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2$$

to the integral (5.17). By (5.14), we have

$$\begin{aligned}
(5.23) \quad & \int_Q |s_\eta s_\mu| \frac{|E_{\eta,\mu}(x_1)|}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 \\
& \ll \int_Q |s_\eta s_\mu| \frac{|\mu_1 h_\eta^3|}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 + \int_Q |s_\eta s_\mu| \frac{|\mu_1 h_\eta h_\mu^2|}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 \\
& + \int_Q |s_\eta s_\mu| \frac{|\eta_1 h_\mu h_\eta^2|}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 + \int_Q |s_\eta s_\mu| \frac{|\eta_1 h_\mu^3|}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2;
\end{aligned}$$

we will bound the 1st integral on the r.h.s. of (5.23), with the last one being bounded along similar lines, and the 2nd and the 3rd ones are easier, as the corresponding numerator is divisible by both h_η and h_μ (see the argument below).

We might bound the 1st integral on the r.h.s. of (5.23) (or the integral w.r.t. x_1), using the Cauchy-Schwarz inequality

$$(5.24) \quad s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2 \gg s_\eta h_\eta s_\mu h_\mu,$$

to bound the denominator from below, so that

$$\int_Q |s_\eta s_\mu| \frac{|\mu_1 h_\eta^3|}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 \leq \int_Q \frac{|s_\eta s_\mu| \cdot |\mu_1 h_\eta^3|}{s_\eta h_\eta s_\mu h_\mu} dx_1 \ll \int_Q \frac{|\mu_1 h_\eta^2|}{|h_\mu|} dx_1$$

which has the unfortunate burden of having to deal with h_μ in the denominator, that could potentially be much smaller than h_η . To deal with this obstacle we exploit the relation (5.10) between h_η and h_μ once again, both being linear functions of x_1 . We write

$$|\mu_1 h_\eta^3| = |\mu_1 h_\eta \cdot h_\eta^2| = |\mu_1 h_\eta \pm \eta_1 \mu_1 d(\eta, \mu)| \cdot h_\eta^2 \leq \eta_1 h_\eta^2 |h_\mu| + \eta_1 \mu_1 d(\eta, \mu) h_\eta^2,$$

so that

$$(5.25) \quad \int_{I_i} |s_\eta s_\mu| \frac{|\mu_1 h_\eta^3|}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 \leq \int_{I_i} \frac{|s_\eta s_\mu| |\eta_1| |h_\mu| h_\eta^2}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 + \int_{I_i} \frac{|s_\eta s_\mu \eta_1 \mu_1| d(\eta, \mu) h_\eta^2}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1.$$

For the former integral on the r.h.s. of (5.25) we use the aforementioned idea (5.24) to bound the denominator from below

$$\int_{I_j} \frac{|s_\eta s_\mu| |\eta_1| |h_\mu| h_\eta^2}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 \ll |\eta_1| \int_{I_j} h_\eta dx_1 \ll \sqrt{n} \cdot \frac{1}{\sqrt{n}} = 1,$$

since $|h_\eta| \leq \frac{\pi}{2}$ is bounded, so that, after the integration w.r.t. x_2 , we obtain

$$(5.26) \quad \int_Q \frac{|s_\eta s_\mu| |\eta_1| |h_\mu| h_\eta^2}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 \ll \frac{1}{\sqrt{n}}.$$

Concerning the latter integral on the r.h.s. of (5.25) (or the double integral on Q), since, as above, h_η is bounded, we have

$$(5.27) \quad \int_Q \frac{|s_\eta s_\mu \eta_1 \mu_1| d(\eta, \mu) h_\eta^2}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 \ll \int_Q \frac{|s_\eta s_\mu \eta_1 \mu_1| d(\eta, \mu)}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 \ll \frac{1}{\sqrt{n}},$$

readily evaluated above as the main term (see (5.22)).

Combining the estimates (5.27) and (5.26), and substituting them into (5.25) (integrated w.r.t. x_2 on I_j) yield the bound

$$(5.28) \quad \int_Q |s_\eta s_\mu| \frac{|\mu_1 h_\eta^3|}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 \ll \frac{1}{\sqrt{n}}$$

for the 1st integral on the r.h.s. of (5.23), and its analogues for the other integrals on the r.h.s. of (5.23) also follow (the 4th is symmetric, whereas the 2nd and the 3rd are easier, with no need to deal with the denominator). Substituting (5.28) and its analogues for the other three integrals in (5.23) into (5.23), we finally obtain a bound for the contribution of the error term

$$(5.29) \quad \int_Q |s_\eta s_\mu| \frac{E_{\eta,\mu}(x_1)}{s_\eta^2 h_\eta^2 + s_\mu^2 h_\mu^2} dx_1 dx_2 \ll \frac{1}{\sqrt{n}}.$$

This, together with (5.22), and (5.17), implies the statement (5.1) of Lemma 5.1 in this non-degenerate case (5.11).

Finally, we treat the degenerate case $x_1^\eta = x_1^\mu$, or, equivalently,

$$(5.30) \quad \eta_1 h_\mu = \mu_1 h_\eta.$$

In this case the situation becomes easier to analyse, and the main terms of the expansion (5.13) vanishes via (5.30), so that here (5.13) reads

$$(5.31) \quad |\mu_1 \sin(\eta_1 x_1) \cos(\mu_1 x_1) - \eta_1 \sin(\mu_1 x_1) \cos(\eta_1 x_1)| = E_{\eta,\mu}(x_1),$$

with the same bound (5.14) for the error term. The argument leading to (5.26) above works unimpaired, with no need to bound the extra term (5.27) here, as the precise identity (5.30) reduces bounding the integral (5.25) (after integrating w.r.t. x_2 on I_j) to bounding (5.26) with no remainder term (5.27). This shows that in this degenerate case, (5.29) holds, and, by (5.31), it also shows that the statement (5.1) of Lemma 5.1 holds here. □

6. PROOF OF PROPOSITION 2.7: PERTURBATIVE ANALYSIS ON THE NON-SINGULAR SET

Proof. The proof of Proposition 2.7 rests on a precise Taylor analysis for the density function K_1 . We exploit the fact that the Gaussian expectation (2.5) is an analytic function with respect to the parameters of the corresponding covariance matrix outside its singularities. It is then possible to Taylor expand K_1 explicitly, in the domain $\mathcal{Q} \setminus \mathcal{Q}_s$, where both $s_n(x)$ and the all the entries of $\Gamma_n(x)$ are small.

We first expand the factor

$$(6.1) \quad \frac{1}{\sqrt{\text{Var}(f_n(x))}} = \frac{1}{\sqrt{1 - s_n(x)}} = 1 + \frac{1}{2}s_n(x) + \frac{3}{8}s_n^2(x) + O(s_n^3(x)).$$

that appear in (2.5). Next, we consider the Gaussian integral

$$(6.2) \quad \mathcal{I}(\Gamma_n(x)) = \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z (I_2 + \Gamma_n(x))^{-1} z^t \right\} dz.$$

Observing that

$$(I_2 + \Gamma_n(x))^{-1} = I_2 - \Gamma_n(x) + \Gamma_n^2(x) + O(\Gamma_n^3(x)),$$

we expand the exponential in (6.2) as:

$$\begin{aligned} \exp \left\{ -\frac{1}{2} z (I_2 + \Gamma_n(x))^{-1} z^t \right\} &= \exp \left\{ -\frac{z z^t}{2} \right\} \cdot \exp \left\{ \frac{1}{2} z (\Gamma_n(x) - \Gamma_n^2(x) + O(\Gamma_n^3(x))) z^t \right\} \\ &= \exp \left\{ -\frac{z z^t}{2} \right\} \left[1 + \frac{1}{2} z (\Gamma_n(x) - \Gamma_n^2(x) + O(\Gamma_n^3(x))) z^t \right. \\ &\quad \left. + \frac{1}{2 \cdot 4} (z (\Gamma_n(x) - \Gamma_n^2(x) + O(\Gamma_n^3(x))) z^t)^2 + O(z (\Gamma_n(x) - \Gamma_n^2(x) + O(\Gamma_n^3(x))) z^t)^3 \right], \end{aligned}$$

so that the Gaussian integral (6.2) is such that

$$\begin{aligned} \mathcal{I}(\Gamma_n(x)) &= \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{zz^t}{2} \right\} \left[1 + \frac{1}{2} z\Gamma_n(x)z^t - \frac{1}{2} z\Gamma_n^2(x)z^t + \frac{1}{8} (z\Gamma_n(x)z^t)^2 \right] dz \\ &\quad + O(\Gamma_n^3(x)). \end{aligned}$$

We introduce the following notation:

$$(6.3) \quad \begin{aligned} \mathcal{I}_0(\Gamma_n(x)) &= \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{zz^t}{2} \right\}, \\ \mathcal{I}_1(\Gamma_n(x)) &= \frac{1}{2} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{zz^t}{2} \right\} z\Gamma_n(x)z^t dz, \\ \mathcal{I}_2(\Gamma_n(x)) &= -\frac{1}{2} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{zz^t}{2} \right\} z\Gamma_n^2(x)z^t dz, \end{aligned}$$

and

$$(6.4) \quad \mathcal{I}_3(\Gamma_n(x)) = \frac{1}{8} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{zz^t}{2} \right\} (z\Gamma_n(x)z^t)^2 dz.$$

In Lemma 6.1, postponed at the end of this section, we evaluate the terms $\mathcal{I}_i(\Gamma_n(x))$; we use Lemma 6.1 to write

$$(6.5) \quad \begin{aligned} \mathcal{I}(\Gamma_n(x)) &= \mathcal{I}_0(\Gamma_n(x)) + \mathcal{I}_1(\Gamma_n(x)) + \mathcal{I}_2(\Gamma_n(x)) + \mathcal{I}_3(\Gamma_n(x)) + O(\Gamma_n^3(x)) \\ &= \mathcal{I}_0(\Gamma_n(x)) + \mathcal{I}_1(\Gamma_n(x)) + \beta_1 \operatorname{tr}(\Gamma_n^2(x)) + \beta_2 [\operatorname{tr}(\Gamma_n(x))]^2 + O(\Gamma_n^3(x)) \end{aligned}$$

where

$$\beta_1 = -\frac{3}{2^{3/2}}\pi^{3/2} + \frac{1}{8} \frac{15\sqrt{2}}{4}\pi^{3/2} = -\frac{9}{16\sqrt{2}}\pi^{3/2}, \quad \beta_2 = \frac{15\sqrt{2}}{64}\pi^{3/2}.$$

Finally we expand the factor

$$\frac{1}{\sqrt{\det(I_2 + \Gamma_n(x))}}.$$

We know that $\Gamma_n(x)$ is symmetric and hence diagonalizable, we denote with $g_{1;n}(x)$ and $g_{2;n}(x)$ its eigenvalues. The eigenvalues of $I_2 + \Gamma_n(x)$ are then $1 + g_{1;n}(x)$ and $1 + g_{2;n}(x)$, and we have

$$\det(I_2 + \Gamma_n(x)) = [1 + g_{1;n}(x)][1 + g_{2;n}(x)] = 1 + \operatorname{tr}(\Gamma_n(x)) + \det(\Gamma_n(x)).$$

This implies that we can write

$$\frac{1}{\sqrt{\det(I_2 + \Gamma_n(x))}} = 1 - \frac{1}{2} [\operatorname{tr}(\Gamma_n(x)) + \det(\Gamma_n(x))] + \frac{3}{8} [\operatorname{tr}(\Gamma_n(x)) + \det(\Gamma_n(x))]^2 + O(\Gamma_n^3(x)).$$

We rewrite the third term on the right hand side of the last equation as follows

$$\begin{aligned} &[\operatorname{tr}(\Gamma_n(x)) + \det(\Gamma_n(x))]^2 \\ &= [g_{1;n}(x) + g_{2;n}(x) + g_{1;n}(x)g_{2;n}(x)]^2 \\ &= g_{1;n}^2(x) + g_{2;n}^2(x) + g_{1;n}^2(x)g_{2;n}^2(x) + 2g_{1;n}^2(x)g_{2;n}(x) + 2g_{1;n}(x)g_{2;n}^2(x) + 2g_{1;n}(x)g_{2;n}(x) \\ &= g_{1;n}^2(x) + g_{2;n}^2(x) + 2g_{1;n}(x)g_{2;n}(x) + O(\Gamma_n^3(x)) \\ &= [\operatorname{tr}(\Gamma_n(x))]^2 + O(\Gamma_n^3(x)) \end{aligned}$$

and we have that

$$\frac{1}{\sqrt{\det(I_2 + \Gamma_n(x))}} = 1 - \frac{1}{2} \operatorname{tr}(\Gamma_n(x)) - \frac{1}{2} \det(\Gamma_n(x)) + \frac{3}{8} [\operatorname{tr}(\Gamma_n(x))]^2 + O(\Gamma_n^3(x)).$$

Observing that $g_{1;n}^2(x)$ and $g_{2;n}^2(x)$ are eigenvalues of $\Gamma_n^2(x)$, we rewrite $\det(\Gamma_n(x))$ as follows

$$\begin{aligned}\det(\Gamma_n(x)) &= g_{1;n}(x)g_{2;n}(x) \\ &= \frac{1}{2} \{ [g_{1;n}(x) + g_{2;n}(x)]^2 - [g_{1;n}^2(x) + g_{2;n}^2(x)] \} \\ &= \frac{1}{2} \{ [\operatorname{tr}(\Gamma_n(x))]^2 - \operatorname{tr}(\Gamma_n^2(x)) \},\end{aligned}$$

and we arrive at the following expansion

$$(6.6) \quad \frac{1}{\sqrt{\det(\Gamma_n(x) + I_2)}} = 1 - \frac{1}{2} \operatorname{tr}(\Gamma_n(x)) + \frac{1}{8} [\operatorname{tr}(\Gamma_n(x))]^2 + \frac{1}{4} \operatorname{tr}(\Gamma_n^2(x)) + O(\Gamma_n^3(x)).$$

In conclusion, in view of (6.1), (6.5) and (6.6), we obtain the Taylor expansion of the density function K_1 :

$$\begin{aligned}K_{1,n}(x) &= \frac{\sqrt{n}}{2^2\sqrt{\pi}} \left[1 + \frac{1}{2}s_n(x) + \frac{3}{8}s_n^2(x) \right] \times \\ &\times \left[\sqrt{2}\pi^{3/2} + \frac{3}{2^{3/2}}\pi^{3/2} \operatorname{tr}(\Gamma_n(x)) - \frac{9}{8 \cdot 2^{3/2}}\pi^{3/2} \operatorname{tr}(\Gamma_n^2(x)) + \frac{15\sqrt{2}}{8^2}\pi^{3/2} [\operatorname{tr}(\Gamma_n(x))]^2 \right] \\ &\times \left[1 - \frac{1}{2} \operatorname{tr}(\Gamma_n(x)) + \frac{1}{8} [\operatorname{tr}(\Gamma_n(x))]^2 + \frac{1}{4} \operatorname{tr}(\Gamma_n^2(x)) \right] + O(\sqrt{n} \cdot s_n^3(x)) + O(\sqrt{n} \cdot \Gamma_n^3(x)) \\ &= \frac{\sqrt{n}\pi^{3/2}}{2^2\sqrt{\pi}} \left[\sqrt{2} + \frac{s_n(x)}{\sqrt{2}} + \frac{1}{2\sqrt{2}} \operatorname{tr}(\Gamma_n(x)) + \frac{3}{4\sqrt{2}}s_n^2(x) + \frac{1}{4\sqrt{2}}s_n(x) \operatorname{tr}(\Gamma_n(x)) \right. \\ &\quad \left. - \frac{1}{16\sqrt{2}} \operatorname{tr}(\Gamma_n^2(x)) - \frac{1}{32\sqrt{2}} [\operatorname{tr}(\Gamma_n(x))]^2 \right] + O(\sqrt{n} \cdot s_n^3(x)) + O(\sqrt{n} \cdot \Gamma_n^3(x)).\end{aligned}$$

□

We state now Lemma 6.1; the proof of this lemma is relegated to Appendix B.

Lemma 6.1. *We have*

$$\begin{aligned}\mathcal{I}_0(\Gamma_n(x)) &= \sqrt{2}\pi^{3/2}, \\ \mathcal{I}_1(\Gamma_n(x)) &= \frac{3}{2^{3/2}}\pi^{3/2} \operatorname{tr}(\Gamma_n(x)), \\ \mathcal{I}_2(\Gamma_n(x)) &= -\frac{3}{2^{3/2}}\pi^{3/2} \operatorname{tr}(\Gamma_n^2(x)), \\ \mathcal{I}_3(\Gamma_n(x)) &= \frac{15\sqrt{2}}{64}\pi^{3/2} \{ 2 \operatorname{tr}(\Gamma_n^2(x)) + [\operatorname{tr}(\Gamma_n(x))]^2 \}.\end{aligned}$$

7. PROOF OF LEMMA 2.8: BOUNDARY EFFECT AND ERROR TERM

We first prove Lemma 2.8a. By the definition (2.17) of $L_n(x)$, we have

$$(7.1) \quad \int_{\mathcal{Q}} L_n(x) dx = \frac{\sqrt{n}\pi}{4\sqrt{2}} \int_{\mathcal{Q}} \left[s_n(x) + \frac{1}{2} \operatorname{tr}(\Gamma_n(x)) + \frac{3}{4}s_n^2(x) + \frac{1}{4}s_n(x) \operatorname{tr}(\Gamma_n(x)) \right] dx \\ + \frac{\sqrt{n}\pi}{4\sqrt{2}} \int_{\mathcal{Q}} \left[-\frac{1}{16} \operatorname{tr}(\Gamma_n^2(x)) - \frac{1}{32} [\operatorname{tr}(\Gamma_n(x))]^2 \right] dx.$$

The following technical lemma evaluates the individual integrals encountered within (7.1).

Lemma 7.1. *The integrals of the individual terms on the r.h.s. of (7.1) admit the following asymptotics.*

a.

$$\int_{\mathcal{Q}} s_n(x) dx = 0.$$

b.

$$\int_{\mathcal{Q}} \text{tr}(\Gamma_n(x)) dx = -\frac{6}{N_n} + O\left(\frac{1}{N_n^2}\right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|).$$

c.

$$\int_{\mathcal{Q}} s_n^2(x) dx = \frac{5}{N_n}.$$

d.

$$\int_{\mathcal{Q}} s_n(x) \text{tr}(\Gamma_n(x)) dx = \frac{2}{N_n} + O\left(\frac{1}{N_n^2}\right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|).$$

e.

$$\int_{\mathcal{Q}} \text{tr}(\Gamma_n^2(x)) dx = \frac{2^2}{N_n} [1 + 2^5 M_4(n)] + O\left(\frac{1}{N_n^2}\right) + O(n^{-1-1/2} N_n^{-l-3} |\mathcal{M}_l(n)|).$$

f.

$$\int_{\mathcal{Q}} [\text{tr}(\Gamma_n(x))]^2 dx = \frac{2^2}{N_n} [2^6 M_4(n) - 3] + O\left(\frac{1}{N_n^2}\right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|).$$

g.

$$\int_{\mathcal{Q}} s_n^3(x) dx = O\left(\frac{1}{N_n^2}\right).$$

h.

$$\int_{\mathcal{Q}} [\text{tr}(\Gamma_n(x))]^3 dx = O\left(\frac{1}{N_n^2}\right).$$

The proof of Lemma 7.1 is postponed till Section C.

Proof of Lemma 2.8 assuming Lemma 7.1. We substitute the various asymptotic statements of Lemma 7.1 into (7.1) to obtain

$$\begin{aligned} & \int_{\mathcal{Q}} L_n(x) dx \\ &= \frac{\sqrt{n}\pi}{4\sqrt{2}} \left\{ -\frac{1}{2} \frac{6}{N_n} + \frac{3}{4} \frac{5}{N_n} + \frac{1}{4} \frac{2}{N_n} - \frac{1}{16} \frac{2^2}{N_n} [1 + 2^5 M_4(n)] - \frac{1}{32} \frac{2^2}{N_n} [2^6 M_4(n) - 3] \right\} \\ &+ O\left(\frac{\sqrt{n}}{N_n^2}\right) + O(n^{-1/2} N_n^{-l-1} |\mathcal{M}_l(n)|) \\ &= \frac{\sqrt{n}\pi}{4\sqrt{2}} \frac{1}{N_n} \left[\frac{11}{2^3} - 2^4 M_4(n) \right] + O\left(\frac{\sqrt{n}}{N_n^2}\right) + O(n^{1/2} N_n^{-l-1} |\mathcal{M}_l(n)|). \end{aligned}$$

The latter formula, expressed in terms of the sequence $\{\widehat{\nu}_n(4)\}$ of the 4th Fourier coefficients of ν_n gives the statement of Lemma 2.8a. The proof of Lemma 2.8b follows immediately from (2.18) and Lemma 7.1 parts g and h.

□

8. PROOF OF THEOREM 1.3: LENGTH- l SPECTRAL SEMI-CORRELATIONS

We begin by proving Theorem 1.3 in the case of square-free numbers. To this end, for any fixed $K \geq 1$ we introduce the set

$$\Omega_{M,K} = \{n \leq M, \text{rad}(n) = n, p|n \in S \Rightarrow p \geq K\},$$

and let $\Omega_M := \Omega_{M,1} = S \cap [1, M]$. The following lemma is borrowed from [4].

Lemma 8.1. *For $m \in \Omega_{M,K}$, let $m = p_1 \cdot p_2 \cdots p_r$ be its factorization with $K < p_1 < p_2 < \cdots < p_r$. Then as $M \rightarrow \infty$ we have $p_s > 2^{s\Phi(s)}$ for $1 \leq s \leq r$ holds for all $m \in \Omega_{M,K} \setminus \Omega_{M,K}^{(1)}$, where the exceptional set $\Omega_{M,K}^{(1)}$ has cardinality*

$$|\Omega_{M,K}^{(1)}| \leq \eta(K, \Phi) |\Omega_{M,K}|$$

with $\eta(K, \Phi) \rightarrow 0$ as $K \rightarrow \infty$. If $\Phi(x) = o(\log x)$, then we can choose $\eta(K, \Phi) = K^{-1+\delta}$ for every fixed $\delta > 0$.

The next proposition shows that the number of solutions of (1.30) is small for almost all $m \in \Omega_{M,K}$.

Proposition 8.2. *Let $\delta > 0$ be fixed. If $K \geq K(\delta)$ and $M \rightarrow \infty$, then for all but $K^{-1+\delta} |\Omega_{M,K}|$ elements $m \in \Omega_{M,K}$ the equation (1.30) has $O(N_m^k)$ solutions.*

Proof. Let $\tilde{S} \in S$ be the subset for which (1.30) has a nontrivial solution. For any prime p we write $p = \pi \cdot \bar{\pi}$ where π is the corresponding Gaussian prime with $\arg(\pi) \in [0, \pi/2]$. For any integer $s \geq 1$ we introduce the set

$$\mathcal{F}_s = \left\{ n \in \Omega_{M,K}, \omega(n) = s, n \in \tilde{S}; \forall d \neq n, d|n \Rightarrow d \in S \setminus \tilde{S} \right\}.$$

Fix $s \geq 1$ and consider $n \in \mathcal{F}_s$ with a given factorization $n = p_1 \cdot p_2 \cdots p_s$, $K < p_1 < p_2 < \cdots < p_s$. We have that there exist integer points $\{\xi\}_{j=1}^{2k}$ with $\|\xi_j\| = \sqrt{n}$ and $\varepsilon_j \in \{-1, 0, 1\}$, $1 \leq j \leq 2k$ with vanishing linear combination

$$\sum_{j=1}^{2k} \varepsilon_j \text{Re}(\xi_j) = 0.$$

Each point ξ_r can be uniquely written as a product $\xi_r = i^{k_{\xi_r}} \prod_{j \leq s} \pi_{j,r}^*$ where each $\pi_{j,r}^* \in \{\pi_j, \bar{\pi}_j\}$ and $k_{\xi_r} \in \{0, 1, 2, 3\}$. We now regroup the terms in the last expression by collecting π_s and $\bar{\pi}_s$ into different summands to end up with an equivalent form

$$(8.1) \quad \text{Re}(\pi_s A_{s-1}) + \text{Re}(\bar{\pi}_s B_{s-1}) = 0,$$

where each A_{s-1}, B_{s-1} consists of the sum of at most $2k-1$ terms composed of first $(s-1)$ Gaussian primes. Setting $\pi_s = |\pi_s| e^{i\phi_s}$, $A_{s-1} = |A_{s-1}| e^{ia_{s-1}}$ and $B_{s-1} = |B_{s-1}| e^{ib_{s-1}}$, the equation (8.1) can be rewritten as

$$|A_{s-1}| \cos(\phi_s + a_{s-1}) + |B_{s-1}| \cos(b_{s-1} - \phi_s) = 0.$$

Using elementary trigonometric identities we get

$$\cos(\phi_s) (|A_{s-1}| \cos(a_{s-1}) + |B_{s-1}| \cos(b_{s-1})) = \sin(\phi_s) (|A_{s-1}| \sin(a_{s-1}) - |B_{s-1}| \sin(b_{s-1})).$$

Consequently,

$$\tan(\phi_s) = \frac{|A_{s-1}| \cos(a_{s-1}) + |B_{s-1}| \cos(b_{s-1})}{|A_{s-1}| \sin(a_{s-1}) - |B_{s-1}| \sin(b_{s-1})}$$

and so $\tan(\phi_s)$ is determined uniquely unless

$$|A_{s-1}| \cos(a_{s-1}) + |B_{s-1}| \cos(b_{s-1}) = |A_{s-1}| \sin(a_{s-1}) - |B_{s-1}| \sin(b_{s-1}) = 0.$$

In the latter case we must have

$$(8.2) \quad |A_{s-1}| \cos(a_{s-1}) + |B_{s-1}| \cos(b_{s-1}) = 0.$$

Since $n \in \mathcal{F}_s$ we have $p_1 p_2 \dots p_{s-1} = \tilde{n} |n$ and so by definition $\tilde{n} \in S \setminus \tilde{S}$. This implies that (8.2) and consequently (8.1) must be trivial with $A_{s-1} = B_{s-1} = 0$. This contradicts the definition of $n \in \mathcal{F}_s$. Hence $\tan(\phi_s)$ is determined uniquely and so is the corresponding prime p_s . Indeed, if $\pi_s^{(1)} = x^2 + y^2$ and $p_s^{(2)} = a^2 + b^2$ are two primes corresponding to the same angle ϕ_s , then

$$\tan(\phi_s) = \frac{a}{b} = \frac{x}{y}.$$

Since $(a, b) = (x, y) = 1$, we have that $|a| = |x|$ and $|b| = |y|$ and therefore $a^2 + b^2 = x^2 + y^2 = p_s^{(1)} = p_s^{(2)} := p_s$.

We are now ready to estimate the number of $m \in \Omega_{M,K}$ which give rise to a nontrivial solution of (1.30). By Lemma 8.1, we can restrict ourselves to $m = p_1 p_2 \dots p_r \in \Omega_{M,K}$, with $K < p_1 < p_2 < \dots < p_r$ and $p_j \geq 2^{j\Phi(j)}$ for any $1 \leq j \leq r$ and some slowly growing function $\Phi(x)$ to be determined later. Clearly for each such m , there exists unique $1 \leq s \leq r$ such that the product $p_1 p_2 \dots p_s \in \mathcal{F}_s$. Given $K < p_1 < p_2 < \dots < p_{s-1}$ we can form at most $2^{2k(s-1)}$ sums A_{s-1} and B_{s-1} and thus produce at most $2^{2k(s-1)}$ distinct $n = p_1 p_2 \dots p_{s-1} p_s \in \Omega_{M,K}$. By Lemma 8.1, $p_s \geq \max\{2^{s\Phi(s)}, K\}$ and therefore the total number of elements in $\tilde{S} \cap [1, M]$ induced by the elements in \mathcal{F}_s is at most

$$\ll 2^{2ks} \left\{ m \leq \frac{M}{\max\{K, 2^{s\Phi(s)}\}}, m \in \Omega_{M,K} \right\},$$

where the last bound comes from ‘‘conditioning’’ on the value of p_s and noting that $m/p_s \in \Omega_{M,K}$. Summing this over different ranges for s and choosing $\Phi(x) = o(\log x)$ in the same way as in [4] yields the desired conclusion. \square

We are now ready to handle the general case.

Proposition 8.3. *For all but $o\left(\frac{M}{\sqrt{\log M}}\right)$ elements $m \in S \cap [1, M]$, the equation (1.30) has $O(N_m^k)$ solutions.*

Proof. Fix large $K > 0$ and consider $\mathcal{P}_K = \prod_{p \leq K} p$. We decompose $n = n_K n_b$ where $(n_b, \mathcal{P}_K) = 1$ and $\text{rad}(n_K) | \mathcal{P}_K$. For any fixed part $n_K = \xi \tilde{\xi}$ we count the number of $n \in \tilde{S} \cap [1, M]$ with fixed $n_K | n$ and $(\frac{n}{n_K}, \mathcal{P}_K) = 1$. In this way (1.30) reduces to

$$\sum_{i=1}^{2k} \text{Re}(\alpha_i \xi_i) = 0$$

with $\alpha_i | n_K$ and $\|\xi_i\| = \sqrt{n_b}$ for $1 \leq i \leq 2k$. We now follow the proof of Proposition 8.2 regarding α_i as fixed coefficients. Let $\Omega_{4,1}(n)$ denote the number of prime divisors $p = 1 \pmod{4}$ of n counting multiplicity. Given n_K , we have at most $2^{2k\Omega_{4,1}(n_K)}$ choices for the coefficients α_i . Note that the number of integers $n \in \Omega_M$ for which $p^2 | n$ for some prime $p \geq K$, is bounded above by

$$\sum_{p > K} |\Omega_{\frac{M}{p^2}}| \ll \sum_{p > K} \frac{M}{p^2 \sqrt{\log M}} \ll \frac{M}{K \log K \sqrt{\log M}}$$

and thus give negligible contribution. Consequently, we can restrict ourselves to the set of integers with $\text{rad}(n_b) = n_b$. The number of $n \in \tilde{S} \cap [1, M]$ induced in this way, after appealing to Proposition 8.2 is bounded above by

$$\begin{aligned} \sum_{\text{rad}(n_K) | \mathcal{P}_K} 2^{2k\Omega_{4,1}(n_K)} |\Omega_{M/(n_K), K}| &\ll \sum_{\text{rad}(n_K) \in \mathcal{P}_K} \frac{4^{k\Omega_{4,1}(n_K)}}{n_K} \left((K^{-1+\delta} + o(1)) \frac{M}{\sqrt{\log M}} \right) \\ &\ll \prod_{p \leq K, p \equiv 1 \pmod{4}} \left(\sum_{j \geq 0} \frac{4^{kj}}{p^j} \right) (K^{-1+\delta} + o(1)) \frac{M}{\sqrt{\log M}} \\ &\ll \left(\frac{(\log K)^{2k+1}}{K^{1-\delta}} + o(1) \right) \frac{M}{\sqrt{\log M}}. \end{aligned}$$

The result now follows by letting $K \rightarrow \infty$. \square

Proposition 8.3 now implies that $|\tilde{S} \cap [1, M]| = o\left(\frac{M}{\sqrt{\log M}}\right)$ and we may take $S' = S \setminus \tilde{S}$. In order to prove the last part of Theorem 1.3, we require the following classical result of Kubilius.

Lemma 8.4. *The number of Gaussian primes ω in the sector $0 \leq \alpha \leq \arg \omega \leq \beta \leq 2\pi$, $|\omega|^2 \leq u$ is equal to*

$$\frac{2}{\pi}(\beta - \alpha) \int_2^u \frac{1}{\log v} dv + O\left(u \exp(-b\sqrt{\log u})\right),$$

for some positive constant $b \in \mathbb{R}$.

We now choose particular “thin” subset of S , to guarantee the desired limiting behaviour of $\{\widehat{\nu}_n(4)\}_{n \in S'}$.

Proposition 8.5. *For any $s \in [-1, 1]$, there exists a sequence $\{n_i\}_{i \geq 1}$, with $N_{n_i} \rightarrow \infty$ whenever $i \rightarrow \infty$, such that $\widehat{\nu}_{n_i}(4) \rightarrow s$ and equation (1.30) has $O(N_{n_i}^k)$ solutions for any $i \geq 1$.*

Proof. Fix large $m \geq 1$ and small $\varepsilon > 0$. By Lemma 8.4, we can select prime $p_n = \pi_n \bar{\pi}_n$, $p_n \equiv 1 \pmod{4}$ and $|\arg(\pi_n)| \leq \frac{\varepsilon}{100m}$. We further select prime p such that

$$(8.3) \quad |\widehat{\nu}_p(4) - s| \leq \frac{\varepsilon}{2}.$$

Consider the number of the form $n = p_n^m p$. It is easy to see that

$$(8.4) \quad |\widehat{\nu}_{p_n}(4) - 1| \leq \frac{\varepsilon}{2m}$$

and $N_n > 2^m$. Using an elementary inequality

$$\left| \prod_{j \leq r} x_j - \prod_{j \leq r} y_j \right| \leq \sum_{j \leq r} |x_j - y_j|,$$

valid for $|x_j|, |y_j| \leq 1$ and the fact that $\widehat{\nu}_n(4) = (\widehat{\nu}_{p_n}(4))^m \widehat{\nu}_p(4)$, we can estimate

$$|\widehat{\nu}_n(4) - s| \leq m|\widehat{\nu}_{p_n}(4) - 1| + |\widehat{\nu}_p(4) - s| \leq m \cdot \frac{\varepsilon}{2m} + \frac{\varepsilon}{2} = \varepsilon.$$

We are left to show that equation (1.30) has only trivial solutions for appropriately chosen values of p_n, p , which satisfy (8.3) and (8.4). Let $\pi_n = r_n e^{i\phi}$ and $p = \pi \cdot \bar{\pi}$ with $\arg \pi = \alpha$. Clearly each integer point on the circle of radius \sqrt{n} can be written as $\xi_j = \sqrt{n} e^{i(j\phi \pm \alpha + r\frac{\pi}{2})}$ for some $|j| \leq m$ and $r = \{0, 1, 2, 3\}$. For such defined n , upon taking real parts, equation (1.30) can be rewritten in the form

$$(8.5) \quad \sum_{j=1}^{2k} \varepsilon_j \cos\left(l_j \phi \pm \alpha + \frac{\pi r_j}{2}\right) = 0,$$

where $\varepsilon_j = \{+1, -1\}$ and $|l_j| \leq m$ for $1 \leq j \leq 2k$. By collecting terms with equal phases $l_j \phi$ and using elementary trigonometric identities, we can rewrite (8.5) in the form

$$(8.6) \quad F_r(\phi) = \sum_{j=1}^r \cos(m_j \phi) (\alpha_j \cos \alpha + \beta_j \sin \alpha) + \sin(m_j \phi) (\alpha_j^{(1)} \cos \alpha + \beta_j^{(1)} \sin \alpha) = 0,$$

where $1 \leq r \leq 2k$ and $0 \leq m_1 < m_2 < \dots < m_r$ with $-2k \leq \alpha_j, \alpha_j^{(1)}, \beta_j, \beta_j^{(1)} \leq 2k$. Since k is fixed, there are only finitely many choices for the coefficients $\alpha_j, \alpha_j^{(1)}, \beta_j, \beta_j^{(1)}$ and therefore we can select angle α for which the corresponding prime p satisfies (8.3) and such that $a \sin \alpha + b \cos(\alpha) \neq 0$ for all $a, b \in \mathbb{Z}$ with $|a| + |b| \neq 0$ and $|a|, |b| \leq 2k$. Now since each $F_r(\phi)$ is a trigonometric polynomial of a total degree at most $4k$, each non degenerate equation (8.6) has at most $4k$ solutions. Since there are only finitely many of such equations, we can select p_n sufficiently large which satisfies (8.4) and the corresponding equation (8.5) has only trivial solutions. This concludes the proof. \square

Combining Proposition 8.3 and Proposition 8.5 completes the proof of Theorem 1.3.

Remark 8.6. It is possible to give a construction of n in Proposition 8.5 which is square-free. The idea is to use Lemma 8.4 and choose inductively sequence of primes $p_1 < p_2 < \dots < p_m$ such that $p_j = \pi_j \bar{\pi}_j$ and $|\arg(\pi_j)| \leq \frac{1}{m^2}$ with the property that for any $1 \leq r \leq m$ we have $\cos(r \cdot \arg(\pi_m)) \notin \text{span}_{\mathbb{N}}\{\cos(\sum_{i \leq m-1} a_i \arg(\pi_i))\}$ where $-m \leq a_i \leq m$, $a_i \in \mathbb{N}$. The latter can be ensured by taking sufficiently sparse sequence of primes p_1, p_2, \dots . Now select $n = p \cdot p_1 \dots p_m$ with $|\hat{\mu}_p(4) - s| \leq \delta$ and δ sufficiently small. We leave the details to the interested reader.

APPENDIX A. PROOF OF LEMMA 2.2: EVALUATING $\Omega_n(\cdot)$, THE NORMALISED CONDITIONAL COVARIANCE MATRIX

The ultimate goal of this section is evaluating the 2×2 (normalised) covariance matrix $\Omega_n(\cdot)$ in (2.9) of $\nabla f_n(x)$, conditioned upon $f_n(x) = 0$. First, we will need to evaluate (§A.1) the (unconditional) 3×3 covariance matrix $\Sigma_n(x)$ of $(f_n(x), \nabla f_n(x))$, and then apply (§A.2) the well-known procedure for conditioning in the Gaussian case.

A.1. Evaluating the unconditional covariance matrix.

Lemma A.1. *Let $\Sigma_n(x)$ be the 3×3 covariance matrix of $(f_n(x), \nabla f_n(x))$. Then $\Sigma_n(x)$ could be expressed as*

$$\Sigma_n(x) = \begin{pmatrix} v_n(x) & B_n(x) \\ B_n^t(x) & C_n(x) \end{pmatrix},$$

where $v_n(\cdot) = \text{Var}(f_n(x))$ is as in (2.7),

$$B_n(x) = \mathbb{E}[f_n(x) \cdot \nabla_y f_n(y)] \Big|_{x=y}$$

and

$$C_n(x) = \mathbb{E}[\nabla_x f_n(x) \otimes \nabla_y f_n(y)] \Big|_{x=y}.$$

The 2×1 matrix $B_n^t(x)$ is:

$$(A.1) \quad B_n^t(x) = \frac{4\pi}{N_n} \left(\frac{\sum_{\mu \in \mathcal{E}_n/\sim} \mu_1 \sin(2\mu_1 \pi x_1) [1 - \cos(2\mu_2 \pi x_2)]}{\sum_{\mu \in \mathcal{E}_n/\sim} \mu_2 \sin(2\mu_2 \pi x_2) [1 - \cos(2\mu_1 \pi x_1)]} \right),$$

and the entries of the 2×2 matrix $C_n(x)$ is given by:

$$C_{n;11}(x) = \frac{n\pi^2}{2} + \frac{4\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1^2 \left[\cos(2\mu_1 \pi x_1) - \cos(2\mu_2 \pi x_2) - \cos(2\mu_1 \pi x_1) \cos(2\mu_2 \pi x_2) \right]$$

$$C_{n;12}(x) = C_{n;21}(x) = \frac{4\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1 \mu_2 \sin(2\mu_1 \pi x_1) \sin(2\mu_2 \pi x_2),$$

and

$$C_{n;22}(x) = \frac{n\pi^2}{2} + \frac{4\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_2^2 \left[\cos(2\mu_2 \pi x_2) - \cos(2\mu_1 \pi x_1) - \cos(2\mu_2 \pi x_2) \cos(2\mu_1 \pi x_1) \right].$$

Proof. First, that the $(1, 1)$ entry of $\Sigma_n(x)$ is $v_n(x)$ is self-evident, being the variance of $f_n(x)$. For the other elements of $\Sigma_n(x)$, recalling the covariance function $r_n(x, y)$ in (1.22), we have that

$$(A.2) \quad B_n(x) = \nabla_y r_n(x, y) \Big|_{x=y},$$

and

$$(A.3) \quad C_n(x) = (\nabla_x \otimes \nabla_y) r_n(x, y) \Big|_{x=y}.$$

What follows below is the, somewhat technically demanding, routine, evaluation of the derivatives on the r.h.s. (A.2) and (A.3).

We first compute the entries of the 1×2 matrix $B_n(x)$. Note that, without stationarity (nor unit variance), successive derivatives are not uncorrelated at a fixed point. We have

$$\begin{aligned} \frac{\partial}{\partial y_1} r_n(x, y) \Big|_{x=y} &= \frac{16}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \sin(\mu_1 \pi x_1) \sin(\mu_2 \pi x_2) \frac{\partial}{\partial y_1} \sin(\mu_1 \pi y_1) \sin(\mu_2 \pi y_2) \Big|_{x=y} \\ &= \frac{16}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \sin(\mu_1 \pi x_1) \sin(\mu_2 \pi x_2) \cos(\mu_1 \pi y_1) \sin(\mu_2 \pi y_2) \mu_1 \pi \Big|_{x=y} \\ &= \frac{16\pi}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \sin(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \cos(\mu_1 \pi x_1) \mu_1, \end{aligned}$$

since $\sin(2\theta) = 2 \sin \theta \cos \theta$, we can also write

$$\frac{\partial}{\partial y_1} r_n(x, y) \Big|_{x=y} = \frac{8\pi}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \sin(2\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \mu_1.$$

Similarly

$$\frac{\partial}{\partial y_2} r_n(x, y) \Big|_{x=y} = \frac{8\pi}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \sin(2\mu_2 \pi x_2) \sin^2(\mu_1 \pi x_1) \mu_2.$$

Further, since $2 \sin^2 \theta = 1 - \cos(2\theta)$, we have

$$\frac{\partial}{\partial y_1} r_n(x, y) \Big|_{x=y} = \frac{4\pi}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1 \sin(2\mu_1 \pi x_1) [1 - \cos(2\mu_2 \pi x_2)]$$

and

$$\frac{\partial}{\partial y_2} r_n(x, y) \Big|_{x=y} = \frac{4\pi}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_2 \sin(2\mu_2 \pi x_2) [1 - \cos(2\mu_1 \pi x_1)],$$

which is (A.1). Note that the both entries of $B_n(x)$ are purely oscillatory (though do not vanish identically, as in the stationary case).

The entries of the matrix $C_n(x)$ are

$$\begin{aligned} \frac{\partial}{\partial y_1} \frac{\partial}{\partial x_1} r_n(x, y) \Big|_{x=y} &= \frac{16\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 \cos(\mu_1 \pi x_1) \sin(\mu_2 \pi x_2) \cos(\mu_1 \pi y_1) \sin(\mu_2 \pi y_2) \Big|_{x=y} \\ &= \frac{16\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 \cos^2(\mu_1 \pi x_1) \sin^2(\mu_2 \pi x_2) \\ &= \frac{4\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 + \frac{4\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 [\cos(2\mu_1 \pi x_1) - \cos(2\mu_2 \pi x_2) - \cos(2\mu_1 \pi x_1) \cos(2\mu_2 \pi x_2)] \\ &= \frac{n\pi^2}{2} + \frac{4\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 [\cos(2\mu_1 \pi x_1) - \cos(2\mu_2 \pi x_2) - \cos(2\mu_1 \pi x_1) \cos(2\mu_2 \pi x_2)], \end{aligned}$$

where we reused the identity $2 \sin^2 \theta = 1 - \cos(2\theta)$ and $2 \cos^2 \theta = 1 + \cos(2\theta)$, and applied the equality

$$\sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 = \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 \pm \sum_{\mu \in \mathcal{E}_n / \sim} \mu_2^2 = n \frac{N_n}{4} - \sum_{\mu \in \mathcal{E}_n / \sim} \mu_2^2 = n \frac{N_n}{4} - \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2.$$

Similarly we have

$$\begin{aligned} \frac{\partial}{\partial y_2} \frac{\partial}{\partial x_2} r_n(x, y) \Big|_{x=y} &= \frac{16\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_2^2 \sin^2(\mu_1 \pi x_1) \cos^2(\mu_2 \pi x_2) \\ &= \frac{n\pi^2}{2} + \frac{4\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_2^2 \left[\cos(2\mu_2 \pi x_2) - \cos(2\mu_1 \pi x_1) - \cos(2\mu_2 \pi x_2) \cos(2\mu_1 \pi x_1) \right]. \end{aligned}$$

The off-diagonal entries of $C_n(x)$ are both equal to

$$\begin{aligned} \frac{\partial}{\partial y_1} \frac{\partial}{\partial x_2} r_n(x, y) \Big|_{x=y} &= \frac{16\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1 \mu_2 \sin(\mu_1 \pi x_1) \cos(\mu_2 \pi x_2) \cos(\mu_1 \pi y_1) \sin(\mu_2 \pi y_2) \Big|_{x=y} \\ &= \frac{16\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1 \mu_2 \sin(\mu_1 \pi x_1) \cos(\mu_2 \pi x_2) \cos(\mu_1 \pi x_1) \sin(\mu_2 \pi x_2) \\ &= \frac{4\pi^2}{N_n} \sum_{\mu \in \mathcal{E}_n/\sim} \mu_1 \mu_2 \sin(2\mu_1 \pi x_1) \sin(2\mu_2 \pi x_2). \end{aligned}$$

□

A.2. Proof of Lemma 2.2.

Proof. Let $\Theta_n(x)$ be the conditional covariance matrix of the Gaussian vector $(\nabla f_n(x) | f_n(x) = 0)$, related to $\Omega_n(x)$ via the normalisation

$$(A.4) \quad \Omega_n(x) = \frac{2}{n\pi^2} \Theta_n(x).$$

Once the (unconditional) covariance matrix Σ_n of $(f_n(x), \nabla f_n(x))$ is, following Lemma A.1, known, the conditional covariance matrix $\Theta_n(x)$ is given by the standard Gaussian transition formula:

$$\Theta_n(x) = C_n(x) - \frac{1}{v_n(x)} B_n^t(x) B_n(x).$$

We use (A.1) to compute

$$\frac{1}{v_n(x)} B_n^t(x) B_n(x) = \frac{64\pi^2}{N_n^2 v_n(x)} \begin{pmatrix} d_{1;n}(x)^2 & d_{1;n}(x) \cdot d_{2;n}(x) \\ d_{1;n}(x) \cdot d_{2;n}(x) & d_{2;n}(x)^2 \end{pmatrix},$$

and then, on recalling the notation (2.11)-(2.12), we have

$$\Theta_n(x) = \frac{n\pi^2}{2} I_2 + \frac{4\pi^2}{N_n} \begin{pmatrix} b_{11;n}(x) & b_{12;n}(x) \\ b_{21;n}(x) & b_{22;n}(x) \end{pmatrix} - \frac{64\pi^2}{N_n^2 v_n(x)} \begin{pmatrix} d_{1;n}(x)^2 & d_{1;n}(x) \cdot d_{2;n}(x) \\ d_{1;n}(x) \cdot d_{2;n}(x) & d_{2;n}(x)^2 \end{pmatrix}.$$

All that remains to obtain an explicit expression for $\Omega_n(x)$ from $\Theta_n(x)$ in the form prescribed in (2.9) (and (2.10)-(2.12)) is merely to invoke the normalisation (A.4):

$$\Omega_n(x) = \frac{2}{n\pi^2} \Theta_n(x) = I_2 + \Gamma_n(x).$$

□

APPENDIX B. PROOF OF LEMMA 6.1: AUXILIARY RESULTS FOR PROPOSITION 2.7

Proof. Recall the notation (6.3)-(6.4) for whatever quantities are been evaluated in turn. A transformation of variables shows at once that

$$\mathcal{I}_0(\Gamma_n(x)) = \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} dz = 2\pi \int_0^\infty \exp \left\{ -\frac{1}{2} \rho^2 \right\} \rho^2 d\rho = 2\pi \sqrt{\frac{\pi}{2}} = \pi^{3/2} \sqrt{2}.$$

Let $\Gamma_n(x) = (\gamma_{ij;n}(x))_{ij=1,2}$, with $\gamma_{ij;n}(x)$ the entries of $\Gamma_n(x)$ (cf. (2.10)). We have

$$\begin{aligned} \mathcal{I}_1(\Gamma_n(x)) &= \frac{1}{2} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} z \Gamma_n(x) z^t dz \\ &= \frac{1}{2} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} (z_1^2 \gamma_{11;n}(x) + z_2^2 \gamma_{22;n}(x) + 2z_1 z_2 \gamma_{12;n}(x)) dz \\ &= a_1 \operatorname{tr}(\Gamma_n(x)) + a_2 \gamma_{12;n}(x), \end{aligned}$$

where

$$\begin{aligned} a_1 &= \frac{1}{2} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} z_1^2 dz = \frac{1}{2} 2\pi \int_0^\infty \rho \exp \left\{ -\frac{\rho^2}{2} \right\} \frac{\rho^2}{2} \rho d\rho = 3 \frac{\pi^{3/2}}{2^{3/2}}, \\ a_2 &= \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} z_1 z_2 dz = 0. \end{aligned}$$

Further,

$$\begin{aligned} \mathcal{I}_2(\Gamma_n(x)) &= -\frac{1}{2} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} z \Gamma_n^2(x) z^t dz \\ &= -\frac{1}{2} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} (\gamma_{11;n}^2(x) z_1^2 + \gamma_{12;n}^2(x) z_1^2 + 2\gamma_{11;n}(x) \gamma_{12;n}(x) z_1 z_2 \\ &\quad + 2\gamma_{12;n}(x) \gamma_{22;n}(x) z_1 z_2 + \gamma_{12;n}^2(x) z_2^2 + \gamma_{22;n}^2(x) z_2^2) dz \\ &= -\frac{1}{2} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} \{ z_1^2 [\gamma_{11;n}^2(x) + \gamma_{12;n}^2(x)] + z_2^2 [\gamma_{12;n}^2(x) + \gamma_{22;n}^2(x)] \} dz \\ &= -a_1 [\gamma_{11;n}^2(x) + \gamma_{12;n}^2(x)] - a_1 [\gamma_{22;n}^2(x) + \gamma_{12;n}^2(x)] = -a_1 \operatorname{tr}(\Gamma_n^2(x)). \end{aligned}$$

Finally,

$$\begin{aligned} \mathcal{I}_3(\Gamma_n(x)) &= \frac{1}{8} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{z z^t}{2} \right\} (z \Gamma_n(x) z^t)^2 dz \\ &= \frac{1}{8} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{z z^t}{2} \right\} [\gamma_{11;n}^2(x) z_1^4 + \gamma_{22;n}^2(x) z_2^4 + 4\gamma_{12;n}^2(x) z_1^2 z_2^2 \\ &\quad + 2\gamma_{11;n}(x) \gamma_{22;n}(x) z_1^2 z_2^2 + 4\gamma_{11;n}(x) \gamma_{12;n}(x) z_1^3 z_2 + 4\gamma_{12;n}(x) \gamma_{22;n}(x) z_1 z_2^3] dz \\ &= \frac{1}{8} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{z z^t}{2} \right\} [\gamma_{11;n}^2(x) z_1^4 + \gamma_{22;n}^2(x) z_2^4 + 4\gamma_{12;n}^2(x) z_1^2 z_2^2 \\ &\quad + 2\gamma_{11;n}(x) \gamma_{22;n}(x) z_1^2 z_2^2] dz. \end{aligned} \tag{B.1}$$

We write

$$\begin{aligned} \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} (z_1^2 + z_2^2)^2 dz &= 2\pi \int_0^\infty \rho \exp \left\{ -\frac{\rho^2}{2} \right\} \rho^4 \rho d\rho = 2 \frac{15}{\sqrt{2}} \pi^{3/2}, \\ \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} z_1^4 dz &= \frac{15}{\sqrt{2}} \frac{3}{4} \pi^{3/2}, \end{aligned} \tag{B.2}$$

and, using these, we also have

$$(B.3) \quad \iint_{\mathbb{R}^2} |z| \exp \left\{ -\frac{1}{2} z z^t \right\} z_1^2 z_2^2 dz = \frac{15}{\sqrt{2}} \frac{1}{4} \pi^{3/2}.$$

Substituting (B.2)-(B.3) into (B.1) we obtain:

$$\begin{aligned} \mathcal{I}_3(\Gamma_n(x)) &= \frac{1}{8} \frac{15}{4\sqrt{2}} \pi^{3/2} [3\gamma_{11;n}^2(x) + 3\gamma_{22;n}^2(x) + 4\gamma_{12;n}^2(x) + 2\gamma_{11;n}(x)\gamma_{22;n}(x)] \\ &= \frac{1}{8} \frac{15\sqrt{2}}{8} \pi^{3/2} [(2\gamma_{11;n}^2(x) + 2\gamma_{22;n}^2(x) + 4\gamma_{12;n}^2(x)) + (\gamma_{11;n}^2(x) + \gamma_{22;n}^2(x) \\ &\quad + 2\gamma_{11;n}(x)\gamma_{22;n}(x))] = \frac{15\sqrt{2}}{64} \pi^{3/2} [2 \operatorname{tr}(\Gamma_n^2(x)) + [\operatorname{tr}(\Gamma_n(x))]^2], \end{aligned}$$

precisely as claimed. \square

APPENDIX C. PROOF OF LEMMA 7.1: AUXILIARY RESULTS FOR THE PROOF OF LEMMA 2.8

Proof. We first observe that

$$\int_{\mathcal{Q}} s_n(x) dx = 0, \quad \int_{\mathcal{Q}} s_n^2(x) dx = \frac{4^2}{N_n^2} \sum_{\mu \in \mathcal{E}_n / \sim} \frac{5}{4} = \frac{5}{4} \frac{4^2}{N_n^2} \frac{N_n}{4} = \frac{5}{N_n}.$$

Now we turn to

$$\operatorname{tr}(\Gamma_n(x)) = \frac{8}{nN_n} [b_{11;n}(x) + b_{22;n}(x)] - \frac{128}{nN_n^2 v_n(x)} [d_{1;n}^2(x) + d_{2;n}^2(x)].$$

We observe that

$$\int_{\mathcal{Q}} b_{11;n}(x) dx = 0, \quad \int_{\mathcal{Q}} b_{22;n}(x) dx = 0.$$

Moreover, recalling that [23, Lemma 2.3]

$$\sum_{\mu \in \mathcal{E}_n / \sim} \mu_i \mu_j = \frac{n}{2} \frac{N_n}{4} \delta_{i,j},$$

we have

$$\int_{\mathcal{Q}} d_{2;n}^2(x) dx = \int_{\mathcal{Q}} d_{1;n}^2(x) dx = \frac{3}{16} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 = \frac{3}{16} \frac{n}{2} \frac{N_n}{4} = \frac{3}{2^7} n N_n,$$

and also

$$\int_{\mathcal{Q}} d_{1;n}^2(x) s_n(x) dx = \int_{\mathcal{Q}} d_{2;n}^2(x) s_n(x) dx = O(n).$$

Consolidating all the above estimates, we write

$$\begin{aligned} \int_{\mathcal{Q}} \operatorname{tr}(\Gamma_n(x)) dx &= -\frac{2^7}{nN_n^2} \int_{\mathcal{Q}} \frac{1}{v_n(x)} [d_{1;n}^2(x) + d_{2;n}^2(x)] dx \\ &= -\frac{2^7}{nN_n^2} \int_{\mathcal{Q}} [d_{1;n}^2(x) + d_{2;n}^2(x)] [1 + O(s_n(x))] dx + O(N_n^{2-l} |\mathcal{M}_l(n)|) \\ &= -\frac{2^7}{nN_n^2} \left(2 \frac{3}{2^7} n N_n + O(n) \right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|) \\ &= -\frac{6}{N_n} + O\left(\frac{1}{N_n^2}\right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|). \end{aligned}$$

And, since

$$\int_{\mathcal{Q}} s_n(x) b_{11;n}(x) dx = \int_{\mathcal{Q}} s_n(x) b_{22;n}(x) dx = \frac{4}{N_n} \sum_{\mu \in \mathcal{E}_n / \sim} \frac{\mu_1^2}{4} = \frac{4}{N_n} \frac{1}{4} \frac{n N_n}{8} = \frac{n}{8},$$

we have

$$\begin{aligned} & \int_{\mathcal{Q}} s_n(x) \operatorname{tr}(\Gamma_n(x)) dx \\ &= \frac{8}{n N_n} \int_{\mathcal{Q}} s_n(x) [b_{11;n}(x) + b_{22;n}(x)] dx - \frac{2^7}{n N_n^2} \int_{\mathcal{Q}} \frac{s_n(x)}{v_n(x)} [d_{1;n}^2(x) + d_{2;n}^2(x)] dx \\ &= \frac{8}{n N_n} \int_{\mathcal{Q}} s_n(x) [b_{11;n}(x) + b_{22;n}(x)] dx + O\left(\frac{1}{N_n^2}\right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|) \\ &= \frac{8}{n N_n} 2 \frac{n}{8} + O\left(\frac{1}{N_n^2}\right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|) = \frac{2}{N_n} + O\left(\frac{1}{N_n^2}\right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|). \end{aligned}$$

Moreover,

$$\begin{aligned} [\operatorname{tr}(\Gamma_n(x))]^2 &= \frac{8^2}{n^2 N_n^2} [b_{11;n}^2(x) + b_{22;n}^2(x) + 2b_{11;n}(x)b_{22;n}(x)] \\ &\quad + \frac{128^2}{n^2 N_n^4 v_n^2(x)} [d_{1;n}^4(x) + d_{2;n}^4(x) + 2d_{1;n}^2(x)d_{2;n}^2(x)] \\ &\quad - 2 \frac{8}{n N_n} \frac{128}{n N_n^2 v_n(x)} [b_{11;n}(x) + b_{22;n}(x)] [d_{1;n}^2(x) + d_{2;n}^2(x)]. \end{aligned}$$

and

$$\begin{aligned} \int_{\mathcal{Q}} [\operatorname{tr}(\Gamma_n(x))]^2 dx &= \frac{8^2}{n^2 N_n^2} \int_{\mathcal{Q}} [b_{11;n}^2(x) + b_{22;n}^2(x) + 2b_{11;n}(x)b_{22;n}(x)] dx \\ &\quad + \frac{128^2}{n^2 N_n^4} \int_{\mathcal{Q}} [d_{1;n}^4(x) + d_{2;n}^4(x) + 2d_{1;n}^2(x)d_{2;n}^2(x)] [1 + O(s_n(x))] dx \\ &\quad - 2 \frac{8}{n N_n} \frac{128}{n N_n^2} \int_{\mathcal{Q}} [b_{11;n}(x) + b_{22;n}(x)] [d_{1;n}^2(x) + d_{2;n}^2(x)] [1 + O(s_n(x))] dx \\ &\quad + O(N_n^{-l-1} |\mathcal{M}_l(n)|), \end{aligned}$$

where

$$\int_{\mathcal{Q}} b_{11;n}^2(x) dx = \int_{\mathcal{Q}} b_{22;n}^2(x) dx = \frac{5}{4} \sum_{\mu \in \mathcal{E} / \sim} \mu_1^4 =: \frac{5}{4} n^2 N_n M_4(n),$$

and, since

$$\sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 \mu_2^2 = \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 (\mu_2^2 \pm \mu_1^2) = \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 (n - \mu_1^2) = \frac{n^2 N_n}{8} - n^2 N_n M_4(n),$$

we have

$$\int_{\mathcal{Q}} b_{11;n}(x) b_{22;n}(x) dx = -\frac{3}{4} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 \mu_2^2 = -\frac{3}{4} \left(\frac{n^2 N_n}{8} - n^2 N_n M_4(n) \right).$$

Further, since for $i, j = 1, 2$,

$$\sum_{\substack{\eta, \mu \in \mathcal{E}_n / \sim \\ \eta \neq \mu}} \mu_j^2 \eta_i^2 = \sum_{\mu \in \mathcal{E}_n / \sim} \mu_j^2 \left(\sum_{\eta \in \mathcal{E}_n / \sim} \eta_i^2 - \mu_j^2 \right) = \frac{n^2 N_n^2}{8^2} - n^2 N_n M_4(n),$$

we have

$$\int_{\mathcal{Q}} d_{1;n}^4(x) dx = \int_{\mathcal{Q}} d_{2;n}^4(x) dx = \frac{105}{1024} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^4 + \frac{9}{256} \sum_{\substack{\eta, \mu \in \mathcal{E}_n / \sim \\ \eta \neq \mu}} \eta_1^2 \mu_1^2 = O(n^2 N_n^2),$$

$$\int_{\mathcal{Q}} d_{1;n}^2(x) d_{2;n}^2(x) dx = \frac{25}{1024} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 \mu_2^2 + \frac{9}{256} \sum_{\substack{\eta, \mu \in \mathcal{E}_n / \sim \\ \eta \neq \mu}} \eta_2^2 \mu_1^2 = O(n^2 N_n^2),$$

and similarly, for $i, j = 1, 2$,

$$\int_{\mathcal{Q}} b_{ii;n}(x) d_{j;n}^2(x) dx = O(n^2 N_n^2),$$

that implies

$$\begin{aligned} \int_{\mathcal{Q}} [\text{tr}(\Gamma_n(x))]^2 dx &= \frac{8^2}{n^2 N_n^2} \left[2 \frac{5}{4} n^2 N_n M_4(n) - 2 \frac{3}{4} \left(\frac{n^2 N_n}{8} - n^2 N_n M_4(n) \right) \right] \\ &\quad + O\left(\frac{1}{N_n^2}\right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|) \\ &= \frac{8^2}{n^2 N_n^2} \left[2 \frac{5}{4} n^2 N_n M_4(n) - 2 \frac{3}{4} \left(\frac{n^2 N_n}{8} - n^2 N_n M_4(n) \right) \right] \\ &= \frac{2^2}{N_n} (2^6 M_4(n) - 3) + O\left(\frac{1}{N_n^2}\right) + O(N_n^{-l-1} |\mathcal{M}_l(n)|). \end{aligned}$$

Finally, since for a symmetric matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

$\text{tr}(A^2) = a^2 + 2b^2 + c^2$, one has

$$\begin{aligned} \text{tr}(\Gamma_n^2(x)) &= \left[\frac{8}{n N_n} b_{11;n}(x) - \frac{128}{n N_n^2 v_n(x)} d_{1;n}^2(x) \right]^2 + \left[\frac{8}{n N_n} b_{22;n}(x) - \frac{128}{n N_n^2 v_n(x)} d_{2;n}^2(x) \right]^2 \\ &\quad + 2 \left[\frac{8}{n N_n} b_{12;n}(x) - \frac{128}{n N_n^2 v_n(x)} d_{1;n}(x) d_{2;n}(x) \right]^2 \end{aligned}$$

and, observing that

$$\int_{\mathcal{Q}} b_{12;n}^2(x) dx = \frac{1}{4} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^2 \mu_2^2 = \frac{1}{4} \left(\frac{n^2 N_n}{8} - n^2 N_n M_4(n) \right),$$

$$\int_{\mathcal{Q}} b_{12;n}(x) d_{1;n}(x) d_{2;n}(x) dx = O(n^2 N_n),$$

we have

$$\begin{aligned} \int_{\mathcal{Q}} \operatorname{tr}(\Gamma_n^2(x)) dx &= 2 \frac{8^2}{n^2 N_n^2} \left[\int_{\mathcal{Q}} b_{11;n}^2(x) dx + \int_{\mathcal{Q}} b_{12;n}^2(x) dx \right] + O\left(\frac{1}{N_n^2}\right) + O(n^{-2} N_n^{-l-2} |\mathcal{M}_l(n)|) \\ &= 2 \frac{8^2}{n^2 N_n^2} \left[\frac{5}{4} n^2 N_n M_4(n) + \frac{1}{4} \left(\frac{n^2 N_n}{8} - n^2 N_n M_4(n) \right) \right] + O\left(\frac{1}{N_n^2}\right) + O(n^{-2} N_n^{-l-2} |\mathcal{M}_l(n)|) \\ &= \frac{2^2}{N_n} [1 + 2^5 M_4(n)] + O\left(\frac{1}{N_n^2}\right) + O(n^{-1-1/2} N_n^{-l-3} |\mathcal{M}_l(n)|). \end{aligned}$$

To prove parts g and h of Lemma 7.1 we evaluate

$$\int_{\mathcal{Q}} s_n^3(x) dx = \frac{4^3}{N_n^3} \sum_{\mu \in \mathcal{E}_n / \sim} \left(-\frac{3}{2}\right) = O\left(\frac{1}{N_n^2}\right),$$

and

$$\int_{\mathcal{Q}} [\operatorname{tr}(\Gamma_n(x))]^3 dx = O\left(\frac{1}{n^3 N_n^3} \sum_{\mu \in \mathcal{E}_n / \sim} \mu_1^6\right) = O\left(\frac{1}{N_n^2}\right).$$

□

REFERENCES

- [1] Azaïs, J.-M., Wschebor, W. *Level Sets and Extrema of Random Processes and Fields*. John Wiley & Sons Inc., Hoboken, NJ, 2009.
- [2] Berry, M. V. *Regular and irregular semiclassical wavefunctions*. J. Phys. A **10**, no. 12, (1977), 2083–2091.
- [3] Berry, M. V. *Statistics of nodal lines and points in chaotic quantum billiards: perimeter corrections, fluctuations, curvature*. J. Phys. A **35** (2002), 3025–3038.
- [4] Bombieri, E., Bourgain, J. *A problem on sums of two squares*. International Math. Res. Notices **11** (2015), 3343–3407.
- [5] Benatar, J., Marinucci, D., Wigman, I. *Planck-scale distribution of nodal length of arithmetic random waves*. J. d'Anal. Math., to appear. Available online <https://arxiv.org/abs/1710.06153>.
- [6] Brüning, J. *Über Knoten Eigenfunktionen des Laplace-Beltrami Operators*. Math. Z. **158** (1978), 15–21.
- [7] Brüning, J., Gromes, D. *Über die Länge der Knotenlinien schwingender Membranen*. Math. Z. **124** (1972), 79–82.
- [8] Cann, J. *Counting Nodal Components of Boundary-adapted Arithmetic Random Waves*. PhD thesis submitted at King's College London (2019).
- [9] Cheng, S. Y. *Eigenfunctions and nodal sets*. Comm. Math. Helv. **51** (1976), 43–55.
- [10] Cilleruelo, J. *The distribution of the lattice points on circles*, Journal of Number Theory **43**, no. 2 (1993), 198–202.
- [11] Donnelly, H., Fefferman, C., *Nodal sets of eigenfunctions on Riemannian manifolds*. Inventiones Mathematicae **93** (1988), 161–183.
- [12] Gnuzmann, S., Lois, S. *Remarks on nodal volume statistics for regular and chaotic wave functions in various dimensions*. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **372** (2014), <http://doi.org/10.1098/rsta.2012.0521>
- [13] Krishnapur, M., Kurlberg, P., Wigman, I. *Nodal length fluctuations for arithmetic random waves*. Ann. Math. **177** no. 2 (2013), 699–737.
- [14] Kurlberg, P., Wigman, I. *On probability measures arising from lattice points on circles*. Math. Ann. **367**, no. 3-4 (2017), 1057–1098.
- [15] Kratz, M. F., León, J. R. *Central limit theorems for level functionals of stationary Gaussian processes and fields*. J. Theoret. Probab. **14** no. 3 (2001), 639–672.
- [16] Landau, E. *Über die Einteilung der positiven Zahlen nach vier Klassen nach der Mindestzahl der zu ihrer Addition Zusammensetzung erforderlichen Quadrate*. Arch. Math, und Phys. III, (1908).
- [17] Logunov, A., Malinnikova, E. *Nodal sets of Laplace eigenfunctions: estimates of the Hausdorff measure in dimensions two and three*. 50 Years with Hardy spaces, Oper. Theory Adv. Appl. **261** (2018), 333–344.
- [18] Logunov, A. *Nodal sets of Laplace eigenfunctions: proof of Nadirashvili's conjecture and of the lower bound in Yau's conjecture*. Ann. of Math. (2) **187** no. 1 (2018), 241–262.
- [19] Logunov, A. *Nodal sets of Laplace eigenfunctions: polynomial upper estimates of the Hausdorff measure*. Ann. of Math. (2) **187** no. 1 (2018), 221–239.
- [20] Longuet-Higgins, M. S. *The statistical analysis of a random, moving surface*. Philos. Trans. Roy. Soc. London Ser. A. **249** (1957), 321–387.

- [21] Longuet-Higgins, M. S. *Statistical properties of an isotropic random surface*. Philos. Trans. Roy. Soc. London. Ser. A. **250** (1957), 157–174.
- [22] Marinucci, D., Peccati, G., Rossi, M., Wigman, I. *Non-universality of nodal length distribution for arithmetic random waves*. Geom. and Func. Analysis. (GAFA) **26**, no. 3 (2016), 926–960.
- [23] Rudnick, Z., Wigman, I. *On the volume of nodal sets for eigenfunctions of the Laplacian on the torus*. Ann. Henri Poincaré **9**, no. 1 (2008), 109–130.
- [24] Rudnick, Z., Wigman, I. *Nodal intersections for random eigenfunctions on the torus*. Amer. J. Math. **138** no. 6 (2016), 1605–1644.
- [25] Royen, T. *A simple proof of the Gaussian correlation conjecture extended to multivariate gamma distributions*. arXiv preprint arXiv:1408.1028 (2014).
- [26] Wigman, I. *Fluctuations of the nodal length of random spherical harmonics*. Communications in Mathematical Physics, **298** (3) (2010), 787–831. Erratum published Comm. Math. Phys. **309** no. 1, (2012), 293–294.
- [27] Yau, S.T. *Survey on partial differential equations in differential geometry*. Seminar on Differential Geometry, Annals of Mathematics Studies, **102** (1982), 3–71 Princeton University Press, Princeton, N.J.