Gaussian Approximation of the Distribution of Strongly Repelling Particles on the Unit Circle

Alexander Soshnikov∗, Yuanyuan Xu†

November 1, 2017

Abstract

In this paper, we consider a strongly-repelling model of \( n \) ordered particles \( \{e^{i\theta_j}\}_{j=0}^{n-1} \) with the density

\[
p(\theta_0, \cdots, \theta_{n-1}) = \frac{1}{Z_n} \exp \left\{ -\frac{\beta}{2} \sum_{j \neq k} \sin^{-2} \left( \frac{\theta_j - \theta_k}{2} \right) \right\}, \quad \beta > 0.
\]

Let \( \theta_j = \frac{2\pi j}{n} + \frac{x_j}{n^2} + \text{const} \) such that \( \sum_{j=0}^{n-1} x_j = 0 \). Define \( \zeta_n \left( \frac{2\pi j}{n} \right) = \frac{x_j}{\sqrt{n}} \) and extend \( \zeta_n \) piecewise linearly to \([0, 2\pi]\). We prove the functional convergence of \( \zeta_n(t) \) to

\[
\zeta(t) = \sqrt{\frac{2}{\beta}} \Re \left( \sum_{k=1}^{\infty} \frac{1}{k} e^{ikt} Z_k \right), \quad \text{where } Z_k \text{ are i.i.d. complex standard Gaussian random variables.}
\]

1 Introduction

The study of random matrix theory (RMT) can be traced back to sample covariance matrices introduced by J. Wishart in data analysis in 1920s-1930s. In 1951, E. Wigner associated the energy levels of heavy-nuclei atoms with Hermitian matrices whose components are i.i.d. random variables. In 1960s, F. Dyson and M. Mehta identified three types of matrix ensembles: Gaussian Orthogonal Ensemble (GOE), Gaussian Unitary Ensemble (GUE), and Gaussian Symplectic Ensemble (GSE). In particular, the GUE/GOE/GSE is defined to be the ensemble of \( n \times n \) Hermitian/ Real Symmetric/Hermitian Quaternionic matrices equipped with a probability measure given by

\[
p_{\beta}^{\text{g}}(H) = \frac{1}{Z_n} e^{-\text{Tr}(H^2)} dH,
\]

where \( dH \) is the Lebesgue measure on the appropriate space of matrices.

The joint probability density for the eigenvalues \( \lambda_1, \cdots, \lambda_n \) of GUE/GOE/GSE is given by

\[
\frac{1}{Z_{\beta,n}} \prod_{i=1}^{n} e^{-\frac{\beta}{n} \lambda_i^2} \prod_{j<k} |\lambda_j - \lambda_k|^\beta,
\]

∗Department of Mathematics, University of California, Davis; soshniko@math.ucdavis.edu
†Department of Mathematics, University of California, Davis; yyxu@math.ucdavis.edu
where $\beta = 1$ for GOE, $\beta = 2$ for GUE, and $\beta = 4$ for GSE. The generalized ensemble for $\beta > 0$ is called Gaussian $\beta$-ensemble ([9]). They also introduced equally important circular ensembles, namely Circular Orthogonal Ensemble (COE), Circular Unitary Ensemble (CUE), and Circular Symplectic Ensemble (CSE). CUE is defined to be the ensemble of $n \times n$ unitary matrices equipped with the Haar measure. The ordered eigenvalues are denoted as $\{e^{i\theta_j}\}_{j=0}^{n-1}$, where $0 \leq \theta_0 \leq \cdots \leq \theta_{n-1} \leq 2\pi$.

The joint probability density of the angles $\{\theta_j\}_{j=0}^{n-1}$ is given by

$$p^c_\beta(\theta_1, \cdots, \theta_n) = \frac{1}{Z_{n,\beta}} \prod_{0 \leq j < k \leq n-1} |e^{i\theta_j} - e^{i\theta_k}|^\beta,$$

$$= \frac{1}{Z_{n,\beta}} \exp \left\{ \frac{\beta}{2} \sum_{j \neq k} \log \left| 2 \sin \frac{\theta_j - \theta_k}{2} \right| \right\},$$

with $\beta = 2$ and

$$Z_{n,\beta} = \frac{(2\pi)^n}{n!} \frac{\Gamma \left( \frac{3n}{2} + 1 \right)}{\Gamma \left( \frac{\beta n}{2} + 1 \right)}.$$

Similarly, the joint probability density for COE/CSE is given by (2)-(4) with $\beta = 1$ for COE and $\beta = 4$ for CSE. The generalized ensemble for $\beta > 0$ is named Circular $\beta$-ensemble ([11]).

These matrix ensembles were originally introduced in Physics, but recently have played an important role in linking RMT with Number Theory.

In 1859, B. Riemann conjectured that besides the negative even integers, all other non-trivial zeros of the Riemann zeta function have the form of $\frac{1}{2} + i\gamma_j$, where $\gamma_j \in \mathbb{R}$. Based on Riemann’s hypothesis, in 1970s, H. Montgomery proved (see [19]) that under technical conditions, the two point correlation function of $\gamma_j$’s on the scale of their mean spacing is

$$R_2(x) = 1 - \frac{\sin^2(\pi x)}{\pi^2 x^2}.$$  

Later, A. Odlyzko provided numerical support for Montgomery’s results in [20], and Z. Rudnick and P. Sarnak extended the results of [19] to higher order correlations in [21]. F. Dyson pointed out that (3) is the same as the two point correlation function of the scaled eigenvalues of random Hermitian matrices (GUE) and scaled eigenphases of random unitary matrices(CUE) as $n \to \infty$. In general, the limiting distribution of the non-trivial zeros of L-functions is believed to be related to the eigenvalue statistics of CUE/GUE model. Since then, many efforts were made to find the deep connection between zeta function, prime numbers, and random matrices, see e.g. [3]. In 2000, J. Keating and N. Snaith made an important contribution in applying Random Matrix approach in Number Theory, see [16]. They showed that the distribution of values taken by the characteristic polynomials

$$Z(U_n, s) = \det (I - U_n e^{-is}) \quad U_n \in \text{CUE}$$

is a good approximation to that of the Riemann zeta function.

In 1988, K. Johansson [13] proved the Central Limit Theorem(CLT) for the linear statistics in the Circular $\beta$-ensemble.
Theorem 1. Let $f \in C^{1+\epsilon}(S^1)$, $\epsilon > 0$. Then \( \sum_{j=0}^{n-1} f(\theta_j) - \frac{n}{2\pi} \int_0^{2\pi} f(x) dx \) converges in law to the Gaussian distribution $N(0, \frac{2}{\beta} \sum_{k=-\infty}^{\infty} |k|c_k^2)$, where $c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x)e^{-ikx} dx$.

Remark 2. The proof in [13] does not rely on the determinantal form of (3) for $\beta = 2$. The approach could be applied to all Circular $\beta$-ensembles. For $\beta = 2$, the result holds under the optimal condition $\sum_{k=-\infty}^{\infty} |k|c_k^2 < \infty$ (see also [7], [23] and references therein).

It was proved in [12] that for the CUE ($\beta = 2$), $\sqrt{2}\log|\det(e^{i\varphi} - U_n)|$ converges in distribution to a generalized random function

$$T(s) = \frac{1}{\sqrt{2}} \Re \left( \sum_{k=1}^{\infty} e^{iks}/\sqrt{k} Z_k \right),$$

where $Z_k$ are i.i.d complex standard Gaussian variables (see also [2] and [8]).

The generalized random function $T(s)$ makes another appearance in the Circular $\beta$-ensemble ($\beta = 2$) as follows. One can show that the joint probability density of (3) obtains its maximum at the lattice configuration $\theta_j = \frac{2\pi j}{n} + \text{const} \, (0 \leq j \leq n - 1)$. Write

$$\theta_j = \frac{2\pi j}{n} + t_j + \text{const}.$$ 

Let us choose the constant properly such that $\sum_{j=0}^{n-1} t_j = 0$ and then take Taylor expansion of (3) at this critical configuration. If we ignore the cubic and higher terms, then we get, as an approximation, a multivariate Gaussian distribution on the hyperplane $\sum_{j=0}^{n-1} t_j = 0$

$$\tilde{p}_g(t) = \frac{1}{Z_g} \exp \left\{ -\frac{\beta}{16} \sum_{j \neq k} \sin^2 \left( \frac{\pi(j-k)}{n} \right) \frac{(t_j - t_k)^2}{n^2} \right\}.$$ 

It can be shown that $t_j$ can be expressed as

$$t_j = \frac{2}{\sqrt{\beta}} \Re \left( \sum_{k=1}^{n} \frac{e^{2\pi i j k}/\sqrt{k}}{Z_k} \right) + \epsilon_n,$$

where $\epsilon_n$ is a negligible random error term with $\text{Var}(\epsilon_n) = o_n(1)$. Moreover, the linear statistics $\sum_{j=0}^{n-1} f \left( \frac{2\pi j}{n} + \frac{t_j}{n} \right)$ satisfies the same CLT as in Theorem 1.

This indicates that the generalized random function $T(s)$ defined in (6) gives a good approximation of the eigenvalue statistics of CUE. However, it is not entirely clear in what sense we can ignore cubic and higher order terms of the Taylor expansion of (3). This motivated us to consider a new model of interacting particles on the unit circle with stronger repulsion than that in the Circular $\beta$-ensembles. The purpose of this paper is to establish the Gaussian approximation for the distribution of strongly repelling particles on the unit circle.

Throughout this paper, the letters $C_k, C'_k, c_k$ and $c'_k$ ($k \in \mathbb{N}$) denote positive constants whose values might change in different parts of the paper, but are always independent of $n$. We say $X \ll Y$ if $\frac{X}{Y}\to 0$ as $n\to\infty$ and $X \sim Y$ if there exist positive constants $c$ and $C$ such that $c Y \leq X \leq CY$. Also, we denote $X \lesssim Y$ if $X \ll Y$ or $X \sim Y$. 

3
2 Set up and notations

Consider a strong repulsion model of particles distributed on
\[ \mathbb{T}^n / S_n = \{ \theta = (\theta_0, \cdots, \theta_{n-1}) \in [0, 2\pi]^n \mid \theta_0 \leq \theta_1 \leq \cdots \leq \theta_{n-1} \} . \]

The joint probability density is defined as
\[ q(\theta) = \frac{1}{Z_n} e^{H_{n,\beta}(\theta)} , \]
where
\[ H_{n,\beta}(\theta) = -\frac{\beta}{2} \sum_{i \neq j} \frac{1}{\sin^2 \left( \frac{\theta_i - \theta_j}{2} \right)} , \]
and
\[ Z_n = \int_{\mathbb{T}^n / S_n} e^{H_{n,\beta}(\theta)} d\theta . \]

For any measurable subset \( A \) of \( \mathbb{T}^n / S_n \), let \( P(A) = \int_A q(\theta) d\theta \).

Note that the repulsion in \( H_{n,\beta}(\theta) \) is stronger than the logarithmic one in (3).

Let
\[ \theta_i = \frac{2\pi i}{n} + \psi + \frac{x_i}{n^2} , \]
where \( \psi \) is a constant chosen so that
\[ \sum_{i=0}^{n-1} x_i = 0 . \]

Define \( \alpha_i = \frac{2\pi i}{n} + \psi, \alpha = (\alpha_0, \cdots, \alpha_{n-1}) \), \( x = (x_0, \cdots, x_{n-1}) \), then \( \theta = \alpha + \frac{x}{n^2} \).

Next, we introduce some useful lemmas.

**Lemma 3.** \( q(\theta) \) obtains its maximum at \( \theta = \alpha \), and
\[ H_{n,\beta}(\alpha) - H_{n,\beta}(\theta) = \frac{\beta}{2} \sum_{i \neq j} \frac{(x_i - x_j)^2}{n^4} \int_0^1 \frac{1}{2} + \cos^2 \left( \frac{\pi(i-j)}{n} + \tau \frac{x_i - x_j}{2n^2} \right) \frac{\sin^4 \left( \frac{\pi(i-j)}{n} + \tau \frac{x_i - x_j}{2n^2} \right)}{\sin^4 \left( \frac{\pi(i-j)}{n} \right)} \cdot (1 - \tau) d\tau . \]

**Lemma 4.**
\[ \sum_{k=1}^{n-1} \frac{1}{\sin^2 \left( \frac{\pi k}{n} \right)} = \frac{n^2 - 1}{3} . \]
\[ \sum_{k=1}^{n-1} \frac{1}{\sin^4 \left( \frac{\pi k}{n} \right)} = \frac{(n^2 - 1)(n^2 + 11)}{45} . \]
\[ \sum_{k=1}^{n-1} \frac{\sin^2 \left( \frac{m \pi k}{n} \right)}{\sin^2 \left( \frac{\pi k}{n} \right)} = m(n - m) \quad (1 \leq m \leq n - 1) . \]
\[ \sum_{k=1}^{n-1} \frac{\sin^2 \left( \frac{m \pi k}{n} \right)}{\sin^4 \left( \frac{\pi k}{n} \right)} = \frac{m^2(n - m)^2}{3} + \frac{2}{3} m(n - m) \quad (1 \leq m \leq n - 1) . \]
The proof of Lemma 4 is shown in the Appendix. By (15) in Lemma 4, we have,

\[
H_{n,\beta}(\alpha) = -\frac{\beta}{2} \sum_{i \neq j} \frac{1}{\sin^2 \left( \frac{\pi(i-j)}{n} \right)} = \frac{(n^3 - n)\beta}{6}. \tag{19}
\]

Next lemma shows that typically \(H_{n,\beta}(\theta)\) is not far from \(H_{n,\beta}(\alpha)\).

**Lemma 5.** For any \(C > 1\), define

\[
\Theta = \{ \theta \in \mathbb{T}^n/S_n \mid H_{n,\beta}(\alpha) - H_{n,\beta}(\theta) \leq Cn \log n \}.
\tag{20}
\]

Then there exists some \(c > 0\), such that

\[
P(\Theta) \geq 1 - n^{-c}.
\]

**Remark 6.** If we choose \(C = 1\), then the upper bound should be modified to \(n \log n - C' n\) for some \(C' > 0\).

Using Lemma 4 and Lemma 6, we have

**Lemma 7.** For any \(C > 1\), if \(\theta \in \Theta\), then there exists some constant \(C_0\) such that

\[
\sum_{i \neq j} \frac{(x_i - x_j)^2}{n^4 \sin^4 \left( \frac{\pi(i-j)}{n} \right)} \leq C_0 n \log^3 n. \tag{21}
\]

Moreover, for all \(0 \leq i \neq j \leq n - 1\),

\[
|x_i - x_j| \leq C_0 |i - j| \alpha n^{\frac{3}{2}} \log^2 n, \tag{22}
\]

where \(|i - j|_0 = \min \{|i - j|, n - |i - j|\}\).

Taking the Taylor expansion of the joint probability function \(q(\theta)\) around the critical configuration \(\theta = \alpha\), we have that for some \(\delta \in [0, 1]\),

\[
q(\theta) = \frac{1}{Z_n} e^{H_{n,\beta}(\alpha)} \exp \left\{ \frac{\beta}{4} \sum_{i \neq j} -\frac{3}{2} + \sin^2 \left( \frac{\pi(i-j)}{n} \right) (x_i - x_j)^2 \right\} \tag{24}
\]

\[
\times \exp \left\{ \frac{\beta}{12} \sum_{i \neq j} \frac{3 \cos \left( \frac{\pi(i-j)}{n} + \frac{\delta x_i - x_j}{2n^2} \right) - \cos \left( \frac{\pi(i-j)}{n} + \frac{\delta x_i - x_j}{2n^2} \right)}{n^6 \sin^5 \left( \frac{\pi(i-j)}{n} + \frac{\delta x_i - x_j}{2n^2} \right)} \right\} (x_i - x_j)^3 \right\}.
\]

Denote the quadratic term by

\[
G(x) := \frac{\beta}{4} \sum_{i \neq j} -\frac{3}{2} + \sin^2 \left( \frac{\pi(i-j)}{n} \right) (x_i - x_j)^2, \tag{25}
\]
and the cubic term as
\[
F(x) := \frac{\beta}{12} \left[ \sum_{i \neq j} \frac{3 \cos \left( \frac{\pi(i-j)}{n} + \delta \frac{x_i - x_j}{2n^2} \right)}{n^6 \sin^5 \left( \frac{\pi(i-j)}{n} + \delta \frac{x_i - x_j}{2n^2} \right)} - \frac{\cos \left( \frac{\pi(i-j)}{n} + \delta \frac{x_i - x_j}{2n^2} \right)}{n^6 \sin^3 \left( \frac{\pi(i-j)}{n} + \delta \frac{x_i - x_j}{2n^2} \right)} \right] (x_i - x_j)^3. \tag{26}
\]

Using (12), consider the change of variable \( \theta \to (x, \psi) \), where \( x \) is a degenerate vector on the hyperplane \( \Gamma \),
\[
\Gamma = \left\{ x \in \mathbb{R}^n : \sum_{i=0}^{n-1} x_i = 0 \right\}. \tag{27}
\]
Let
\[
f(x) = q(\theta(x, \phi)). \tag{28}
\]
Note that the joint probability density \( f \) only depends on \( x \). But the domain \( \Omega \) depends on both \( x \) and \( \psi \). If \( \theta \in \mathbb{T}_n / \mathbb{S}_n \), then
\[
(x, \psi) \in \Omega = \left\{ \Gamma \times [-\pi + \frac{\pi}{n}, \pi + \frac{\pi}{n}] : x_i - x_{i-1} \geq -2\pi n; -\frac{x_0}{n^2} \leq \psi \leq \frac{2\pi}{n} - \frac{x_{n-1}}{n^2} \right\}.
\]
Thus, the marginal density function for \( x \) is
\[
p(x) = \int_{-\frac{2\pi}{n} \leq \psi \leq \frac{2\pi}{n} - \frac{x_{n-1}}{n^2}} f(x) d\psi = \left( \frac{2\pi}{n} - \frac{x_{n-1} - x_0}{n^2} \right) f(x) \tag{29}
\]
where
\[
x \in \Lambda = \left\{ x \in \Gamma : x_i - x_{i-1} \geq -2\pi n; x_0 \geq -\pi n(n+1); x_{n-1} \leq \pi n(n+1); x_{n-1} - x_0 \leq 2\pi n \right\}. \tag{30}
\]
Otherwise, if \( x \in \Lambda^c \), \( p(x) = 0 \). For any measurable subset \( A \subset \Gamma \), denote
\[
\mathbb{P}_x(A) = \int_A p(x) dx. \tag{31}
\]
It follows from Lemma 5 and Lemma 7 that there exists a subset of \( \Lambda \), namely
\[
\Gamma_D = \left\{ x \in \Gamma : \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|} \leq Dn^{\frac{1}{4}} \log^\frac{3}{2} n; x_0 \geq -\pi n(n+1); x_{n-1} \leq \pi n(n+1); x_{n-1} - x_0 \leq 2\pi n \right\}, \tag{32}
\]
such that
\[
\mathbb{P}_x(\Gamma_D^c) \leq n^{-c_0}. \tag{33}
\]
In addition, if \( x \in \Gamma_D \), the probability density of \( x \) can be written as
\[
p(x) = \frac{1}{\tilde{Z}_n} e^{G(x) + F(x)} (1 + o_n(1)), \tag{34}
\]
where
\[
\tilde{Z}_n = \int_{\Gamma_D} e^{G(x) + F(x)} dx. \tag{35}
\]
Let us define a Gaussian distribution on the hyperplane $\Gamma$ by

$$p_g(x) = \frac{1}{Z_g} e^{G(x)}, \quad (36)$$

where

$$Z_g = \int_{\Gamma} e^{G(x)} dx. \quad (37)$$

For any measurable subset $A \subset \Gamma$, denote

$$\mathbb{P}_g(A) = \int_A p_g(x) dx. \quad (38)$$

**Remark 8.** Refining the argument used in the proof of Lemma 7, we can also prove that there exists a subset $\Gamma_{D'} \subset \Gamma$ such that $\mathbb{P}_x(\Gamma_{D'}^c) = o_n(1)$. If $x \in \Gamma_{D'}$, we have $x_j \ll n$ for all $0 \leq j \leq n - 1$, and thus $\phi \sim \frac{1}{n}$. In addition,

$$p(\psi|x) = \frac{f(x)}{\int_{-\frac{2\pi}{n}}^{\frac{2\pi}{n}} f(x) d\psi} = \frac{1}{2\pi n^2 - x_0^2} = \frac{n}{2\pi}(1 + o_n(1)).$$

One can show that $n\psi$ and $x$ are asymptotically independent from each other and $n\psi$ converges to the uniform distribution on $[0, 2\pi]$. However, we are not going to use this in the paper.

### 3 Main theorems

In this section, we formulate our main results. We start with an auxiliary proposition. Recall that we have defined $\mathbb{P}_x$ and $\mathbb{P}_g$ in (31) and (38).

**Proposition 9.** There exists a subset $\Gamma' \subset \Gamma$, such that

$$\mathbb{P}_x(\Gamma') = 1 - o_n(1), \quad \mathbb{P}_g(\Gamma') = 1 - o_n(1),$$

and

$$\sup_{x \in \Gamma'} F(x) = o_n(1).$$

Proposition 9 immediately implies that the total variation distance between $\mathbb{P}_x$ and $\mathbb{P}_g$ goes to zero as $n$ goes to infinity.

**Theorem 10.**

$$\sup_{A \subset \Gamma} |\mathbb{P}_x(A) - \mathbb{P}_g(A)| = o_n(1),$$

where the supremum at the LHS is taken over all measurable subsets $A \subset \Gamma$. 
Based on Theorem 10, we obtain the main theorem. For each fixed \( n \), we construct a random function in \( C[0, 2\pi] \), denoted as \( \zeta_n(t) \), by letting \( \zeta_n\left(\frac{2\pi j}{n}\right) = \frac{x_j}{\sqrt{n}} \) and then connecting these lattice points with straight segments. Define the limiting random function to be

\[
\zeta(t) = \sqrt{\frac{2}{\beta}} \Re \left( \sum_{k=1}^{\infty} \frac{1}{k} e^{ikt} Z_k \right) = \sqrt{\frac{2}{\beta}} \sum_{k=1}^{\infty} \left( \frac{\cos kt}{k} p_k - \frac{\sin kt}{k} q_k \right).
\]

where \( p_k, q_k \) are i.i.d. real standard Gaussian random variables, and \( Z_k \) are i.i.d. complex standard Gaussian random variables. Note that this is a well defined random function since the variance is bounded. It can be viewed as an analogue of (6) in the CUE case. Then we have:

**Theorem 11.** \( \zeta_n(t) \) converges to \( \zeta(t) \) in finite dimensional distribution. Furthermore, functional convergence takes place. In other words, \( \zeta_n(t) \) converges to \( \zeta(t) \) in distribution weakly on the space \( C[0, 2\pi] \).

Finally, we finish this section by formulating two corollaries.

**Corollary 12.** Consider periodic function \( g \) on \( S^1 \) with complex Fourier coefficients \( \{c_k\}_{k \geq 0} \), where \( c_k = \frac{1}{2\pi} \int_0^{2\pi} g(x) e^{ikx} dx \), such that \( \sum_{k=-\infty}^{\infty} |k|^{\frac{3}{2}}|c_k| < \infty \), then

\[
\mathbb{E} \exp \left( it \sqrt{n} \sum_{j=0}^{n-1} g(\theta_j) \right) = \exp \left( itn^{\frac{3}{2}} c_0 - \frac{t^2}{\beta} \sum_{k=1}^{\infty} |c_k|^2 \right) (1 + o_n(1)).
\]

In other words, \( \sqrt{n} \sum_{j=0}^{n-1} g(\theta_j) - n^{\frac{3}{2}} c_0 \) converges in distribution to \( N \left( 0, \frac{2}{\beta} \sum_{k=1}^{\infty} |c_k|^2 \right) \).

**Remark 13.** The condition \( \sum_{k=-\infty}^{\infty} |k|^{\frac{3}{2}}|c_k| < \infty \) in Corollary 12 is expected to be weakened to \( \sum_{k=-\infty}^{\infty} |c_k|^2 < \infty \).

**Corollary 14.** \( \max_{0 \leq i \leq n-1} \left| \frac{x_i}{\sqrt{n}} \right| \) converges in distribution to \( \sup_{t \in [0, 2\pi]} |\zeta(t)| \).

To simplify the notations, the proofs of these results are written for \( \beta = 2 \). The general case \( \beta > 0 \) is essentially identical.

## 4 Proofs of the Lemmas in Section 2

We start this section by proving Lemma 3.

**Proof.** (Lemma 3) Define

\[
\phi(\tau) := H_{n,2}(\alpha + \tau \frac{x}{n^2}).
\]

We compute its first and second derivative with respect to \( \tau \),

\[
\phi'(\tau) = \sum_{i \neq j} \frac{\cos \left( \frac{\pi(i-j)}{n} + \tau \frac{x_i - x_j}{2n^2} \right)}{\sin^2 \left( \frac{\pi(i-j)}{n} + \tau \frac{x_i - x_j}{2n^2} \right)} \cdot \frac{x_i - x_j}{n^2};
\]
\[ \phi''(\tau) = - \sum_{i \neq j} \frac{1}{2} \cos^2 \left( \frac{\pi(i-j)}{n} + \frac{\tau x_i - x_j}{2n^2} \right) \cdot \left( \frac{x_i - x_j}{n^2} \right)^2. \] (42)

Note that
\[
\phi'(0) = \sum_{i \neq j} \cos \left( \frac{\pi(i-j)}{n} \right) \cdot \frac{x_i - x_j}{n^2} = \frac{1}{n^2} \left( \sum_{i \neq j} \cos \left( \frac{\pi(i-j)}{n} \right) \cdot x_i - \sum_{i \neq j} \cos \left( \frac{\pi(i-j)}{n} \right) \cdot x_j \right)
\]
\[
= \frac{1}{n^2} \sum_i \left( \sum_{k=1}^{n-1} \cos \frac{2\pi k}{n} \right) x_i - \frac{1}{n^2} \sum_j \left( \sum_{l=1}^{n-1} \cos \frac{2\pi l}{n} \right) x_j = 0,
\]
and \( \phi''(0) \leq 0 \). Thus,
\[
H_{n,2}(\alpha) - H_{n,2}(\theta) = \phi(0) - \phi(1) = - \int_0^1 (1 - \tau) \phi''(\tau) d\tau
\]
\[
= \sum_{i \neq j} \left( \frac{x_i - x_j}{n^2} \right)^2 \int_0^1 \frac{1}{2} \cos^2 \left( \frac{\pi(i-j)}{n} + \frac{\tau x_i - x_j}{2n^2} \right) \cdot (1 - \tau) d\tau \geq 0.
\]
This implies that \( H_{n,2}(\theta) \) obtains its maximum at \( \theta = \alpha \).

Next, we turn our attention to proving Lemma 5.

**Proof.** (Lemma 5) It follows from the definition of \( \Theta \) in (20) and the trigonometric identity (19) that
\[
\mathbb{P}(\Theta^c) = \frac{1}{Z_n} \int_{\Theta^c} e^{H_{n,2}(\theta)} d\theta \leq \left( e^{\frac{3}{n} - Cn \log n} \right) \frac{1}{Z_n} \mu(\mathbb{T}^n / S_n) = \frac{(2\pi)^n}{n!} e^{\frac{3}{n} - Cn \log n} \frac{1}{Z_n},
\] (43)
where \( \mu \) denote the Lebesgue measure on \( \mathbb{R}^n \). Choose any \( 0 < C' < C \) and define a subset of \( \Theta \),
\[
\Theta' = \{ \theta \in \mathbb{T}^n / S_n \mid H_{n,2}(\alpha) - H_{n,2}(\theta) \leq C'n \log n \}.
\] (44)

Then similarly, we have
\[
Z_n = \int_{\mathbb{T}^n / S_n} e^{H_{n,2}(\theta)} d\theta \geq \int_{\Theta'} e^{H_{n,2}(\alpha) - C'n \log n} d\theta = e^{\frac{3}{n} - C'n \log n} \mu(\Theta').
\] (45)

Note that if each entry of \( x \) can be bounded by some constant \( M > 0 \), then
\[
H_{n,2}(\alpha) - H_{n,2}(\theta) \leq \sum_{i \neq j} \frac{M^2}{n^4 \sin^4 \left( \frac{\pi(i-j)}{n} \right)} \leq M^2 \sum_{i \neq j} \frac{1}{\pi^4 |i-j|^4} \leq C'n \log n.
\]
Therefore,
\[ \{ \theta \in \mathbb{T}^n / \mathcal{S}_n \mid \theta_i - \alpha_i \leq \frac{M}{n^2}, \ \forall \ 0 \leq i \leq n-1 \} \subset \Theta', \]
and thus the Lebesgue measure of the set \( \Theta' \) can be bounded from below as
\[ \mu(\Theta') \geq \left( \frac{M}{n^2} \right)^n = e^{n \log M - 2n \log n}. \] (46)

Therefore, by (15) and (46),
\[ Z_n \geq e^{\frac{n^3 - n}{4} - C'n \log n + n \log M - 2n \log n}. \] (47)

Combining it with (19) and (43), we obtain
\[ \mathbb{P}(\Theta^c) \leq \left( \frac{2\pi}{n!} \right)^n e^{-(C-C'n \log n - n \log M + 2n \log n) - C'' e^{-(C-C-1)n \log n} = o_n(1),} \] (48)
provided \( C > 1 \) and \( C' \) is sufficiently small.

Using Lemma 5, we finish this section by giving the proof of Lemma 7.

**Proof.** (Lemma 7)

By Lemma 5, if \( \theta \in \Theta \), then for some constant \( C > 1 \),
\[ \sum_{i \neq j} \frac{(x_i - x_j)^2}{n^4} \int_0^1 \frac{(1 - \tau) d\tau}{\sin^4 \left( \frac{(i-j)}{n} + \frac{x_i-x_j}{2n} \right)} \leq Cn \log n. \] (49)

Let \( I = \{(i, j) \mid 0 \leq i \neq j \leq n-1\} \), \( I_1 = \{(i, j) \in I \mid |x_i - x_j| < n \eta_n |i - j| \} \), and \( I_2 = I \setminus I_1 \). Let \( \eta_n \gg 1 \). Then by (49) we have
\[ Cn \log n \geq \sum_{I_1} \frac{(x_i - x_j)^2}{n^4} \int_0^1 \frac{(1 - \tau) d\tau}{\sin^4 \left( \frac{(i-j)}{n} + \frac{x_i-x_j}{2n} \right)} \geq \sum_{I_1} \frac{(x_i - x_j)^2}{n^4} \int_0^1 \frac{(1 - \tau) d\tau}{\frac{|i-j|_o}{n} + \frac{|x_i-x_j|}{2n^2}}. \]

Thus,
\[ \sum_{I_1} \frac{(x_i - x_j)^2}{n^4 |i-j|_o^4} \leq C'' n \eta_n^4 \log n. \] (50)

Next, it can be shown that \( I_2 = \emptyset \). Note that
\[ \int_0^1 \frac{(1 - \tau) d\tau}{\sin^4 \left( \frac{(i-j)}{n} + \frac{x_i-x_j}{2n^2} \right)} \geq \int_0^1 \frac{(1 - \tau) d\tau}{\left( \frac{(i-j)}{n} + \frac{x_i-x_j}{2n^2} \right)^4} = \frac{3\pi(i-j)}{n} + \frac{x_i-x_j}{n^2}. \] (51)

If \( (i, j) \in I_2 \), then
\[ \frac{|x_i - x_j|}{n^2} \geq \eta_n |i - j|_o. \]
and
\[
\text{RHS of (51)} \geq \text{const} \frac{\eta_n |i-j|_o}{n^3} \left( \frac{|x_i-x_j|}{n^2} \right)^2 \geq \text{const} \frac{\eta_n n^6}{(|i-j|_o)^2 (|x_i-x_j|)^2}. \tag{52}
\]

Note that by the triangle inequality, if \((i, j) \in I_2\), then there exists at least \(|i-j|\) index pairs belonging to \(I_2\), in the form of \((i, k)\) or \((k, j)\) where \(k\) is between \(i\) and \(j\). Thus by (49) and (52),
\[
Cn \log n \geq C' \sum_{i \neq j} \frac{\eta_n n^2}{|i-j|_o} \geq C' n^2 \eta_n \frac{1}{|i-j|_o}.
\]

This implies that \(|i-j|_o \geq C'' n \eta_n \log^{-1} n\), and thus \(|x_i-x_j| \geq C'' n^2 \eta_n^2 \log^{-1} n\). For some sufficient large \(M > 0\), choosing \(\eta_n = M \log^{\frac{1}{2}} n\), we obtain
\[
|x_i-x_j| \geq C'' M^2 n^2.
\]

With a sufficient large \(M\), the last inequality contradicts that \(|x_i-x_j| \leq 2\pi n^2\). Therefore \(I_2 = \emptyset\) and there exists some constant \(C_0\) such that
\[
\sum_{i} \frac{(x_i-x_j)^2}{|i-j|_o^4} \leq C_0 n \log^3 n. \tag{53}
\]

Furthermore, denote \(x_{n+i} = x_i, i \geq 0\), then for \(0 \leq j \leq n-1\),
\[
(x_{j+1}-x_j)^2 \leq C_0 n \log^3 n.
\]

Therefore, by the triangle inequality, we have
\[
|x_i-x_j| \leq C_0 |i-j|_o n^\frac{1}{2} \log^{\frac{3}{2}} n.
\]

Finally, since
\[
\frac{2}{\pi} \leq \frac{\sin x}{x} \leq 1, \quad 0 < x \leq \frac{\pi}{2},
\]

we have
\[
\sum_{i \neq j} \frac{(x_i-x_j)^2}{n^4 \sin^4 \left( \frac{\pi(i-j)}{n} \right)} \leq C'_0 n \log^3 n. \tag{54}
\]

\[\square\]

5 Estimate of the multivariate Gaussian distribution

Note that \(G(x)\) defined in (25) can be written in the quadratic form of \(-\frac{1}{2} x^T A x\), where
\[
A_{i,j} = -\frac{3}{n^4 \sin^4 \left( \frac{\pi(i-j)}{n} \right)} + \frac{2}{n^4 \sin^2 \left( \frac{\pi(i-j)}{n} \right)} \quad (i \neq j), \tag{55}
\]
and, by (15) and (16) in Lemma 4,
\[ A_{ii} = \left( n^2 - 1 \right) \left( n^2 + 11 \right) \frac{15}{15n^4} - \frac{2n^2 - 2}{3n^4} = -\sum_{j \neq i} A_{ij}. \] (56)

Since \( A \) is not invertible, \( p_g \) defined in (36) can be viewed as a degenerate Gaussian distribution on the hyperplane \( \Gamma \) defined in (27),
\[ p_g(x) = \frac{1}{Z_g} e^{-\frac{1}{2} x^T A x}, \]
where \( Z_g = \int_{\Gamma} e^{-\frac{1}{2} x^T A x} dx. \)

Next, we aim to explore the covariance structure of this Gaussian distribution. Note that \( A \) is a circular matrix generated by the vector \((A_{0,0}, A_{0,1}, \cdots, A_{0,n-1})\). Therefore its normalized eigenvectors can be chosen as
\[ v_k = \frac{1}{\sqrt{n}} (\omega_0^k, \omega_1^k, \cdots, \omega_{n-1}^k), \quad k = 0, 1, \cdots, n - 1. \] (57)

where \( \omega_k = e^{\frac{2\pi ik}{n}} \). By using (17) and (18) in Lemma 4 the corresponding eigenvalues are given by
\[ \lambda_k = A_{0,0} + A_{0,1} \omega_k + A_{0,2} \omega_k^2 + \cdots + A_{0,n-1} \omega_k^{n-1} = \left( n^2 - 1 \right) \left( n^2 + 11 \right) \frac{15}{15n^4} - \frac{2n^2 - 2}{3n^4} - \frac{3}{n^4} \sum_{j=1}^{n-1} \cos \frac{2\pi kj}{n} \sin^4 \frac{\pi j}{n} + \frac{2}{n^4} \sum_{j=1}^{n-1} \cos \frac{2\pi kj}{n} \sin^2 \frac{\pi j}{n} \]
\[ = \frac{6}{n^4} \sum_{j=1}^{n-1} \sin^2 \frac{\pi j}{n} \sin^4 \frac{\pi j}{n} - \frac{4}{n^4} \sum_{j=1}^{n-1} \sin^2 \frac{\pi j}{n} \sin^2 \frac{\pi j}{n} = \frac{2k^2 (n-k)^2}{n^4}, \quad 0 \leq k \leq n - 1. \] (58)

Define
\[ U = (v_0^T, v_1^T, \cdots, v_{n-1}^T), \] (59)
then we have \( U^*U = 1 \) and \( A = \Lambda U U^* \), where \( \Lambda \) is the diagonal matrix generated by \((\lambda_0, \lambda_1, \cdots, \lambda_{n-1})\). Let \( s = U^*x \). Then for \( 1 \leq k, j \leq n - 1 \),
\[ \mathbb{E} s_k s_k = \frac{1}{\lambda_k} = \frac{n^4}{2k^2(n-k)^2} \quad \text{and} \quad \mathbb{E} s_k s_j = 0 \quad (k \neq j). \] (60)

Note that \( s_0 = 0 \) because of the definition of \( \Gamma \) in (27). We also note that
\[ s_j = \overline{s}_{n-j} \quad \text{if} \quad j > \frac{n-1}{2}. \] (61)

For simplicity, we assume that \( n \) is odd. (the even case can be treated in a similar way). Then \((s_1, \cdots, s_{\frac{n-1}{2}})\) are \( \frac{n-1}{2} \) independent complex Gaussian random variables. In particular, we can write
\[ s_k = \frac{n^2}{2k(n-k)} p_k + i \frac{n^2}{2k(n-k)} q_k, \quad \text{for} \quad 1 \leq k \leq \frac{n-1}{2}. \] (62)
where \( \{p_k\} \) and \( \{q_k\} \) are independent real standard Gaussian variables.

Since \( x = Us \), we can compute the covariance structure for \( x \),

\[
\mathbb{E} x_k x_j = \frac{1}{n} \sum_{m=1}^{n-1} \frac{1}{\lambda_m} e^{2\pi i m(k-j)} = \frac{1}{2n} \sum_{m=1}^{n-1} e^{2\pi i m(k-j)} \frac{n^4}{m^2(n-m)^2}.
\]

(63)

In particular, \( \text{Var}(x_k) \sim n \). In addition,

\[
x_j = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} e^{2\pi i j k/n} s_k = \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} e^{2\pi i j k/n} s_k + \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} e^{-2\pi i j k/n} s_k = \frac{2}{\sqrt{n}} \sum_{k=1}^{n-1} \Re(e^{2\pi i j k/n} s_k)
\]

\[
= \frac{2}{\sqrt{n}} \sum_{k=1}^{n-1} \left( \cos\left(\frac{2\pi j k}{n}\right) \frac{n^2}{2k(n-k)} p_k - \sin\left(\frac{2\pi j k}{n}\right) \frac{n^2}{2k(n-k)} q_k \right)
\]

(64)

Let

\[
\xi_j^{(l)} = x_{j+l} - x_j, \quad \text{where} \quad x_{n+i} = x_i, \quad i \geq 0
\]

(65)

It is also useful for us to compute the covariance function between \( \xi_j^{(l)} \) and \( \xi_k^{(l)} \).

**Proposition 15.** There exists some universal constant \( C \) such that

\[
|\mathbb{E} \xi_k^{(l)} \xi_j^{(l)}| \leq C \min \left\{ l, \frac{l^2}{|k-j|} \right\}.
\]

(66)

In particular,

\[
\text{Var}(\xi_k^{(l)}) \leq Cl.
\]

(67)

**Proof.** By (63), we have

\[
\mathbb{E} \xi_k^{(l)} \xi_j^{(l)} = \mathbb{E}(x_{k+l} - x_k)(\bar{x}_{j+l} - \bar{x}_j) = \mathbb{E}x_{k+l}\bar{x}_j + \mathbb{E}x_k\bar{x}_{j+l} - \mathbb{E}x_k\bar{x}_j - \mathbb{E}x_{k+l}\bar{x}_{j+l}
\]

\[
= \frac{1}{n} \sum_{m=1}^{n-1} \frac{1}{\lambda_m} e^{2\pi i m(k-j)} \left( 2 - e^{2\pi i m l/n} - e^{-2\pi i m l/n} \right) = \frac{2}{n} \sum_{m=1}^{n-1} \frac{\sin^2(\pi ml/n)}{(m/n)^2 \left(1 - (m/n)^2\right)^2} e^{2\pi i (k-j) m/n}.
\]

(68)

For the RHS of (68), we can find an upper bound as

\[
|\mathbb{E} \xi_k^{(l)} \xi_j^{(l)}| \leq \frac{2}{n} \sum_{m=1}^{n-1} \frac{\sin^2(\pi ml/n)}{(m/n)^2 \left(1 - (m/n)^2\right)^2} \leq \frac{2}{n} \sum_{m=1}^{n-1} \frac{\sin^2(\pi ml/n)}{(m/n)^2 \left(1 - (m/n)^2\right)^2} \leq \frac{16}{3n} \sum_{m=1}^{n-1} \frac{\sin^2(\pi ml/n)}{(m/n)^2 \left(1 - (m/n)^2\right)^2}.
\]

(69)

The RHS of (69) can be viewed as a Riemann sum of the function \( \frac{\sin^2 \pi x}{x^2} \) corresponding to the evenly-spaced partition over \( [0, \frac{1}{2}] \) with the subintervals of length \( \frac{1}{n} \). Since the function \( \frac{\sin^2 \pi x}{x^2} \) can be bounded from above by a monotonic function \( m(x) \) defined as

\[
m(x) = \begin{cases} 
1 & \text{if } 0 \leq x \leq 1 \\
\frac{1}{x^2} & \text{if } x > 1,
\end{cases}
\]

13
the RHS of (69) can be bounded by a Riemann sum of \( m(x) \). Note that \( m(x) \) is monotonic, the error between its upper and lower Riemann sum is at most of the order \( \frac{l}{n} \). Then there exists a universal constant \( C > 0 \), such that

\[
|E_{\xi_k}^{(l)} \xi_j^{(l)}| \leq \frac{16}{3} l \int_0^l m(x)dx + O \left( \frac{l}{n} \right) \leq \frac{16l}{3} \int_0^\infty p(x)dx(1 + o(1)) \leq Cl.
\]

For large \(|k - j|\), the heavy oscillation of the exponential term leads to cancellations between terms in the expression (68) for \( E_{\xi_k}^{(l)} \xi_j^{(l)} \). Thus the upper bound that we have obtained above is not sharp in this case. Let \( a_m = \frac{\sin^2 \left( \frac{\pi ml}{n} \right)}{(\frac{m}{n}) \left( 1 - (\frac{m}{n}) \right)^2} \), and \( b_m = e^{2\pi i (k-j) \frac{m}{n}} \). By summation by parts, we have

\[
\sum_{m=1}^{n-1} a_m b_m = \sum_{m=1}^{n-1} (a_m - a_{m+1}) \left( \sum_{p=1}^{m} b_p \right) + a_n \sum_{p=1}^{n-1} b_p.
\]

Then

\[
|E_{\xi_k}^{(l)} \xi_j^{(l)}| \leq 2 \sum_{m=1}^{n-1} \frac{1}{n} \left| a_m - a_{m+1} \right| \left| \frac{1 - e^{2\pi i (k-j) \frac{m}{n}}}{1 - e^{2\pi i (k-j) \frac{n}{n}}} \right|.
\]

By differentiating the function \( f(x) = \frac{\sin^2 \pi x}{x^2(1-x)^2} \), we find that the derivative is at most of the order \( l^2 \) and we have

\[
|a_m - a_{m+1}| = \left| f \left( \frac{m}{n} \right) - f \left( \frac{m+1}{n} \right) \right| \leq \frac{C' l^2}{n}.
\]

Using the inequality

\[
\left| \frac{1 - e^{2\pi i (k-j) \frac{m}{n}}}{1 - e^{2\pi i (k-j) \frac{n}{n}}} \right| \leq \left| \frac{2}{1 - e^{2\pi i (k-j) \frac{n}{n}}} \right| \leq \frac{C' n}{|k - j|},
\]

we have

\[
|E_{\xi_k}^{(l)} \xi_j^{(l)}| \leq \frac{C' l^2}{|k - j|}.
\]

This finishes the proof of the Proposition 15.

Combining (32) with (67) in Proposition 15 and (63), we have

**Lemma 16.** There exist some constants \( c_1, c_2 \), such that

\[
\mathbb{P}_g(\Gamma_D) \leq e^{-c_1 n^2} + n^2 e^{-c_2 n \log^3 n}.
\]
6 Proof of Proposition 9 and Theorem 10

In this section, we prove Proposition 9 and Theorem 10. We start with some preliminary details. If \( x = (x_0, \ldots, x_{n-1}) \in \Gamma_D \), then by Lemma 7,

\[
G(x) = \frac{1}{2} \sum_{i \neq j} \frac{-3}{2} + \sin^2 \left( \frac{\pi(i-j)}{n} \right) (x_i - x_j)^2 \lesssim -\sum_{i>j} \frac{(x_i - x_j)^2}{|i-j|_o} \sim -n \log^3 n,
\]

and

\[
\frac{|x_i - x_j|}{|i-j|_o} \leq n^{\frac{1}{2}} \log^\frac{3}{2} n. \tag{71}
\]

Then \( \frac{(x_i-x_j)}{2n^2} \lesssim \frac{\log^\frac{3}{2} n}{n^2} \) is negligible, and thus

\[
F(x) = \frac{1}{6} \sum_{i \neq j} \frac{3 \cos \left( \frac{\pi(i-j)}{n} + \delta \frac{x_i-x_j}{2n^2} \right)}{n^6 \sin^5 \left( \frac{\pi(i-j)}{n} \right)} - \frac{\delta \frac{x_i-x_j}{2n^2}}{n^6 \sin^3 \left( \frac{\pi(i-j)}{n} \right)} (x_i - x_j)^3
\]

\[
\sim \sum_{i \neq j} \left[ \frac{(x_i-x_j)^3}{2n^6 \sin^5 \left( \frac{\pi(i-j)}{n} \right)} + \frac{(x_i-x_j)^3}{6n^6 \sin^3 \left( \frac{\pi(i-j)}{n} \right)} \right] = \sum_{i>j} \frac{(x_i-x_j)^3}{n|i-j|_o^5}. \tag{72}
\]

Comparing \( F(x) \) with \( G(x) \) and by (71), we have

\[
|F(x)| = |G(x)|O \left( \max_{i \neq j} \frac{|x_i-x_j|}{n|i-j|_o} \right) = |G(x)|O \left( \frac{\log^\frac{3}{2} n}{n^\frac{1}{2}} \right) \lesssim n^{\frac{1}{2}} \log^\frac{3}{2} n. \tag{73}
\]

In this section, we want to show that \( F(x) = o_n(1) \) with high probability. To be more specific, by (72), we want to show that

\[
F(x) \sim \sum_{i=1}^{n-1} \frac{1}{n} \sum_{j=0}^{n-l-1} \xi_j^{(l)}(x)^3 = o_n(1), \tag{74}
\]

where \( \xi_j^{(l)} \) is defined in (55). We divide the proof into three parts. The first step is to show that the normalized constant \( \tilde{Z}_n \) defined in (35) is not far from \( Z_g \) defined in (37). This implies that the probability distribution of \( x \) is not far from the Gaussian distribution \( p_g(x) \). The second step is to show that under the Gaussian distribution, \( F(x) \) decays to zero polynomially fast as \( n \to \infty \) with high probability. The last step is to combine the first two steps and obtain that \( F(x) = o_n(1) \) with high probability of \( p(x) \).

6.1 Step 1: Comparing \( Z_g \) and \( \tilde{Z}_n \)

For reader’s convenience, recall the definition of the normalized constants \( \tilde{Z}_n \) and \( Z_g \).

\[
\tilde{Z}_n = \int_{\Gamma_D} e^{G(x)+F(x)} dx; \quad Z_g = \int e^{G(x)} dx.
\]
We start with a lemma. Rescale the Gaussian distribution defined in (36) and define two new Gaussian distributions,

\[ p_g^+(x) = \frac{1}{Z_g^+} e^{G(x)(1 + \frac{cB(n)}{n})}, \quad p_g^-(x) = \frac{1}{Z_g^-} e^{G(x)(1 - \frac{cB(n)}{n})}, \]

(75)

**Lemma 17.** If \( B(n) \ll n \), the normalized constants \( Z_g^\pm \) satisfy

\[ Z_g^\pm = Z_g e^{\pm \frac{c}{2} B(n)(1 + o_n(1))}. \]

**Proof.** Change the variable to \( \tilde{x} = \sqrt{1 - \frac{cB(n)}{n}} x \), then

\[ Z_g^- = \int e^{-\frac{1}{2} \tilde{x}^T A \tilde{x} (1 - \frac{cB(n)}{n})} d\tilde{x} = \int e^{-\frac{1}{2} \tilde{x}^T A \tilde{x}} \prod_{j=1}^{n} \frac{1}{\sqrt{1 - \frac{cB(n)}{n}}} d\tilde{x} = Z_g \prod_{j=1}^{n} \frac{1}{\sqrt{1 - \frac{cB(n)}{n}}}. \]

Note that

\[ \left(1 - \frac{cB(n)}{n}\right)^{-\frac{n}{2}} = e^{\frac{cB(n)}{2}(1 + o_n(1))}. \]

Similarly, we have

\[ \left(1 + \frac{cB(n)}{n}\right)^{-\frac{n}{2}} = e^{-\frac{cB(n)}{2}(1 + o_n(1))}. \]

Thus

\[ Z_g^\pm = Z_g e^{\pm \frac{c}{2} B(n)(1 + o_n(1))}. \]

Let \( B(n) = D n^{\frac{1}{8}} \log^\frac{3}{2} n \). By the definition of \( \Gamma_D \) in (32), we have \( \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} \leq B(n) \). Then by (73), there exists a universal constant \( c > 0 \), such that

\[ \frac{cB(n)}{n} G(x) \leq F(x) \leq -\frac{cB(n)}{n} G(x). \]

Using the argument in Lemma 17, one can show that

\[ Z_g e^{-cB(n)} \leq \hat{Z}_n \leq Z_g e^{cB(n)}. \]

When we proceed to the Step 2 and Step 3 presented below, this estimate is not sufficient for us to show \( F(x) = o_n(1) \) because the exponential term \( e^{cB(n)} \) grows faster than any polynomial. In order to get a better upper bound of

\[ \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o}, \]

we come up with the idea of iteration. We need the following two lemmas.
Lemma 18. For some sufficient large $M > 0$, let $M \log^{1/2} n \leq B \leq Dn^{1/2} \log^{1/2} n$. If there exists some $\gamma > 0$ such that
\[
\mathbb{P}_x \left( x \in \Gamma_D : \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} \geq B \right) \leq n^{-\gamma},
\] (76)
then for some constant $c > 0$, we have
\[
Z_g e^{-cB} \leq \tilde{Z}_n \leq Z_g e^{cB}.
\] (77)

Proof. Denote $A = \{ x \in \Gamma_D : \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} \leq B \}$ and then $\mathbb{P}_x (\Gamma_D \setminus A) \leq n^{-\gamma}$. Because of (73), if $x \in A$, there exists a universal constant $c > 0$,
\[
\frac{cB}{n} G(x) \leq F(x) \leq \frac{cB}{n} G(x).
\] (78)
By (78), we have
\[
\tilde{Z}_n = \int_{\Gamma_D} e^{G(x) + F(x)} dx = \int_A e^{G(x) + F(x)} dx + \mathbb{P}_x (\Gamma_D \setminus A) \tilde{Z}_n \leq \int_A e^{G(x)(1 - \frac{cB}{n})} dx + n^{-\gamma} \tilde{Z}_n.
\]
Then
\[
(1 - n^{-\gamma}) \tilde{Z}_n \leq \int_{\Gamma} e^{G(x)(1 - \frac{cB}{n})} dx \leq e^{\frac{cB}{n}(1 + o(n))} Z_g.
\]
Thus
\[
\tilde{Z}_n \leq e^{\frac{cB}{n}(1 + o(n))} Z_g (1 + O(n^{-\gamma})) \leq e^{cB} Z_g.
\] (79)
For the lower bound, similarly,
\[
\tilde{Z}_n \geq \int_A e^{G(x) + F(x)} dx \geq \int_A e^{G(x)(1 + \frac{cB}{n})} dx = \int_{\Gamma} e^{G(x)(1 + \frac{cB}{n})} dx - \int_{\Gamma \setminus A} e^{G(x)(1 + \frac{cB}{n})} dx
\]
\[
\geq \int_{\Gamma} e^{G(x)(1 + \frac{cB}{n})} dx - \int_{\Gamma \setminus A} e^{G(x)} dx = \left( e^{\frac{cB}{n}(1 + o(n))} - \mathbb{P}_g(\Gamma \setminus A) \right) Z_g.
\]
By Lemma 16, we have
\[
\mathbb{P}_g(\Gamma \setminus A) \leq \mathbb{P}_g(\Gamma^c_D) + \mathbb{P}_g \left( \bigcup_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} \geq B \right) \leq e^{-c_1 n^2 + n^2 e^{-c_2 n \log^3 n} + n^2 \max_{j,l} \mathbb{P}_g \left( \frac{\xi_j^{(l)}}{\sqrt{l}} \geq B \right)}.
\]
Since the variance of $\xi_j^{(l)}$ is at most of the order $l$ by Proposition 15, we have
\[
\mathbb{P}_g \left( \frac{\xi_j^{(l)}}{\sqrt{l}} \geq B \right) \leq e^{-c'B^2},
\] (80)
where $c'$ is a universal constant. Thus,
\[
\mathbb{P}_g(\Gamma \setminus A) \leq e^{-c_1 n^2 + n^2 e^{-c_2 n \log^3 n} + n^2 e^{-c'B^2}},
\]
and
\[ \tilde{Z}_n \geq (e^{-\frac{cB}{2}(1+o_1(1))} - n^2 e^{-c'B^2} - e^{-c_1 n^2} - n^2 e^{-c_2 n \log^3 n}) Z_g. \]

If \( B \geq M \log^\frac{3}{2} n \) with some sufficient large \( M > 0 \), then \( n^2 e^{-c'B^2} + e^{-c_1 n^2} + n^2 e^{-c_2 n \log^3 n} \) is much smaller than \( e^{-\frac{cB}{2}} \). We have
\[ \tilde{Z}_n \geq e^{-\frac{cB}{2}(1+o_1(1))} Z_g (1 - o_1(1)) \geq Z_g e^{-cB}. \] (81)

\[ \square \]

Using the result of Lemma 18, we can prove Lemma 19.

**Lemma 19.** For some sufficient large \( M > 0 \), let \( M \log^\frac{3}{2} n \leq B_k \leq Dn^2 \log^\frac{3}{2} n \). If there exists some \( \gamma > 0 \) such that \( kn^{-10} \leq n^{-\gamma} \) and
\[ \mathbb{P}_x \left( x \in \Gamma_D : \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} \geq B_k \right) \leq kn^{-10}, \] (82)
then
\[ \mathbb{P}_x \left( x \in \Gamma_D : \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} \geq B_{k+1} \right) \leq (k+1)n^{-10}, \] (83)
with
\[ B_{k+1} = \sqrt{\frac{4cB_k + 24 \log n}{c'}}. \] (84)

Here \( c, c' \) are universal constants that do not depend on \( n, k \).

**Proof.** Denote \( A_k = \{ x \in \Gamma_D : \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} \leq B_k \} \) and \( A_k^{\text{com}} = \Gamma_D \setminus A_k \). Then \( \mathbb{P}_x(A_k^{\text{com}}) \leq kn^{-10} \) by (82). Since \( B_{k+1} \leq B_k \), then \( A_k^{\text{com}} \subset A_k \subset \Gamma_D \) and \( A_k^{\text{com}} \subset A_{k+1}^{\text{com}} \). By Lemma 18, if \( x \in A_k \), then there exists a universal constant \( c > 0 \), such that
\[ \mathbb{P}_x \left( A_{k+1}^{\text{com}} \right) = \mathbb{P}_x(A_k^{\text{com}}) + \mathbb{P}_x(A_k \cap A_{k+1}^{\text{com}}) \leq kn^{-10} + \frac{1}{Z_n} \int_{A_{k+1}^{\text{com}}} e^{G(x)(1 - \frac{cB_k}{n})} dx \]
\[ = kn^{-10} + \frac{Z_g}{Z_n} \int_{A_{k+1}^{\text{com}}} e^{G(x)(1 - \frac{cB_k}{n})} dx \leq kn^{-10} + e^{cB_k} \left( \frac{1}{Z_g} \int_{A_{k+1}^{\text{com}}} e^{G(x)(1 - \frac{cB_k}{n})} dx \right). \] (85)

Let \( \tilde{x} = \sqrt{1 - \frac{cB_k}{n}} x \). Then
\[ \frac{1}{Z_g} \int_{A_{k+1}^{\text{com}}} e^{G(x)(1 - \frac{cB_k}{n})} dx = \left( 1 - \frac{cB_k}{n} \right)^{\frac{1}{2}} \frac{1}{Z_g} \int_{A_{k+1}^{\text{com}}} e^{G(\tilde{x})} d\tilde{x} \]
\[ \leq e^{cB_k} \frac{1}{Z_g} \int_{A_{k+1}^{\text{com}}} e^{G(\tilde{x})} d\tilde{x} = e^{cB_k} \mathbb{P}_g(A_{k+1}^{\text{com}}). \] (86)

where \( A_{k+1}^{\text{com}} = \left\{ x \in \Gamma_D : \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} > B_{k+1} \sqrt{1 - \frac{cB_k}{n}} \right\} \).
Note that
\[
\Pr_g(A_{k+1}) \leq \Pr \left( \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|} \geq B_{k+1} \sqrt{1 - \frac{cB_k}{n}} \right) \leq n^2 \max_{j,l} \Pr \left( |\xi_j^{(l)}| \geq B_{k+1} \sqrt{1 - \frac{cB_k}{n}} \right) \leq n^2 \max_{j,l} \Pr \left( |\xi_j^{(l)}| \geq B_{k+1} \sqrt{1 - \frac{cB_k}{n}} \right) \leq n^2 \max_{j,l} \Pr \left( \frac{|\xi_j^{(l)}|}{\sqrt{l}} \geq B_{k+1} \sqrt{1 - \frac{cB_k}{n}} \right) \leq n^2 e^{-c'B_{k+1}^2(1 - \frac{cB_k}{n})}.
\]

where \( c' > 0 \) is the universal constant introduced in \( (80) \). Thus

LHS of \( (86) \) \leq n^2 e^{-c'B_{k+1}^2(1 - \frac{cB_k}{n}) + cB_k} \leq n^2 e^{-\frac{c'}{2}B_{k+1}^2 + cB_k + 2 \log n}. \quad (87)

Therefore, combining \( (85), (86), \) and \( (87) \), we have

\[
\Pr_x(A_{k+1}) \leq kn^{-10} + e^{-\frac{c'}{2}B_{k+1}^2 + 2cB_k + 2 \log n}. \quad (88)
\]

By letting the RHS of \( (88) \) equal \( (k+1)n^{-10} \), we can solve

\[-c'B_{k+1}^2 + 4cB_k + 4 \log n = -20 \log n \quad \text{for } B_{k+1}. \]

We obtain

\[
B_{k+1} = \sqrt{\frac{4cB_k + 24 \log n}{c'}}.
\]

Combining Lemma \( 18 \) and \( 19 \) we have

**Proposition 20.** There exist some constants \( C_1, C_2 > 0 \), such that for sufficient large \( n \),

\[
\Pr_x \left( x \in \Gamma_D : \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|} \geq C_1 \log^\frac{1}{2} n \right) \leq n^{-9}, \quad (89)
\]

and

\[
Z_g e^{-C_2 \log^\frac{1}{2} n} \leq \tilde{Z} \leq Z_g e^{C_2 \log^\frac{1}{2} n}. \quad (90)
\]

**Proof.** The definition of \( \Gamma_D \) in \( (32) \) indicates that for \( k = 0 \), \( (32) \) is satisfied with \( B_0 = Dn^{\frac{1}{2}} \log^\frac{3}{2} n. \) Then we use Lemma \( 18 \) and Lemma \( 19 \) to proceed the iteration by setting

\[
B_{k+1} = \sqrt{\frac{4cB_k + 24 \log n}{c'}}. \quad (91)
\]

Note that in Lemma \( 19 \) \( c, c' \) are universal constants.

The fixed point of the iteration \( (91) \) is

\[-c'B^2_f + 4cB_f + 4 \log n = -20 \log n \quad \Rightarrow \quad B_f = \frac{2c + \sqrt{4c^2 + 24c' \log n}}{c'} \sim \log^\frac{1}{2} n. \]

Recall that \( B_0 \sim n^{\frac{3}{2}} \log^\frac{5}{2} n. \) Moreover, if \( B_k \gtrsim \log n \), then \( B_{k+1} \sim \sqrt{B_k}. \) This implies that for \( B_k \) to reach the value of order \( \log^\frac{1}{2} n \), one needs about \( \log \log n \) iteration steps. In other
words, the sequence \( \{B_k\} \) will starts from \( B_0 = Dn^{\frac{3}{2}} \log^2 n \) and after \( C_1 \log \log n \) number of iterations, \( B_k \) decreases below the level

\[
B_\infty = C_1 \log^\frac{1}{2} n.
\]

Finally, we still need to check if the conditions of Lemma 18, 19 are satisfied. To satisfy the first condition of Lemma 19 (also Lemma 18), i.e. \( M \log^2 n \leq B_k \leq Dn^{\frac{3}{2}} \log^\frac{3}{2} n \), we need to modify the stopping time of the iteration process. We will end the iteration right before \( B_k \) falls below \( M \log \frac{1}{2} n \). But the result remains the same. Note that the number of iteration steps is of the order \( \log \log n \), so the second condition of Lemma 19 also holds, i.e. \( kn^{-10} \leq n^{-\gamma} \) for some \( \gamma > 0 \).

Therefore, we can find a subset, denoted as

\[
A_\infty = \left\{ x \in \Gamma_D : \max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} \leq C_1 \log^\frac{1}{2} n \right\},
\]

(92)

such that

\[
P_x(\Gamma_D \setminus A_\infty) \leq n^{-9}.
\]

(93)

Moreover, by Lemma 18 there exists some constant \( C_2 \), such that

\[
Z_g e^{-C_2 \log^\frac{1}{2} n} \leq \tilde{Z}_n \leq Z_g e^{C_2 \log^\frac{1}{2} n}.
\]

Combining (93) and (33), we have

**Corollary 21.** There exists a subset \( A_\infty \) defined in (92) such that

\[
P_x(A_\infty^c) \leq n^{-8}.
\]

(94)

### 6.2 Step 2: Estimate of \( F(x) \) under Gaussian distribution

If \( x \in A_\infty \), which is defined in (92), then

\[
\max_{i \neq j} \frac{|x_i - x_j|}{|i - j|_o} \leq C_1 \log^\frac{1}{2} n,
\]

(95)

and by (72), we obtain that

\[
F(x) = O \left( \sum_{i > j} \frac{(x_i - x_j)^3}{n|i - j|_o^{\frac{5}{2}}} \right) = O \left( \log^\frac{3}{2} n \right).
\]

(96)

Note that

\[
\sum_{i > j, |i - j|_o > \log^3 n} \frac{|x_i - x_j|^3}{n|i - j|_o^{\frac{5}{2}}} \leq C \log^\frac{3}{2} n \sum_{i > \log^3 n} \frac{1}{l^2} \leq C \log^{-\frac{3}{2}} n.
\]

(97)
Thus, it is sufficient for us to estimate
\[\sum_{i > j, |i-j|_o \leq \log^3 n} \frac{(x_i - x_j)^3}{n|i-j|_o^5} = \sum_{k=1}^{\log^3 n} \frac{1}{15} \left(\frac{1}{n} \sum_{j=0}^{n-l-1} \xi_j^{(l)^3}\right),\] (98)
where \(\xi_j^{(l)}\) is defined in (65). Denote
\[\Omega_l = \left\{ \left| \frac{1}{n} \sum_{i=1}^{n-l-1} \xi_i^{(l)^3} \right| \leq n^{-\frac{1}{4}} \right\}, \quad 1 \leq l \leq \log^3 n,\] (99)
and \(\Omega_\infty = \bigcap_{l \leq \log^3 n} \Omega_l\). Note that if \(x \in \Omega' := \Omega_\infty \cap A_\infty\), then
\[\sum_{l=1}^{\log^3 n} \frac{1}{15} \left(\frac{1}{n} \sum_{j=0}^{n-l-1} \xi_j^{(l)^3}\right) \leq n^{-\frac{1}{4}} \sum_{l=1}^{\log^3 n} \frac{1}{15} = O(n^{-\frac{1}{4}}).\] (100)
Combining (98), (100), and (97), we have
\[F(x) = O(\log^{-\frac{3}{2}} n).\] (101)
Using Proposition 15 and Lemma 16 one can show that
\[\mathbb{P}_g(A^c_\infty) \leq \mathbb{P}_g \left( \text{max}_{i \neq j} \frac{|x_i - x_j|}{|i-j|_o} \geq C_1 \log \frac{1}{n} \right) + \mathbb{P}_g (\Gamma^c_D) \leq n^2 e^{-C_1' \log n} + e^{-c_1 n^2} + n^2 e^{-c_2 n \log^3 n} \leq n^{-\frac{1}{4}},\] (102)
provided that \(C_1\) is chosen sufficiently large.
Next, we want to show that \(\mathbb{P}_g(\Omega^c_\infty) = o_n(1)\). The following lemma is useful.

**Lemma 22.** There exists some constant \(C > 0\) such that
\[\mathbb{P}_g \left( \left| \frac{1}{n} \sum_{i=1}^{n-1} \xi_i^{(l)^3} \right| > n^{-\frac{1}{4}} \right) \leq \frac{C l^4 \log n}{n^{\frac{1}{2}}}.\] (103)

**Proof.** By Wick’s formula, we have
\[\mathbb{E} \xi_j^{(l)^3} \xi_k^{(l)^3} = 9 \left( \mathbb{E} \xi_j^{(l)^2} \xi_k^{(l)^2} \mathbb{E} \xi_j^{(l)^3} \xi_k^{(l)^3} \right) + 6 \left( \mathbb{E} \xi_j^{(l)^3} \xi_k^{(l)^3} \right)^2.\]

By Proposition 15 \(\mathbb{E} \xi_j^{(l)^2} \leq C l^2\) and \(\mathbb{E} \xi_j^{(l)^3} \leq \frac{C l^2}{|j-k|_o}\) for \(|j-k| \geq l\). Then
\[\mathbb{E} \xi_j^{(l)^3} \xi_k^{(l)^3} \leq C^3 \frac{l^4}{|j-k|} + C^3 \frac{l^6}{|j-k|^3} \leq \frac{C l^4}{|j-k|}.\]
Similarly, if \(|j-k| \leq l\), then \(\mathbb{E} \xi_j^{(l)^3} \xi_k^{(l)^3} \leq C l^4\), and thus \(\mathbb{E} \xi_j^{(l)^3} \xi_k^{(l)^3} \leq 15C^3 l^3\). Therefore, there exists some constant \(C' > 0\) such that
\[\mathbb{E} \left( \sum_{i=1}^{n-1} \xi_i^{(l)^3} \right)^2 \leq C' l^4 n \log n.\]
It follows from the Markov Inequality that
\[
\mathbb{P}_g \left( \left| \frac{1}{n} \sum_{i=1}^{n-1} \xi_i^{(l)} \right| > n^{-\frac{1}{2}} \right) \leq \frac{\mathbb{E} \left( \sum_{i=1}^{n-1} \xi_i^{(l)} \right)^2}{n^2} \leq \frac{C' l^4 \log n}{n^2}.
\] (104)

By using Lemma 22, it can be shown directly that
\[
\mathbb{P}_g (\Omega_{\infty}^c) \leq \mathbb{P}_g \left( \bigcup_{l \leq \log^3 n} \Omega_c^l \right) \leq \frac{C' l^4 \log^3 n}{n^2} \leq \frac{C' \log^4 n}{n^2} \leq n^{-\frac{1}{4}}.
\] (105)

Let \( x \in \Omega' = \Omega_{\infty} \cap A_{\infty} \). Combining (102) and (105), we have
\[
\mathbb{P}_g (\Omega_{\infty}^c) \leq \mathbb{P}_g (\Omega_{\infty}^c) + \mathbb{P}_g (A_{\infty}^c) \leq 2n^{-\frac{1}{4}}.
\] (106)

Therefore, we have the following lemma.

**Lemma 23.** There exists a subset of \( \Omega' \subset \Gamma \) such that
\[
\mathbb{P}_g (\Omega_{\infty}^c) = o_n(1),
\]
and if \( x \in \Omega' \), then \( F(x) = o_n(1) \).

### 6.3 Step 3: Combining Step 1 and Step 2

In this subsection, we finish the proofs of Proposition 9 and Theorem 10 by combining the results in Step 1 and Step 2. In Step 1, we have showed that \( \mathbb{P}_x (A_{\infty}^c) = o_n(1) \). In Step 2, we have obtained that \( \mathbb{P}_g (\Omega_{\infty}^c) = o_n(1) \) and \( \mathbb{P}_g (A_{\infty}^c) = o_n(1) \).

**Proof.** (Proposition 9 and Theorem 10)

We start with showing \( \mathbb{P}_x (\Omega_{\infty}^c) = o_n(1) \). Recall the notations \( \Omega_l = \left\{ \left| \frac{1}{n} \sum_{i=1}^{n-l-1} \xi_i^{(l)} \right| \leq n^{-\frac{1}{2}} \right\} \), \( 1 \leq l \leq \log^3 n \), and \( \Omega_{\infty} = \bigcap_{l \leq \log^3 n} \Omega_l \). Then
\[
\mathbb{P}_x (\Omega_{l}^c) \leq \mathbb{P}_x (\Gamma_{\infty}^c) + \mathbb{P}_x (A_{\infty}^c) + \mathbb{P}_x (\Omega_{l}^c \cap A_{\infty} \cap \Gamma_D).
\]

By (33) and (94),
\[
\mathbb{P}_x (\Omega_{l}^c) \leq n^{-cn} + n^{-8} + \frac{1}{Z_n} \int_{\Omega_l} e^{G(x)(1 - C_1 \log \frac{1}{n})} dx \leq 2n^{-8} + \frac{Z_g}{Z_n} \frac{1}{Z_g} \int_{\Omega_l} e^{G(x)(1 - C_1 \log \frac{1}{n})} dx.
\] (107)

Changing the variable \( \tilde{x} = \sqrt{1 - \frac{C_1 \log \frac{1}{n}}{n}} x \), by Lemma 17, we have
\[
\mathbb{P}_x (\Omega_{l}^c) \leq 2n^{-8} + e^{C_1 \log \frac{1}{n}} \cdot \frac{Z_g}{Z_n} \cdot \mathbb{P}_g (\tilde{\Omega}_{l}^c),
\] (108)
where \( \tilde{\Omega}_t^c = \left\{ \frac{1}{n} \sum_{i=1}^{n-l-1} \tilde{\xi}^{(i)}_t \geq n^{-\frac{1}{4}} \left( 1 - \frac{C_1 \log^{\frac{2}{3}}_n}{n} \right)^{\frac{3}{4}} \right\} \).

Using similar argument in (104) in Lemma 22, we have

\[
\mathbb{P}_g(\tilde{\Omega}_t^c) = \frac{1}{Z_g} \int_{\tilde{\Omega}_t^c} e^{G(x)} dx \leq \frac{C' l^4 \log(n - l)}{n^{\frac{1}{2}} \left( 1 - \frac{C_1 \log^{\frac{2}{3}}_n}{n} \right)^{\frac{3}{4}}} \leq \frac{C' l^4 \log n}{n^{\frac{1}{2}}}.
\]  

(109)

Combing (109), (108), and (90), we obtain

\[
\mathbb{P}_x(\Omega_t^c) \leq 2 n^{-8} + \frac{C' l^4 \log n e^{(C_1 + C_2) \log^{\frac{2}{3}}_n}}{n^{\frac{1}{2}}}. 
\]

Thus,

\[
\mathbb{P}_x(\Omega_t^c) = \mathbb{P}_x \left( \bigcup_{l \leq \log^{\frac{3}{2}} n} \Omega_t^c \right) \leq 2 n^{-8} \log^3 n + \frac{C' l^4 \log^{13} n e C_3 \log^{\frac{2}{3}}_n}{n^{\frac{1}{2}}} = O(n^{-\frac{3}{4}}). 
\]  

(110)

Repeating the arguments from Step 2, we have that (96), (97) and (98) hold for \( x \in A_\infty \). If \( x \in \Omega_\infty \), (100) also holds. Therefore, if \( x \in \Omega' = A_\infty \cap \Omega_\infty \), the bound (101) on \( F \) still holds. In addition, by (94) and (110), we have

\[
\mathbb{P}_x(\Omega_t^c) \leq \mathbb{P}_x(\Omega_\infty^c) + \mathbb{P}_x(A_\infty^c) = O(n^{-\frac{3}{4}}). 
\]  

(111)

Thus, we have proved Proposition 9. Combining (111), (106), and (101), one can show that

\[
(1 - C' \log^{-\frac{3}{2}} n)Z_g \leq \tilde{Z}_n \leq (1 + C' \log^{-\frac{3}{2}} n)Z_g. 
\]

If \( A \) is some measurable subset of \( \Omega' \), then

\[
\mathbb{P}_x(A) = \frac{1}{\tilde{Z}_n} \int_A e^{G(x) + F(x)} \leq (1 + C'' \log^{-\frac{3}{2}} n) \frac{1}{Z_g} \int_A e^{G(x)} dx = \mathbb{P}_g(A) \left( 1 + O \left( \log^{-\frac{3}{2}} n \right) \right). 
\]

Similarly, we have

\[
\mathbb{P}_x(A) \geq (1 - C'' \log^{-\frac{3}{2}} n) \frac{1}{Z_g} \int_A e^{G(x)} dx = \mathbb{P}_g(A) \left( 1 - O \left( \log^{-\frac{3}{2}} n \right) \right). 
\]

Combining with (111) and (106), we conclude that for any measurable set \( A \subset \Gamma \),

\[
|\mathbb{P}_x(A) - \mathbb{P}_g(A)| = O \left( \log^{-\frac{3}{4}} n \right). 
\]  

(112)
7 Proof of Theorem 11

In this section, we prove functional convergence in distribution of \( \zeta_n(t) \) to \( \zeta(t) \).

**Proof.** Fix finitely many 0 \( \leq t_1, \ldots, t_m \leq 2\pi \). Let \( j_l = \lfloor n t_l \rfloor \), \( l = 1, \ldots, m \). Because of the construction of \( \zeta_n(t) \) and \((95)\), with high probability, we have

\[
\left| \zeta_n(t_l) - \frac{x_{j_l}}{\sqrt{n}} \right| \leq \left| \frac{x_{j_l}}{\sqrt{n}} - \frac{x_{j_{l+1}}}{\sqrt{n}} \right| = o_n(1)
\]

By Theorem 10, the finite-dimensional distribution of \((x_{j_1}, \ldots, x_{j_m})\) can be approximated by the finite-dimensional distribution of the Gaussian law defined in \((36)\). Without loss of generality, assume that \( n \) is odd. For even case, similar considerations hold. Using the representation \((64)\) for \( x_j \), we have

\[
x_j = \frac{2}{n} \sum_{k=1}^{n-1} \left( \cos \left( \frac{2\pi jk}{n} \right) \frac{n^2}{2k(n-k)} p_k - \sin \left( \frac{2\pi jk}{n} \right) \frac{n^2}{2k(n-k)} q_k \right)
\]

\[
= \sum_{k=1}^{n-1} \left( \cos \frac{2\pi jk}{n} p_k - \sin \frac{2\pi jk}{n} q_k \right) + \sum_{k=1}^{n-1} \left( \cos \frac{2\pi jk}{n-k} p_k - \sin \frac{2\pi jk}{n-k} q_k \right)
\]

\[
= \sum_{k=1}^{\infty} \left( \cos \frac{2\pi jk}{n} p_k - \sin \frac{2\pi jk}{n} q_k \right) + e_n, \quad \text{for } 0 \leq j \leq n - 1,
\] \((113)\)

where \( p_k \) and \( q_k \) are i.i.d. real standard normal random variables. Here

\[
e_n = \sum_{k=\frac{n+1}{2}}^{\infty} \left( \cos \frac{2\pi jk}{n} p_k - \sin \frac{2\pi jk}{n} q_k \right) + \sum_{k=1}^{n-1} \left( \cos \frac{2\pi jk}{n-k} p_k - \sin \frac{2\pi jk}{n-k} q_k \right)
\]

is negligible because

\[
\text{Var}_g(e_n) = \sum_{k=\frac{n+1}{2}}^{\infty} \frac{1}{k^2} + \sum_{k=1}^{n-1} \left( \frac{1}{n-k} \right)^2 = O \left( \frac{1}{n} \right).
\]

Therefore, by \((112)\), for \( 1 \leq l \leq m \),

\[
\zeta_n \left( \frac{2\pi j_l}{n} \right) = \frac{x_{j_l}}{\sqrt{n}} = \sum_{k=1}^{\infty} \left( \cos \frac{2\pi j_l k}{n} p_k - \sin \frac{2\pi j_l k}{n} q_k \right) + e_n = \zeta \left( \frac{2\pi j_l}{n} \right) + e_n,
\] \((114)\)

where \( \{p_k\}, \{q_k\} \) are iid real Gaussian variables, \( e_n \) is a random error term with \( \text{Var}(e_n) = o_n(1) \). Therefore, one proves that \( \zeta_n(t) \) converges in finite dimensional distribution to \( \zeta(t) \).

Now, we turn our attention to functional convergence. Note that the sequence of the distributions of \( \zeta_n(t) \) gives a family of probability measures on the space \( C[0, 2\pi] \). Because of the finite-dimension distribution convergence, it is sufficient for us to show the tightness of the distribution sequence. A sequence of probability measures \( \{\mathcal{P}_n\} \) is tight if only if the following two conditions hold \((4)\):

24
1. For any small $\eta > 0$, there exist corresponding $a$ and $n_0$, such that

$$\mathcal{P}_n(f : |f(0)| \geq a) \leq \eta, \text{ for } n \geq n_0.$$ 

2. For any small $\epsilon, \eta > 0$, there exist corresponding $\delta_0$ and $n_0$, such that

$$\mathcal{P}_n(f : \omega_f(\delta_0) \geq \epsilon) \leq \eta, \text{ for } n \geq n_0.$$ 

where

$$\omega_f(\delta) = \sup\{|f(s) - f(t)| : 0 \leq s, t \leq 2\pi, |s - t| < \delta\}.$$ 

To check the first condition, by (114), we note that

$$\zeta_n(0) = \frac{x_0}{\sqrt{n}} = \sum_{k=1}^{\infty} \frac{1}{k} p_k + e_n,$$ 

and thus

$$\text{Var}(\zeta_n(0)) = \sum_{k=1}^{\infty} \frac{1}{k^2} + o_n(1) = \frac{\pi^2}{6} + o_n(1).$$ 

Then

$$\mathbb{P}_x(|\zeta_n(0)| \geq a) \leq \frac{\pi^2}{6a^2},$$ 

we choose $a = \sqrt{\frac{\pi^2}{6\eta}}$ and $n_0 = 1$. 

Next, to check the second condition, we need the following lemma.

**Lemma 24.** There exist positive constants $c_k (1 \leq k \leq 6)$ such that

$$\mathbb{P}_x \left( |\zeta_n(t) - \zeta_n(s)| \leq c_1 |t - s|^{\frac{1}{30}} + c_2 n^{-\frac{1}{10}} \log n, \forall t, s \in [0, 2\pi] : |t - s| \leq \delta \right)$$

$$> 1 - (c_3 e^{-c_4 n^{\frac{1}{6}}} + c_5 \delta^{\frac{1}{6}} + c_6 n^{\frac{1}{10}} \log^{-\frac{3}{2}} n).$$

Assuming that Lemma [24] is proved, we can finish the proof of Theorem [11] by choosing $\delta_0$ and $n_0$ such that

$$c_1 \delta_0^{\frac{1}{30}} \leq \epsilon, \quad c_2 n_0^{-\frac{1}{10}} \log n_0 \leq \epsilon;$$

$$c_3 e^{-c_4 n_0^{\frac{1}{6}}} \leq \frac{\eta}{3}, \quad c_5 \delta_0^{\frac{1}{6}} \leq \frac{\eta}{3}, \quad c_6 n_0^{\frac{1}{10}} \log^{-\frac{3}{2}} n_0 \leq \frac{\eta}{3}.$$ 

Finally, we only need to prove Lemma [24].

**Proof.** (Lemma [24]) Because of (112), it is sufficient to prove that

$$\mathbb{P}_g \left( |\zeta_n(t) - \zeta_n(s)| \leq c_1 |t - s|^{\frac{1}{30}} + c_2 n^{-\frac{1}{10}} \log n, \forall t, s \in [0, 2\pi] : |t - s| \leq \delta \right)$$

$$> 1 - (c_3 e^{-c_4 n^{\frac{1}{6}}} + c_5 \delta^{\frac{1}{6}}).$$
Without loss of generality, assume that $s < t$.

Case 1: Let us fixe $C_0 > 0$ and assume that $|t - s| \leq \frac{C_0}{n}$, then there exist $i, j$ with $|i - j| \leq C_0 + 2$, such that

$$s \in \left[ \frac{2\pi(i - 1)}{n}, \frac{2\pi i}{n} \right], \quad t \in \left[ \frac{2\pi j}{n}, \frac{2\pi(j + 1)}{n} \right]. \tag{120}$$

By the triangle inequality, we have

$$P_p \left( |\zeta_n(t) - \zeta_n(s)| > \epsilon \right) \leq \sum_{i \leq k \leq j + 1} P_p \left( \left| \zeta_n \left( \frac{2\pi(k - 1)}{n} \right) - \zeta_n \left( \frac{2\pi k}{n} \right) \right| \geq C_2 \epsilon \right).$$

Note that $\zeta_n \left( \frac{2\pi k}{n} \right) = \frac{x_k}{\sqrt{n}}$ and $\text{Var}(x_k - x_{k-1}) \sim 1$, then for some constant $C_2, C_3$ depending on $C_0$, we have

$$P_p \left( \left| \zeta_n \left( \frac{2\pi(k - 1)}{n} \right) - \zeta_n \left( \frac{2\pi k}{n} \right) \right| \geq C_2 \epsilon \right) = P_p \left( |x_{k-1} - x_k| \geq C_2 \epsilon \sqrt{n} \right) \leq C_3 e^{-C_2^2 \epsilon^2 n}.$$ 

Thus

$$P_p \left( |\zeta_n(t) - \zeta_n(s)| > \epsilon \right) \leq (C_0 + 2)C_3 e^{-C_2^2 \epsilon^2 n}. \tag{121}$$

Furthermore, there exist constants $C_4, C_5$ which only depend on $C_0$, such that

$$P_p \left( |\zeta_n(t) - \zeta_n(s)| > \epsilon \right) \leq \frac{C_0}{n} \leq C_4 e^{-C_2^2 \epsilon^2 n} \leq C_5 e^{-C_4 n^2 \epsilon^2}. \tag{122}$$

Case 2: If $|t - s| \gg \frac{1}{n}$, we introduce a new partition of $[s, t]$. We start by dividing the interval $[0, 2\pi]$ into $2^k$ disjoint subintervals

$$\Delta_l^{(k)} = \left[ \frac{2\pi l}{2^k}, \frac{2\pi}{2^k} (l + 1) \right], \quad l = 0, 1, \ldots, 2^k - 1. \tag{123}$$

There exists the smallest $k = k_0$, and related $l_0$, such that $\Delta_{l_0}^{(k_0)} \subset [s, t]$. Note that $k_0$ and $l_0$ are unique. Let $S_0 = \Delta_{l_0}^{(k_0)}$. If $[s, t] \neq S_0$, then there exists the unique smallest $k_1 > k_0$, and one or two values of $l_1$, such that $\Delta_{l_1}^{(k_1)} \subset [s, t] \setminus S_0$ (we could potentially add $\Delta_{l_1}^{(k_1)}$ on the left of $\Delta_{l_0}^{(k_0)}$ or add one on the right). If there is only one value of $l_1$, let $\Delta_{a_{l_1}}^{(k_1)} = \Delta_{l_1}^{(k_1)}$ and $\Delta_{b_{l_1}}^{(k_1)} = \emptyset$. If there are two values of $l_1$, let $a_1$ be the smallest of the two and $b_1$ the largest. Set $S_1 = S_0 \cup \Delta_{a_1}^{(k_1)} \cup \Delta_{b_1}^{(k_1)}$. We continue this process. For each $m \geq 2$, find a unique smallest $k_m > k_{m-1}$, $\Delta_{a_m}^{(k_m)} \subset [s, t] \setminus S_{m-1}$ (recall that $\Delta_{b_m}^{(k_m)}$ might be empty). Let $S_m = S_{m-1} \cup \Delta_{b_m}^{(k_m)} \cup \Delta_{a_m}^{(k_m)}$. We will stop at $m = r$, when either $[s, t] = S_r$ or the length of $S_{r+1} := [s, t] \setminus S_r \leq \frac{C_0}{n}$, where $C_0$ is some fixed constant. Thus,

$$[s, t] = S_0 \cup S_{r+1} \bigcup_{m=1}^{r} \bigcup_{m=1}^{r} \Delta_{b_m}^{(k_m)} \cup \Delta_{a_m}^{(k_m)}. \tag{124}$$

Note that

$$\frac{1}{2k_0} \leq |s - t| \leq \frac{3}{2k_0}, \quad \frac{1}{2k_r} \leq \frac{C_0}{n}, \tag{125}$$

26
Lemma 25. Fix $D$ where the last inequality comes from $2\zeta_n\left(\frac{2\pi}{n}(l+1)\right) - \zeta_n\left(\frac{2\pi}{n}l\right)$. Let $a_0 = l_0$. Then by the triangle inequality, we have

$$D_n([s, t]) \leq D_n(\Delta_{l_0}^{(k_0)}) + \sum_{m=1}^{r} D_n(\Delta_{a_m}^{(k_m)}) + D_n(S_{r+1}) \leq 2 \sum_{m=0}^{r} D_n(\Delta_{a_m}^{(k_m)}) + D_n(S_{r+1}).$$

Since $|S_{r+1}| \leq \frac{C_0}{n}$, using the same argument in Case 1, we have for some constant $C_1, C_2$,

$$\mathbb{P}_g\left(D_n(S_{r+1}) > n^{-\frac{1}{10}}, \exists S_{r+1} \subset [0, 2\pi] : |S_{r+1}| \leq \frac{C_0}{n}\right) \leq C_1 e^{-C_2 n^{\frac{4}{5}}}.$$  \hspace{1cm} (126)

To estimate $D_n(\Delta_{a_m}^{(k_m)})$, we need the following lemma.

**Lemma 25.** Fix $s, t \in [0, 2\pi]$. Then there exist some constants $C_1, C_2, C_4 > 0$, such that

$$\mathbb{P}_g\left(|\zeta_n(t) - \zeta_n(s)| > |t - s|^{\frac{1}{30}} + 2n^{-\frac{1}{10}}\right) \leq 2C_1 e^{-C_2 n^{\frac{4}{5}}} + C_4 |t - s|^{\frac{5}{2}}.$$

**Proof.** (Lemma 25) Fix $s, t$. Then there exist $i, j$, such that

$$s \in \left[\frac{2\pi(i - 1)}{n}, \frac{2\pi i}{n}\right], \quad t \in \left[\frac{2\pi j}{n}, \frac{2\pi(j + 1)}{n}\right].$$

Thus, we have

$$\mathbb{P}_g\left(|\zeta_n(t) - \zeta_n(s)| > |t - s|^{\frac{1}{30}} + 2n^{-\frac{1}{10}}\right) \leq \mathbb{P}_g\left(|\zeta_n\left(\frac{2\pi(i - 1)}{n}\right) - \zeta_n\left(\frac{2\pi i}{n}\right)| > n^{-\frac{1}{10}}\right)
+ \mathbb{P}_g\left(|\zeta_n\left(\frac{2\pi j}{n}\right) - \zeta_n\left(\frac{2\pi(j + 1)}{n}\right)| > n^{-\frac{1}{10}}\right) + \mathbb{P}_g\left(|\zeta_n\left(\frac{2\pi i}{n}\right) - \zeta_n\left(\frac{2\pi j}{n}\right)| > |t - s|^{\frac{1}{30}}\right).$$

Note that $\zeta_n\left(\frac{2\pi k}{n}\right) = \frac{2\pi k}{n}$, and $\text{Var}(x_k - x_i) \sim |k - l|$, then we have

$$\mathbb{P}_g\left(|\zeta_n\left(\frac{2\pi i}{n}\right) - \zeta_n\left(\frac{2\pi j}{n}\right)| > |t - s|^{\frac{1}{30}}\right) \leq C_3 |i - j|^2 \leq C_4 |t - s|^{\frac{5}{2}};$$

where the last inequality comes from $\frac{2\pi|i - j|}{n} \leq |t - s|$. Thus,

$$\mathbb{P}_g\left(|\zeta_n(t) - \zeta_n(s)| > |t - s|^{\frac{1}{30}} + 2n^{-\frac{1}{10}}\right) \leq 2C_1 e^{-C_2 n^{\frac{4}{5}}} + C_4 |t - s|^{\frac{5}{2}}.$$
By Lemma 25, we have
\[
\mathbb{P}_g \left( \bigcup_{l=0}^{k-1} \left\{ D(\Delta_l^{(k)}) > \left( \frac{2\pi}{2^k} \right)^{\frac{1}{20}} + 2n^{-\frac{1}{10}} \right\} \right) \leq 2^k \left( 2C_1 e^{-C_2 n^{\frac{2}{5}}} + C_4 \left( \frac{2\pi}{2^k} \right)^{\frac{9}{2}} \right)
\]
and thus
\[
\mathbb{P}_g \left( \bigcup_{l=0}^{k-1} \left\{ D(\Delta_l^{(k)}) > \left( \frac{2\pi}{2^k} \right)^{\frac{1}{20}} + 2n^{-\frac{1}{10}} \right\} \right) \leq 2^{k+1} e^{-C_2 n^{\frac{2}{5}}} + \frac{C_5}{(2^k)^{\frac{3}{2}}}, \tag{127}
\]
Combining with (126), we have
\[
\mathbb{P}_g \left( \bigcap_{l=0}^{k-1} \bigcap_{l=0}^{k-1} \left\{ D(\Delta_l^{(k)}) \leq \left( \frac{2\pi}{2^k} \right)^{\frac{1}{20}} + 2n^{-\frac{1}{10}} \right\} \cap \left\{ D(S_{r+1}) \leq n^{-\frac{1}{10}} \forall S_{r+1} \subset [0, 2\pi] : |S_{r+1}| \leq \frac{C_0}{n} \right\} \right)
\geq 1 - \left( C_6 2^k e^{-C_2 n^{\frac{2}{5}}} + \frac{C_7}{(2^k)^{\frac{3}{2}}} \right). \tag{129}
\]
Note that
LHS of (129) \leq \mathbb{P}_x \left( |\zeta_n(s) - \zeta_n(t)| \leq C_8 \left( \frac{1}{2^k_0} \right)^{\frac{1}{20}} + C_9 k_r n^{-\frac{1}{5}}, \forall s, t \in [0, 2\pi] : |t - s| \leq \delta \right).

Therefore, combining with (125), we have
\[
\mathbb{P}_g \left( |\zeta_n(s) - \zeta_n(t)| \leq c_1 |s - t|^{\frac{1}{20}} + c_2 n^{-\frac{1}{10} \log n}, \forall s, t \in [0, 2\pi] : |t - s| \leq \delta \right) \geq 1 - (c_3 e^{-c_4 n^{\frac{2}{5}}} + c_5 \delta^{\frac{1}{10}}).
\]

\section{Proof of Corollary 12}

\textit{Proof.} Taking the Taylor expansion, we have for 0 \leq \delta \leq 1,
\[
\sqrt{n} \sum_{j=0}^{n-1} g(\theta_j) = \sqrt{n} \sum_{j=0}^{n-1} g \left( \frac{2\pi j}{n} + \psi \right) + \sqrt{n} \sum_{j=0}^{n-1} g' \left( \frac{2\pi j}{n} + \psi + \delta \frac{x_j}{n^2} \right) \frac{x_j}{n^2}. \tag{130}
\]
\[
\sum_{k=-\infty}^{\infty} k^{\frac{3}{2}} |c_k| < \infty \text{ implies that } g' \in C^{\frac{3}{2}}(S^1). \text{ Then there exists some constant } A \text{ such that } |g'(x) - g'(y)| \leq A |x - y|^{\frac{3}{2}}. \text{ Since } \text{Var}(x_j) \sim n, \text{ with high probability, } x_j \leq n^{\frac{1}{2} + \gamma} (\gamma > 0). \text{ Then we have}
\[
\left| g' \left( \frac{2\pi j}{n} + \psi + \delta \frac{x_j}{n^2} \right) - g' \left( \frac{2\pi j}{n} + \psi \right) \right| \leq A\delta^{\frac{3}{2}} \frac{x_j^{\frac{3}{2}}}{n} \leq A\delta^{\frac{3}{2}} \left( \frac{n^{\frac{3}{2} + \gamma}}{n} \right)^{\frac{3}{2}} = O(n^{-\frac{3}{2} + \frac{3}{2} \gamma}).
\]

Thus,
\[
\left| \sqrt{n} \sum_{j=0}^{n-1} g'(\frac{2\pi j}{n} + \psi) \frac{x_j}{n^2} - \sqrt{n} \sum_{j=0}^{n-1} g'(\frac{2\pi j}{n} + \psi) \frac{x_j}{n^2} \right| = O(n^{-\frac{1}{2} + \frac{3}{2}}). \tag{131}
\]
\(\gamma\) can be chosen sufficiently small such that the RHS of (131) is \(o_n(1)\).
Therefore, with high probability, we have
\[
\sqrt{n} \sum_{j=0}^{n-1} g(\theta_j) = \sqrt{n} \sum_{j=0}^{n-1} g\left(\frac{2\pi j}{n} + \psi\right) + \sqrt{n} \sum_{j=0}^{n-1} g'\left(\frac{2\pi j}{n} + \psi\right) \frac{x_j}{n^2} + o_n(1). \tag{132}
\]
In addition, \(g(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}\) where \(c_k\) are complex Fourier coefficients. The first term of (132) is almost deterministic and can be written as
\[
\sqrt{n} \sum_{j=0}^{n-1} g\left(\frac{2\pi j}{n} + \psi\right) = \sqrt{n} \sum_{k=\pm n, \pm 2n, \ldots} n c_k + n^\frac{1}{2} \frac{1}{2\pi} \int_0^{2\pi} g(x + \psi) dx. \tag{133}
\]
Since \(g\) is periodic and \(\sum_{k=-\infty}^{\infty} k^\frac{3}{2} |c_k| < \infty\), we have
\[
\sqrt{n} \sum_{j=0}^{n-1} g\left(\frac{2\pi j}{n} + \psi\right) = n^\frac{1}{2} \frac{1}{2\pi} \int_0^{2\pi} g(x) dx + o_n(1) = n^\frac{1}{2} c_0 + o_n(1). \tag{134}
\]
For the second term of (132), denote \(g_n(x) = \exp\left(it \sqrt{n} \sum_{j=0}^{n-1} g'(\alpha_j) \frac{x_j}{n^2}\right)\), where \(\alpha_j = \frac{2\pi j}{n} + \psi\).
Since \(g_n(x)\) is a bounded function, then
\[
\mathbb{E}_x \left(g_n(x)\right) \leq \frac{1}{Z_n} \int_{\Omega'} g_n(x) e^{G(x) + o_n(1)} dx + \mathbb{P}_x(\Omega^c) \leq \frac{1 + o_n(1)}{1 - o_n(1)} \frac{1}{Z_g} \int_{\Gamma} g_n(x) e^{G(x)} dx + \mathbb{P}_x(\Omega^c)
\]
\[
\leq \mathbb{E}_g \left(g_n(x)\right) (1 + o_n(1)).
\]
Here \(\mathbb{E}_g\) indicates taking expectation under Gaussian distribution of (36). Similarly, we can show that
\[
\mathbb{E}_x \left(g_n(x)\right) \geq \frac{1 - o_n(1)}{1 + o_n(1)} \frac{1}{Z_g} \int_{\Omega} g_n(x) e^{G(x)} dx = \frac{1 - o_n(1)}{1 + o_n(1)} \frac{1}{Z_g} \int_{\Gamma} g_n(x) e^{G(x)} dx - \frac{1 - o_n(1)}{1 + o_n(1)} \mathbb{P}_g(\Omega^c)
\]
\[
\geq \mathbb{E}_g \left(g_n(x)\right) (1 - o_n(1)).
\]
Therefore, we obtain that
\[
\mathbb{E}_x \exp \left(it \sqrt{n} \sum_{j=0}^{n-1} g'(\alpha_j) \frac{x_j}{n^2}\right) = \mathbb{E}_g \exp \left(it \sqrt{n} \sum_{j=0}^{n-1} g'(\alpha_j) \frac{x_j}{n^2}\right) (1 + o_n(1)).
\]
By (132) and (134), we can conclude that \(\left(\sqrt{n} \sum_{j=0}^{n-1} g(\theta_j) - n^\frac{1}{2} c_0\right) / \text{Var}_g \left(\sqrt{n} \sum_{j=0}^{n-1} g'(\alpha_j) \frac{x_j}{n^2}\right)\) converges in distribution to standard normal distribution. Finally, we only need to compute the variance.
Recall (113), then
\[
\sqrt{n} \sum_{j=0}^{n-1} g'(\alpha_j) \frac{x_j}{n^2} = \frac{1}{n} \sum_{j=0}^{n-1} g'(\alpha_j) \left( \sum_{k=1}^{n-1} \left( \frac{\cos \frac{2\pi jk}{n}}{k} p_k - \frac{\sin \frac{2\pi jk}{n}}{k} q_k \right) + e_n \right)
\]
\[
= \sum_{k=1}^{n-1} \frac{1}{k} \left( \frac{1}{n} \sum_{j=0}^{n-1} g'(\alpha_j) \cos \frac{2\pi jk}{n} \right) p_k - \sum_{k=1}^{n-1} \frac{1}{k} \left( \frac{1}{n} \sum_{j=0}^{n-1} g'(\alpha_j) \sin \frac{2\pi jk}{n} \right) q_k + r_n,
\]
where \( r_n \) is a random error term with \( \text{Var}(r_n) = O\left(\frac{1}{n}\right) \). Note that \( g' \in C^4(\mathbb{S}^1) \), then
\[
g'(x) = \sum_{i=-\infty}^{\infty} ilc_1 e^{ix}.
\]
Thus,
\[
\frac{1}{n} \sum_{j=0}^{n-1} g'(\alpha_j) \cos \frac{2\pi jk}{n} = O \left( \sum_{l=\pm \frac{1}{2}, \pm n, \ldots} l c_l \right) + \frac{1}{2\pi} \int_{0}^{2\pi} g'(x) \cos kxdx
\]
\[
= -\frac{k}{2\pi} \int_{0}^{2\pi} g(x) \sin kxdx + o_n(1) = -\frac{k}{2} b_k + o_n(1).
\]
(136)
Similarly,
\[
\frac{1}{n} \sum_{j=0}^{n-1} g'(\alpha_j) \cos \frac{2\pi jk}{n} = \frac{k}{2} a_k + o_n(1),
\]
(137)
where \( a_k = \frac{1}{\pi} \int_{0}^{2\pi} g(x) \cos kxdx \) and \( b_k = \frac{1}{\pi} \int_{0}^{2\pi} g(x) \sin kxdx \). Thus,
\[
\text{Var}_g \left( \sqrt{n} \sum_{j=0}^{n-1} g'(\alpha_j) \frac{x_j}{n^2} \right) = \frac{1}{4} \sum_{k=1}^{\infty} (a_k^2 + b_k^2) + o_n(1) = \sum_{k=1}^{\infty} |c_k|^2 + o_n(1).
\]
We have finished the proof of Corollary 12.

\[\square\]

9 Appendix

In this section, we will give the proof of the formulas in Lemma 4. We start with the proof of the formula (13). Let \( z_k = e^{i\frac{2\pi k}{n}} \) (0 \( \leq k \) \( \leq n - 1 \)). Then,
\[
\sum_{k=1}^{n-1} \frac{1}{\sin^2 \left( \frac{\pi k}{n} \right)} = \sum_{k=1}^{n-1} \left( \frac{2i}{e^{i\frac{2\pi k}{n}} - e^{-i\frac{2\pi k}{n}}} \right)^2 = -4 \sum_{k=1}^{n-1} \frac{z_k}{(z_k - 1)^2}.
\]
Let
\[
F_1(z) = \frac{1}{(z - 1)^2(z^n - 1)} = \frac{1}{(z - 1)^3(1 + z + \cdots + z^{n-1})},
\]
which is a holomorphic function on \( \mathbb{C} \) except \( z = z_k \) (0 \( \leq k \) \( \leq n - 1 \)). Note that \( z = z_k \) (1 \( \leq k \) \( \leq n - 1 \)) are simple poles, and
\[
z^n - 1 = \prod_{j=0}^{n-1} (z - z_j), \quad \prod_{j:j \neq k} (z_k - z_j) = nz_k^{-1}.
\]
30
Combining it with the Residue Theorem, we have

\[ \sum_{k=1}^{n-1} \frac{1}{\sin^2 \left( \frac{\pi k}{n} \right)} = -4n \sum_{k=1}^{n-1} \text{Res}_{z=z_k} F_1(z) = 4n \text{Res}_{z=1} F_1(z). \]

Since \( z = 1 \) is a pole of order 3, we obtain that

\[ \text{Res}_{z=1} F_1(z) = \frac{1}{2!} \left. \left( \frac{1}{1 + z + \cdots + z^{n-1}} \right)'' \right|_{z=1} = \frac{n^2 - 1}{12n}. \]

Thus we have proved the formula (15).

Similarly, we can prove (16). Let

\[ F_2(z) = \frac{z}{(z - 1)^4(z^n - 1)} = \frac{1}{(z - 1)^3(z^n - 1)} + \frac{1}{(z - 1)^4(z^n - 1)}. \]

Then

\[ \sum_{k=1}^{n-1} \frac{1}{\sin^4 \left( \frac{1}{n} \sum \right)} = 16 \sum_{k=1}^{n-1} \frac{z_k^2}{(z_k - 1)^4} = 16n \sum_{k=1}^{n-1} \text{Res}_{z=z_k} F_2(z) = -16n \text{Res}_{z=1} F_2(z) = \frac{(n^2 - 1)(n^2 + 11)}{45}. \]

For the formula (17), let

\[ F_3(z) = \frac{(z^m - 1)^2}{z^m(z - 1)^2(z^n - 1)} = \frac{(1 + z + \cdots + z^{m-1})^2}{z^m(z - 1)(1 + z + \cdots + z^{n-1})}. \]

Since \( F_3 \) is holomorphic on \( \mathbb{C} \) except \( z = 0 \) and \( z = z_k \) (\( 0 \leq k \leq n - 1 \)). Similarly to the previous computations, we have

\[ \sum_{k=1}^{n-1} \frac{\sin^2 \left( m \frac{\pi k}{n} \right)}{\sin^2 \left( \frac{\pi k}{n} \right)} = \sum_{k=1}^{n-1} \left( \frac{z_k^m - 1}{z_k - 1} \right)^2 z_k^{1-m} = -n \left( \text{Res}_{z=1} F_3(z) + \text{Res}_{z=0} F_3(z) \right). \]

Since \( z = 1 \) is simple pole,

\[ \text{Res}_{z=1} F_3(z) = \left. \frac{(1 + z + \cdots + z^{m-1})^2}{z^m(1 + z + \cdots + z^{n-1})} \right|_{z=1} = \frac{m^2}{n}. \]

Also note that

\[ F_3(z) = -z^{-m} (z^{2m} - 2z^m + 1) \left( \sum_{k=0}^{\infty} z^k \right)^2 \left( \sum_{k=0}^{\infty} z^{kn} \right). \]

Since \( 1 \leq m \leq n - 1 \),

\[ \text{Res}_{z=0} F_3(z) = -\text{Res}_{z=0} \frac{1}{(z - 1)^2 z^m} = -m. \]
Thus, we finish the proof of (17).

Finally, for (18), let

\[ F_4(z) = \frac{(z^m - 1)^2}{(z - 1)^4(z^n - 1)} \cdot z^{1-m}. \]

Similar with the proof of the formula (17), we obtain that

\[
\sum_{k=1}^{n-1} \frac{\sin^2 \left( \frac{m\pi k}{n} \right)}{\sin^4 \left( \frac{\pi k}{n} \right)} = -4 \sum_{k=1}^{n-1} \frac{(z_k^m - 1)^2}{(z_k - 1)^4} z_k^{2-m} = 4n \left( \text{Res}_{z=1} F_4(z) + \text{Res}_{z=0} F_4(z) \right) = \frac{m^2(n - m)^2}{3} + \frac{2}{3} m(n - m).
\]
References

[1] G. W. Anderson, A. Guionnet, O. Zeitouni. An Introduction to Random Matrices. Cambridge University Press, 2009.

[2] T. H. Baker, P. J. Forrester. Finite-\(N\) fluctuations formulas for random matrices. J. Stat. Phys., Vol. 88, Nos. 5/6, 1371-1386, 1997.

[3] M. V. Berry, J. P. Keating. The Riemann Zeros and Eigenvalues Asymptotics. SIAM Review, Vol. 41, No. 2, pp 236-266, 1999.

[4] P. Billingsley. Convergence of Probability Measures. Wiley, New York, 1968.

[5] E. B. Bogomolny, J. P. Keating. Quantum chaos, random matrix theory, and the Riemann \(\zeta\)-function. Séminaire Poincaré XIV, 115 153, 2010.

[6] P. Bourgade, J. P. Keating. Quantum chaos, random matrix theory, and the Riemann \(\zeta\)-function. Séminaire Poincaré XIV, 115 153, 2010.

[7] P. Diaconis, M. Shahshahani. On the eigenvalues of random matrices. J. Appl. Prob., 31A, 49, 1994.

[8] P. Diaconis, S. N. Evans. Linear functional of eigenvalues of random matrices. Transactions of the American Mathematical Society, Vol. 353, No. 7, pp. 2615-2633, 2001.

[9] I. Dumitriu, A. Edelman. Matrix models for beta ensembles. Journal of Mathematical Physics, Vol. 43, No. 11, 2002.

[10] F. J. Dyson. Correlations between the eigenvalues of a random matrix. Comm. Math. Phys., 19, 235, 1970.

[11] P. Forrester. Log-Gases and Random Matrices. Princeton University Press, 2010.

[12] C. P. Hughes, J. P. Keating, and N. O'Connell. On the characteristic polynomial of a random unitary matrix. Comm. Math. Phys., 220(2):429451, 2001.

[13] K. Johansson. On Szego’s asymptotic formula for Toeplitz determinants and generalizations. Bull. Sc. math., Serie 2, 112, pp.257-304, 1988.

[14] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. Duke Mathematical Journal., Vol. 91, No. 1, 1998.

[15] K. Johansson. On random matrices from the compact classical groups. Annals of Mathematics, Second Series, Vol. 145, No. 3, pp. 519-545, 1997.

[16] J. P. Keating, and N. Snaith. Random Matrix Theory and (1/2+it). Communications in Mathematical Physics., Volume 214, Issue 1, pp 5789, 2000.

[17] G. Lambert, D. Ostrovsky, N. Simm. Subcritical multiplicative chaos for regularized counting statistics from random matrix theory. arXiv:1612.02367, 2016.
[18] M. Mehta. *Random Matrices*. Pure and Applied Mathematics, Vol.142, 3rd version, Academic Press, 2004.

[19] H. L. Montgomery. The pair correlations of zeroes of the zeta-function. *Proc. Sympos. Pure Math.*, 24, Amer. Math. Soc, 1973.

[20] A.M. Odlyzko. The $10^{20}$-th zero of the Riemann zeta function and 175 million of its neighbors. *Preprint*, 1992.

[21] Z. Rudnick, P. Sarnak. Zeros of principal L-functions and random matrix theory. *Duke Mathematical Journal*, 81, pp. 269-322, 1996.

[22] A. Soshnikov. Level spacings Distribution for Large Random Matrices: Gaussian Fluctuations. *Annals of Mathematics*, vol. 148, pp.573-617, 1998.

[23] A. Soshnikov. Central Limit Theorem for Linear Statistics in Classical Compact Groups and Related Combinatorial Identities. *Annals of Probability*, v. 28, pp. 1353-1370, 2000.

[24] A. Soshnikov. Determinantal Random Point Fields. *Russian Mathematical Surveys*, vol.55, No.5, pp.923-975, 2000.

[25] T. Tao. *Topics in Random Matrix Theory*. Graduate Studies in Mathematics, Vol.132, American Mathematical Society, 2012.