Large Deviation Principles via Spherical Integrals

Serban Belinschi  
CNRS-Université de Toulouse  
Serban.Belinschi@math.univ-toulouse.fr

Alice Guionnet  
CNRS-ENS Lyon  
aguionne@ens-lyon.fr

Jiaoyang Huang  
IAS  
jiaoyang@ias.edu

Abstract

In this article, we develop a framework to study the large deviation principle for matrix models and their quantized versions, by tilting the measures using the limits of spherical integrals obtained in [?, ?]. As examples, we obtain

1. the large deviation principle for the empirical distribution of the diagonal entries of $UB_N U^*$, for a sequence of $N \times N$ diagonal matrices $B_N$ and unitary Haar distributed matrices $U$,
2. a large deviation upper bound for the empirical eigenvalue distribution of $A_N + UB_N U^*$, for two sequences of $N \times N$ diagonal matrices $A_N, B_N$, and their complementary lower bounds at “good” probability distributions,
3. the large deviation principle for the Kostka number $K_{\lambda_N \eta_N}$, for two sequences of partitions $\lambda_N, \eta_N$ with at most $N$ rows,
4. a large deviation upper bound for the Littlewood-Richardson coefficients $c_{\lambda_N \eta_N \kappa_N}$, for three sequences of partitions $\lambda_N, \eta_N, \kappa_N$ with at most $N$ rows, and their complementary lower bounds at “good” probability distributions.

1 Introduction

During the last thirty years, random matrix theory has grown into a sophisticated branch of mathematics, interacting profoundly with physics and other areas of mathematics such as statistics, probability and operator algebra. Following the initial breakthroughs by Wishart [?] and Wigner [?], the convergence of the empirical measure of the eigenvalues and the extreme eigenvalues could be established for many models of random matrices. The fluctuations of the eigenvalues both in the local and global scale were investigated. However, the understanding of the probabilities that the spectrum has an unlikely behavior, as measured by large deviations principles, is still very scarce, even at a conjectural level.

For instance, let us consider Wigner matrices, that are self-adjoint matrices with independent (modulo the symmetry constraint) centered entries with covariance given by the inverse of the dimension. It has been shown that the empirical measure of such random matrices converges towards a non-random limit and that the extreme eigenvalues “stick” to the bulk in the sense that they converge towards the boundary...
of the support of this limiting distribution as soon as their fourth moment is finite \([?, ?, ?]\). Fluctuations around this limit could be studied. It was shown that generically, the empirical eigenvalue distribution has small fluctuations. The central limit theorem for the empirical measure holds without the celebrated normalization by the square root of the dimension required for sums of independent random variables \([?, ?, ?, ?]\). Local fluctuations of the eigenvalues were first proven for Gaussian ensembles \([?, ?]\) and more recently generalized to Wigner matrices \([?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?]\). Large deviation principles, which allow to estimate the probability to deviate from the almost sure asymptotic behavior, are still much less understood. They were first derived for random matrices with Gaussian entries thanks to the explicit joint law of the eigenvalues \([?, ?, ?]\). Large deviation principles were then obtained for random matrices with entries heavier than Gaussian \([?, ?, ?]\), by using the fact that the matrix can more easily create deviations by having a few large entries. The case of sub-Gaussian entries thus stayed open until recently. F. Augeri, J. Husson and one of the author tackled the large deviations for the largest eigenvalue \([?, ?, ?]\). They showed that the large deviation rate function is the same as in the Gaussian case if the entries have a Laplace transform which can be evaluated. It amounts to tilt the law of the matrix in a random direction to make the desired deviation more likely. The tilted measure eventually be evaluated. It amounts to tilt the law of the matrix in a random direction to make the typical events under the tilted measure for an appropriate choice of \(\theta\), and can be studied through the following relation

\[
P(X_N \in x + dx) = P_\theta(X_N \in x + dx) \exp \left\{ -\langle \theta, x \rangle + \log \mathbb{E}[\exp\{\langle \theta, X_N \rangle\}] \right\},
\]

where the asymptotics of the last term on the right hand side can eventually be computed (for instance if it is the sum of independent i.i.d random vectors). Unfortunately, if \(X_N\) is a sequence of random matrices and one considers the law of its empirical measure or its largest eigenvalue \(\lambda_{X_N}\), it is not clear how to make such a computation. Because in general, either the tilt is a function of the largest eigenvalue and then we do not know how to compute its Laplace transform, or it is for instance a tilt on each individual entry but then we do not know how it is related with the largest eigenvalue. The idea in \([?, ?]\) to study the large deviation for the largest eigenvalue was to tilt the probability measure by spherical integrals, because we know \([?]\) it becomes a function of the largest eigenvalue when the dimension goes to infinity, but also it produces independent (but random) tilt on each entry with a Laplace transform which can eventually be evaluated. It amounts to tilt the law of the matrix in a random direction to make the desired deviation more likely. The tilted measure \(P_\theta\) is given by

\[
P_\theta(\lambda_{X_N} \in \lambda + d\lambda) = \frac{1}{\int I_N(\theta, X_N)d\mathbb{P}(X_N)} \int \mathbb{1}_{\{\lambda_{X_N} \in \lambda + d\lambda\}} I_N(\theta, X_N)d\mathbb{P}(X_N),
\]

where \(I_N\) is the one dimensional spherical integral

\[
I_N(\theta, X_N) = \mathbb{E}_u[e^{\theta N(u, X_N u)}],
\]

and \(u\) follows the uniform law on the unit sphere in \(\mathbb{R}^N\). It was proven that given a real number \(\lambda\), one can often (for instance when \(\lambda\) is sufficiently large) find a tilt \(\theta_\lambda\) so that under \(P_{\theta_\lambda}\), the largest eigenvalue \(\lambda_{X_N}\) is close to \(\lambda\), whereas the cost of this tilt can be computed thanks to \([?]\) and the expectation of the spherical integral over \(X_N\) can be estimated \([?, ?, ?]\). Yet, deriving the large deviations for the distribution of the empirical eigenvalue density of Wigner matrices is still an open problem. Even though the large
Schur-Horn theorem and Horn’s problem are complicated and hard to analyze directly. The classical Schur-Horn theorem [?] states that the diagonal entries of $UBN^*$ as a vector is in the permutation polytope generated by $(b_1, b_2, \cdots, b_N)$. A more challenging problem of the same flavor is Horn’s problem: Given two $N \times N$ diagonal matrices $A_N, B_N$ with eigenvalues $a_1 \geq a_2 \geq \cdots \geq a_N$ and $b_1 \geq b_2, \cdots \geq b_N$, what can be said about the eigenvalues of $A_N + UB^N$? Besides the trivial relation $\text{Tr} A_N + \text{Tr} B_N = \text{Tr}(A_N + UB_N^*)$, Horn [?] had conjectured the form of a set of necessary and sufficient inequalities to be satisfied for the eigenvalues of $A_N + UB_N^*$. After contributions by several authors, see in particular [?], these conjectures were proven by Knutson and Tao [?, ?]. See [?] for a nice survey of this problem. This result however do not say anything about the probability that the spectrum of $A_N + UB_N^*$ has some given distribution.

Let $U$ follow the Haar measure on the orthogonal group when $\beta = 1$ and on the unitary group when $\beta = 2$. The randomized Schur-Horn theorem and Horn’s problem ask the distribution of the diagonal entries of $UB_N^*$ and the empirical eigenvalue distribution of $A_N + UB_N^*$:

$$
\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{(UB_N^*)_{ii}}, \quad \bar{\mu}_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{(A_N + UB_N^*)_{ii}}. \tag{1.1}
$$

The randomized Schur-Horn theorem is equivalent to computing the Duistermaat–Heckman measures of the coadjoint orbit $UB_N^*$. More generally, they are defined using the push-forward of the Liouville measure on a symplectic manifold along the moment map [?, ?, ?, ?, ?]. The randomized version of Horn’s problem has been first studied in [?] for $N = 3$. Recent years, there has seen a surge of interest in this problem: [?, ?, ?, ?, ?] for general dimension; [?] for an extension to other Lie groups; [?] for the multiplicative version of randomized Horn’s problem. However, those distribution densities of randomized Schur-Horn theorem and Horn’s problem are complicated and hard to analyze directly.

For the large $N$ limit, we will assume that the spectral measures of $A_N, B_N$ converge towards $\mu_A$ and $\mu_B$ respectively. It is well known that $\mu_N$ converges weakly almost surely towards a delta mass at $\int x d\mu_B(x)$. On the other hand, Voiculescu [?, ?] proved that $\bar{\mu}_N$ converges towards the free convolution of $\mu_A$ and $\mu_B$. For the second model, fluctuations of the empirical measure where studied in [?] and large deviations for the largest eigenvalue in [?]. To study the large deviations for $\mu_N$ and $\bar{\mu}_N$ from this asymptotic behavior, we propose to tilt the original measure using the spherical integrals, which are Fourier transforms over Unitary/Orthogonal groups. We recall that given two sequences $A_N, B_N$ of deterministic self-adjoint matrices, the $N$ dimensional spherical integral is defined as

$$
I_N(A_N, B_N) = \int e^{\frac{2i}{N} \text{Tr}(A_N UB_N^*)} dU,
$$

where $U$ follows the Haar measure on the unitary group (resp. orthogonal group) when $\beta = 2$ (resp. $\beta = 1$). If the spectral measures $\mu_{A_N}, \mu_{B_N}$ of $A_N, B_N$ converge weakly towards $\mu_A$ and $\mu_B$ respectively, under mild assumptions, it was proven in [?, ?, ?, ?, ?], see also [?, ?], that the spherical integral converges

$$
I(\mu_A, \mu_B) = \lim_{N \to \infty} \frac{1}{\beta N^2} \log \int e^{\frac{2i}{N} \text{Tr}(A_N UB_N^*)} dU. \tag{1.2}
$$

The density of the spectrum of the eigenvalues of $A + UB^N$ and of the diagonal entries of $UB^N$ was expressed in terms of spherical integrals in [?, ?, ?, ?] by using Fourier analysis. This however requires to take spherical integrals evaluated at complex entries and use Fourier analysis. It is hard to see how to use such an approach to obtain asymptotics, as spherical integrals need to be evaluated at complex matrices.
for which no asymptotics are known. Our approach by tilting the original measure using the spherical integrals, requires only the derivatives for the limiting spherical integral, which we derive in Section 2. As a consequence, it gives anew understandings of the Schur-Horn theorem and Horn’s problem, as well as the evaluation of the asymptotics of Kostka numbers and Littlewood-Richardson coefficients.

As the first application of our spherical integral approach, we study large deviations of the randomized Schur-Horn theorem and Horn’s problem. We obtain the large deviation principle for the empirical distribution of the diagonal entries of $UB_N U^*$ and the large deviation upper bound for the empirical eigenvalue distribution of $A_N + U B_N U^*$.

The quantized versions of the Horn-Schur theorem and Horn’s problem ask when a Kostka number or a Littlewood-Richardson coefficient is non-zero. We denote by $\mathcal{Y}_N$ the set of partitions with at most $N$ rows. We recall that given a partition $\lambda_N \in \mathcal{Y}_N$, the Kostka numbers $K_{\lambda_N \eta_N}$ are the coefficients that arise when one expresses the Schur symmetric polynomial $S_{\lambda_N}$ as a linear combination of monomial symmetric functions $m_{\eta_N}$:

$$S_{\lambda_N}(x_1, x_2, \ldots, x_N) = \sum_{\mu_N \in \mathcal{Y}_N} K_{\lambda_N \eta_N} m_{\eta_N}(x_1, x_2, \ldots, x_N).$$

(1.3)

It is known that the Kostka number $K_{\lambda_N \mu_N}$ is positive if and only if $\lambda_N$ and $\eta_N$ are of the same size, and $\lambda_N$ is larger than $\eta_N$ in the dominance order:

$$\lambda_1 + \lambda_2 + \cdots + \lambda_i \geq \eta_1 + \eta_2 + \cdots + \eta_i, \quad 1 \leq i \leq N.$$

Given a pair of partitions $\lambda_N, \eta_N \in \mathcal{Y}_N$, the Littlewood-Richardson coefficients $c_{\lambda_N \eta_N}^{\kappa_N}$ are the coefficients that arise when one expresses the product of the Schur symmetric polynomials $S_{\lambda_N} S_{\eta_N}$ as a linear combination of Schur symmetric polynomials $S_{\kappa_N}$:

$$S_{\lambda_N}(x_1, x_2, \ldots, x_N) S_{\eta_N}(x_1, x_2, \ldots, x_N) = \sum_{\kappa_N \in \mathcal{Y}_N} c_{\lambda_N \eta_N}^{\kappa_N} S_{\kappa_N}(x_1, x_2, \ldots, x_N).$$

(1.4)

Horn’s problem is equivalent to deciding the conditions on the triples $(\lambda_N, \eta_N, \kappa_N)$, such that the Littlewood-Richardson coefficient $c_{\lambda_N \eta_N}^{\kappa_N}$ is positive. This result has previously been obtained by Klyachko using geometric invariant theory [?].

As the second application of our spherical integral approach, we derive the large deviation principle of Kostka numbers and the large deviation upper bound for the Littlewood-Richardson coefficients. The asymptotics of certain extreme Kostka numbers and the Littlewood-Richardson coefficients were derived in [?], [?], [?], [?].

Acknowledgement. The authors heartily thank Denis Serre for his help to guess the derivative of spherical integrals at the early stage of this work. H.J. wants to thank Ofer Zeitouni for enlightening discussions. A.G. wants to thank Vadim Gorin for the references on Duistermaat–Heckman measures, and the support from the Institute for Advanced Study. AG was partially supported by the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program "Investissements d’Avenir" (ANR-11-IDEX- 0007) operated by the French National Research Agency (ANR).

1.1 Main Results

Before stating our main results, we need to introduce some notations and definitions. In this paper, we fix a large constant $\mathcal{R} > 0$. We denote by $\mathcal{M}(\mathbb{R})$ (resp. $\mathcal{M}([-\mathcal{R}, \mathcal{R}])$) the space of probability measures on $\mathbb{R}$ (resp. $[-\mathcal{R}, \mathcal{R}]$), with bounded first moment:

$$\mathcal{M} = \{\nu : \text{supp}\nu \subset \mathbb{R}, \nu(|x|) < \infty\}, \quad \mathcal{M}([-\mathcal{R}, \mathcal{R}]) = \{\nu : \text{supp}\nu \subset [-\mathcal{R}, \mathcal{R}], \nu(|x|) < \infty\}.$$
We equip $\mathcal{M}$ and $\mathcal{M}([-\mathbb{R}, \mathbb{R}])$, with the weak topology. The weak topology is compatible with the following distance

$$d(\mu, \nu) := \sup_{\|f\|_\infty \leq 1} \left| \int f(x) d\mu - \int f(x) d\nu \right|, \quad (1.5)$$

where $\|f\|_\infty$ is the Lipschitz constant of $f$. For probability measures with bounded first moment, a more natural distance is the Wasserstein distance:

$$d_W(\mu, \nu) := \sup_{\|f\|_\infty \leq 1} \left| \int f(x) d\mu - \int f(x) d\nu \right|. \quad (1.6)$$

We remark the convergence in the Wasserstein distance $d_W$ is equivalent to weak convergence and the convergence of the first moment. On the set of uniformly compactly supported measures, i.e. $\mathcal{M}([-\mathbb{R}, \mathbb{R}])$, convergences in the Wasserstein distance and weak convergence are equivalent. We also denote by $\mathcal{M}^b([0, \mathbb{R}])$ the set of probability measures on $[0, \mathbb{R}]$ with density bounded by 1, i.e. the set of probabilities $d\nu = \rho(x) dx$, such that $\rho(x)$ is supported on $[0, \mathbb{R}]$ and $\rho(x) \leq 1$.

Given a probability measure $\mu$, let $T_\mu : (0, 1) \to (-\infty, \infty)$ be the right continuous increasing function, such that $\mu$ is the push-forward of the uniform distribution on $(0, 1)$ by $T_\mu$. In other words, for all bounded continuous function $f$, we have

$$\int_0^1 f(T_\mu(x)) dx = \int f(x) d\mu(x). \quad (1.7)$$

More explicitly, $T_\mu$ is the functional inverse of the cumulative density function $F_\mu$ of $\mu$. With this notation, we can rewrite the Wasserstein distance (1.6) as

$$d_W(\mu, \nu) := \sup_{\|f\|_\infty \leq 1} \left| \int f(x) d\mu - \int f(x) d\nu \right| = \int \left| T_\mu(x) - T_\nu(x) \right| dx. \quad (1.8)$$

**Theorem 1.1.** Let $B_N$ be a sequence of deterministic self-adjoint matrices such that the spectral measures $\mu_{BN}$ of $B_N$ converge weakly towards $\mu_B$ as $N \to \infty$. Assume there exists a constant $\mathbb{R} > 0$, such that $\text{supp} \mu_{BN} \subset [-\mathbb{R}, \mathbb{R}]$.

1. Let $\mu \in \mathcal{M}([-\mathbb{R}, \mathbb{R}])$, $\nu \in \mathcal{M}$ and set

$$H^D_\mu(\nu) = \frac{1}{2} \int T_\nu(x) T_\mu(x) dx - I(\nu, \mu_B), \quad (1.9)$$

where $I(\cdot, \cdot)$ is defined in (1.2) and $T_\mu, T_\nu$ are as defined in (1.7). Then, the functional $I^D(\cdot)$

$$I^D(\mu) := \sup_{\nu \in \mathcal{M}} H^D_\mu(\nu), \quad (1.10)$$

is non-negative, lower semicontinuous on $\mathcal{M}([-\mathbb{R}, \mathbb{R}])$ and vanishes only at the Dirac mass at $\int x d\mu_B$. $I^D(\mu) = +\infty$ unless $\int x d\mu = \int x d\mu_B$ and $\mu$ satisfies the limiting Schur-Horn inequalities:

$$\int_y^1 (T_\mu(x) - T_{\mu_B}(x)) dx \leq 0 \quad \text{for all } y \in [0, 1]. \quad (1.11)$$

2. The distribution of the empirical measure of the diagonal entries of $UB_N U^*$,

$$\mu_N = \frac{1}{N} \sum_{i=1}^N \delta(UB_N U^*)_{ii},$$
satisfies a large deviation principle with good rate function $\mathcal{I}^D$. In other words, for any $\mu \in \mathcal{M}([-\mathcal{R}, \mathcal{R}])$, if $\mathbb{B}_\delta(\mu)$ denotes the open ball $\{\nu \in \mathcal{M}([-\mathcal{R}, \mathcal{R}]) : d(\nu, \mu) < \delta\}$,

$$\lim_{\delta \to 0} \liminf_{N \to \infty} \frac{1}{\beta N^2} \log \mathbb{P}(\mu_N \in \mathbb{B}_\delta(\mu)) = \lim_{\delta \to 0} \limsup_{N \to \infty} \frac{1}{\beta N^2} \log \mathbb{P}(\mu_N \in \mathbb{B}_\delta(\mu)) = -\mathcal{I}^D(\mu).$$ (1.12)

**Remark 1.2.** Since $\mu_{B_N}$ is supported on $[-\mathcal{R}, \mathcal{R}]$, then deterministically we have $|(UB_NU^*)_{ii}| \leq \mathcal{R}$ for all $1 \leq i \leq N$. Therefore, for our study of the large deviation principle of the empirical measure $\mu_N$ in Theorem 1.1, we have restricted ourselves in the set of measures supported on $[-\mathcal{R}, \mathcal{R}]$, which is compact in weak topology. If we do not restrict ourselves in the set of measures supported on $[-\mathcal{R}, \mathcal{R}]$, then the large deviation principle (1.12) holds for any probability $\mu$. We also notice that since $\mathcal{M}([-\mathcal{R}, \mathcal{R}])$ is compact, the above weak large deviation principle is equivalent to a full large deviations principle.

**Theorem 1.3.** Let $A_N, B_N$ be a sequence of deterministic self-adjoint matrices, such that their spectral measures $\mu_{A_N}, \mu_{B_N}$ converge weakly towards $\mu_A, \mu_B$ respectively. Assume there exists a constant $\mathcal{R}$ such that $\text{sup} \mu_{A_N}, \text{sup} \mu_{B_N} \subset [-\mathcal{R}, \mathcal{R}]$. Then

1. Let $\mu \in \mathcal{M}([-2\mathcal{R}, 2\mathcal{R}])$, $\nu \in \mathcal{M}$ and set

$$H^{A+B}_\nu(\nu) = I(\nu, \mu) - I(\nu, \mu_A) - I(\nu, \mu_B),$$ (1.13)

where $I(\cdot, \cdot)$ is defined in (1.2). The functional $\mathcal{I}^{A+B}(\cdot)$

$$\mathcal{I}^{A+B}(\mu) := \sup_{\nu \in \mathcal{M}} H^{A+B}_\nu(\nu)$$ (1.14)

is non-negative and lower semicontinuous on $\mathcal{M}([-2\mathcal{R}, 2\mathcal{R}])$. $\mathcal{I}^{A+B}(\mu) = \infty$ unless $\int x d\mu = \int x d\mu_A + \int x d\mu_B$ and the limiting Ky Fan inequalities hold:

$$\int_y^1 T_{\mu_A}(x) dx + \int_y^1 T_{\mu_B}(x) dx \geq \int_y^1 T_{\mu}(x) dx, \quad \text{for all } y \in [0, 1].$$

2. Let $\mathbb{B}_\delta(\mu)$ denote the open ball $\{\nu \in \mathcal{M}([-2\mathcal{R}, 2\mathcal{R}]) : d(\nu, \mu) < \delta\}$. The empirical eigenvalue distribution $\mu_N$ of $A_N + UB_NU^*$ satisfies the large deviation upper bound

$$\limsup_{\delta \to 0} \limsup_{N \to \infty} \frac{1}{\beta N^2} \log \mathbb{P}(\mu_N \in \mathbb{B}_\delta(\mu)) \leq -\mathcal{I}^{A+B}(\mu),$$ (1.15)

and the complementary lower bound holds if the sup in (1.14) is achieved at a probability measure $\nu$ which is compactly supported and has a non trivial absolutely continuous part in each of its connected components.

**Remark 1.4.** If $\mu_{A_N}$ and $\mu_{B_N}$ are supported on $[-\mathcal{R}, \mathcal{R}]$, then deterministically the spectral measure of $A_N + UB_NU^*$ is supported on $[-2\mathcal{R}, 2\mathcal{R}]$. Therefore, for our study of the large deviation principle of the empirical measure of $A_N + UB_NU^*$, we can restrict ourselves in the set of measures supported on $[-2\mathcal{R}, 2\mathcal{R}]$, which is compact in weak topology.

Note here that for any measure $\nu$ which is compactly supported and has absolute continuous part in each of its connected component, we prove that there exists a unique probability measure $\mu_\nu$ such that the supremum in (1.14) is achieved at $\nu$. In this case, $H^{A+B}_\nu(\nu)$ is finite. Hence, we construct in this article a family of probability measures, the set $\mathcal{H}^{A+B} = \{\mu_\nu : \nu \text{ is compactly supported and has absolutely continuous part in each of its connected components}\}$, which are limit points of the spectrum of $A_N + UB_NU^*$, since the probability that the empirical measure of this matrix is close to $\mu_\nu$ does not vanish (but is exponentially small). The set of measures in $\mathcal{H}^{A+B}$ are measures which have to satisfy

6
in (1.17) converges weakly to then (1.2) implies the following asymptotics of Schur symmetric polynomials: $m \lambda$ produce more notations. Given a partition $\lambda$, let $m_\lambda$ denote the measure

$$m_\lambda = \frac{1}{N} \sum_{i=1}^{N} \delta \left( \frac{\lambda + N - i}{N} \right).$$

(1.17)

To get the Schur symmetric polynomial $S_\lambda$ parametrized by the partition $\lambda$ from the Harish-Chandra-Itzykson-Zuber integral formula (1.16), we take

$$e^{Y_N} = (e^{y_1}, e^{y_2}, \ldots, e^{y_N}),$$

and denote $S_\lambda(e^{Y_N})$ for all $\lambda$, that is, we encode it through the counting measure $m[\lambda_N]$ as

$$S_\lambda(e^{Y_N}) = \frac{\det [e^{iy_i + N - 1}]_{1 \leq i, j \leq N}}{\prod_{1 \leq i < j \leq N} (y_i - y_j)} \int e^{Y_N T U} dU.$$

(1.18)

Let $\lambda_N \in \mathcal{Y}_N$ be a sequence of deterministic partitions, such that its counting measure $m[\lambda_N]$ as defined in (1.17) converges weakly to $m_\lambda$, and the spectral measure $\mu_{Y_N}$ of $Y_N$ converges weakly towards $\mu_Y$, then (1.2) implies the following asymptotics of Schur symmetric polynomials

$$\lim_{N \to \infty} \frac{1}{N^2} \log S_\lambda(e^{Y_N}) = J(\mu_Y, m_\lambda),$$

$$J(\mu_Y, m_\lambda) = 2I(\mu_Y, m_\lambda) + \frac{1}{2} \int \log |x - y| dm_\lambda(x) dm_\lambda(y) + \frac{1}{4} \int \log \left( \frac{e^x - e^y}{x - y} \right) d\mu_Y(x) d\mu_Y(y) + \frac{3}{4}.$$

(1.19)

Theorem 1.5. Let $\lambda_N \in \mathcal{Y}_N$ be a sequence of deterministic partitions, such that its counting measure $m[\lambda_N]$ as defined in (1.17), converges weakly towards $m_\lambda$. Assume there exists a constant $R > 0$, such that $\text{supp } m[\lambda_N] \subset [0, R]$ for all $N \in \mathbb{N}$.

1. Let $\mu \in \mathcal{M}^+(0, R)$, $\nu \in \mathcal{M}$ and set

$$H^\nu_{\mu}(\nu) := \int (T_\mu - x) T_\nu dx - J(\nu, m_\lambda),$$

(1.20)
where the functional $J(\cdot, \cdot)$ has been defined in (1.19) and $T_\nu$, $T_\nu^*$ are as defined in (1.7). The functional $T^K(\cdot)$

$$T^K(\mu) := \sup_{\nu \in \mathcal{M}} H^K_\mu(\nu), \quad (1.21)$$

is lower semicontinuous on $\mathcal{M}^b([0, \Re])$ and achieves its maximum only at the uniform measure $\text{uni}([\int xdm_\lambda - 1/2, \int xdm_\lambda + 1/2])$. $T^K(\mu) = +\infty$ unless $\int xd\mu = \int xdm_\lambda$ and $\mu$ satisfies the following inequalities:

$$\int_y^1 (T_\mu(x) - T_{m_\lambda}(x))dx \leq 0 \quad \text{for all } y \in [0, 1]. \quad (1.22)$$

2. The Kostka numbers $K_{\lambda N} \eta_N$ in (1.3) satisfy, for any $\mu \in \mathcal{M}^b([0, \Re])$,

$$\lim \sup_{\delta \to 0} N^{-2} \log \sup_{m[\eta_N] \in B(\mu)} K_{\lambda N} \eta_N = \lim \inf_{N \to \infty} N^{-2} \log \sup_{m[\eta_N] \in B(\mu)} K_{\lambda N} \eta_N = -T^K(\mu), \quad (1.23)$$

where $B(\mu)$ is the ball $\{\nu \in \mathcal{M}^b([0, \Re]) : d(\nu, \mu) < \delta\}$.

**Remark 1.6.** If $m[\lambda_N]$ are supported on $[0, \Re]$, and $K_{\lambda N} \eta_N \neq 0$, then deterministically $m[\eta_N]$ is supported on $[0, \Re]$. Moreover, from our construction of $m[\eta_N]$ as in (1.17), the limit of $m[\eta_N]$ necessarily has a density bounded by 1. Therefore in Theorem 1.7, we have restricted ourselves in the set of measures supported on $[0, \Re]$ with density bounded by 1.

We can also derive the following asymptotic formulas for the Littlewood-Richardson coefficients.

**Theorem 1.7.** Let $\lambda_N, \eta_N \in \mathcal{Y}_N$ be two sequences of deterministic partitions, such that their counting measures $m[\lambda_N], m[\eta_N]$, as defined in (1.17), converge weakly towards $m_\lambda, m_\eta$ respectively. Assume there exists a constant $A > 0$, such that $\sup m[\lambda_N], \sup m[\eta_N] \subset [0, \Re]$. Then

1. Let $\mu \in \mathcal{M}^b([0, 2\Re]), \nu \in \mathcal{M}$ and set

$$H^{LR}_\mu(\nu) = J(\nu, \mu) - J(\nu, m_\lambda) - J(\nu, m_\eta),$$

where the functional $J(\cdot, \cdot)$ has been defined in (1.19). The functional $T^{LR}(\cdot)$

$$T^{LR}(\mu) := \sup_{\nu \in \mathcal{M}} H^{LR}_\mu(\nu) \quad (1.24)$$

is lower semicontinuous on $\mathcal{M}^b([0, 2\Re])$. $T^{LR}(\mu) = +\infty$ unless $\int xdm_\mu = \int xdm_\lambda + \int xdm_\eta$ and the following inequalities hold:

$$\int_y^1 T_{m_\lambda}(x)dx + \int_y^1 T_{m_\eta}(x)dx \geq \int_y^1 T_\mu(x)dx, \quad \text{for all } y \in [0, 1].$$

2. Let $B(\mu)$ denote the ball $\{\nu \in \mathcal{M}^b([0, 2\Re]) : d(\nu, \mu) < \delta\}$. The Littlewood-Richardson coefficients $c^{KN}_{\lambda N} \eta_N (1.4)$ are asymptotically bounded above as follows

$$\lim \sup_{\delta \to 0} N^{-2} \log \sup_{m[\lambda_N] \in B(\mu)} c^{KN}_{\lambda N} \eta_N \leq -T^{LR}(\mu), \quad (1.25)$$

and the complementary lower bound holds if the sup in (1.24) is achieved at a probability measure $\nu$ which is compactly supported and has absolutely continuous part in each of its connected components.

**Remark 1.8.** If $m[\lambda_N]$ and $m[\eta_N]$ are supported on $[0, \Re]$, and $c^{KN}_{\lambda N} \eta_N \neq 0$, then deterministically $m[\kappa_N]$ is supported on $[0, 2\Re]$. Moreover, from our construction of $m[\eta_N]$ as in (1.17), the limit of $m[\kappa_N]$ necessarily has a density bounded by 1. Therefore in Theorem 1.7, we have restricted ourselves in the set of measures supported on $[0, 2\Re]$ with density bounded by 1.

8
2 Spherical Integral

In this section we study the spherical integral and the limit function \( I(\mu_A, \mu_B) \) as in (1.2). In Section 2.1, we collect some estimates of the spherical integral and its limit \( I(\mu_A, \mu_B) \) from [?, ?, ?, ?], where it was shown that \( I(\mu_A, \mu_B) \) is related to a variational problem. The results in [?, ?] requires that one of \( \{\mu_A, \mu_B\} \) is compactly supported and the other has bounded second moment and free energy. However, by a continuity argument, it is easy to see that \( I(\mu_A, \mu_B) \) is well-defined when one of \( \{\mu_A, \mu_B\} \) is compactly supported and another has bounded first moment. In Section 2.2, we extend the results in [?] for the setting that one measure is compactly supported and the other has bounded first moment. We remark this is the largest possible set where \( I(\mu_A, \mu_B) \) is well-defined. In this setting we show that the solution of the variational principle converges to the free Brownian bridge. Moreover, we characterize the limiting joint law of \((A_N, U B_N U)\) under the shifted measure

\[
\exp \left( \frac{2N}{\beta} \text{Tr}(A_N U B_N U^*) \right) dU.
\]

Using the joint law of \((A_N, U B_N U)\) under \(\mu_{A_N, B_N}\) as input, in Section 2.3, we compute the derivatives of the limit function \( I(\cdot, \cdot) \). In Section 2.4, we give a more precise description of the solutions of the variational problem characterizing \( I(\cdot, \cdot) \), by transforming the equations for the solution into a Beltrami equation.

2.1 Preliminaries

For any probability measure \( \mu \in \mathcal{M}(\mathbb{R}) \), we denote \( \Sigma(\mu) \) the energy of its logarithmic potential, or its non-commutative entropy,

\[
\Sigma(\mu) = \int \int \log |x - y| d\mu(x) d\mu(y).
\]

We recall the following Theorem from [?], where it is proven that the limit of the spherical integral exists, provided one measure has bounded \( L^2 \) moment and logarithmic potential, another measure is compactly supported.

**Theorem 2.1** ([?, Theorem 1.1]). Let \( A_N, B_N \) be two sequences of deterministic self-adjoint matrices, such that their spectral measures \( \mu_{A_N} \) and \( \mu_{B_N} \) converge weakly to \( \mu_A \) and \( \mu_B \), respectively. If there exists a constant \( \mathcal{R} > 0 \), such that \( \text{supp} \mu_{A_N} \subset [-\mathcal{R}, \mathcal{R}] \) and \( \mu_{B_N}(|x|^2) \leq \mathcal{R}, \Sigma(\mu_B) \geq -\mathcal{R} \), then the spherical integral converges

\[
\lim_{N \to \infty} I_N(A_N, B_N) := \lim_{N \to \infty} \frac{1}{\beta N^2} \int e^{\frac{2N}{\beta} \text{Tr}(A_N U B_N U^*)} dU = I(\mu_A, \mu_B),
\]

where \( U \) follows the Haar measure on the unitary group (resp. orthogonal group) when \( \beta = 2 \) (resp. \( \beta = 1 \)).

The proof of Theorem 2.1 is intimately related with the following large deviation principle based on the Hermitian (resp. symmetric) matrix Brownian motion (that is the process of \( N \times N \) matrices filled with independent Brownian motion entries above the diagonal).

**Theorem 2.2** ([?, Theorem 3.2 and Theorem 3.3]). Let \( A_N \) be a sequence of deterministic diagonal matrices, with diagonal entries \( a_1 \leq a_2 \leq \cdots \leq a_N \) whose spectral measures converge towards \( \mu_A \). Assume that there exist a constant \( \mathcal{R} > 0 \) and \( \epsilon > 0 \) such that

\[
\mu_{A_N}(|x|^2) \leq \mathcal{R}, \quad \mu_A(|x|^2 + \epsilon) \leq \mathcal{R}.
\]
Let \( H_N(t) \) be a Hermitian (resp. Symmetric) Brownian motion. Let \((\lambda_i(t))_{1 \leq i \leq N}\) be the eigenvalues of the self-adjoint matrix
\[
X_N(t) = A_N + H_N(t), \quad t \in [0,1],
\]
and denote by \((\mu^N_\lambda)_{1 \leq i \leq N}\) the empirical measure of these eigenvalues. Then, the law of \((\mu^N_\lambda)_{t \in [0,1]}\), seen as a continuous process with values in the space \( P(\mathbb{R}) \) of probability measures, satisfies a large deviation principle with speed \( N^2 \) and good rate function which is infinite if \( \mu_0 \neq \mu_A \) and otherwise given by \( \beta S_{\mu_A} \),
\[
S_{\mu_A}(\mu) = \sup_{f \in C^{2,1}(\mathbb{R}[0,1])} \left( S^{0,1}(f,\mu) - \frac{1}{2} \langle f,f \rangle_{\mu} \right),
\]
where \( C^{2,1}(\mathbb{R}[0,1]) \) is the set of functions \( f(x,t) \) which is twice differentiable in \( x \) and differentiable in \( t \), \( (f,g)_\mu = \int_0^1 \int \partial_x f(s,x) \partial_x g(s,x) d\mu_s(x) ds \)
\[
S^{0,1}(f,\nu) = \int f(1,x) d\nu_1(x) - \int f(0,x) d\mu_A(x) - \int_0^1 \int \partial_t f(t,x) d\nu_t(x) dt - \frac{1}{2} \int_0^1 \int \frac{\partial_x f(s,x) - \partial_x f(s,y)}{x-y} d\nu_s(x) d\nu_s(y) ds.
\]
As a consequence, the law of \( \mu^N_\lambda \) satisfies a large deviation principle in the scale \( N^2 \) with the rate function
\[
J_\beta(\mu_A,\mu) := \beta \inf \{ S_{\mu_A}(\mu) : \mu_1 = \mu \}.
\]

Theorem 2.1 can then be deduced from Theorem 2.2 by noticing that the law for the eigenvalues of \( X_N(1) \) is given by
\[
e^{-\frac{2N}{\beta} \text{Tr}(A_N^2)} \prod_{i < j} |x_i - x_j|^\beta e^{-\frac{2N}{\beta} \text{Tr}X_N(1)^2} \prod_{i=1}^N dx_i \int e^{\frac{2N}{\beta} \text{Tr}(X_N(1)U_ANU_A^*)} dU,
\]
where \( x_1, x_2, \cdots, x_N \) are the eigenvalues of \( X_N(1) \). Indeed, this formula (2.4) asymptotically yields
\[
J_\beta(\mu_A,\mu) = \frac{\beta}{4} \left( \int x^2 d\mu_A(x) + \int x^2 d\mu(x) \right) - \frac{\beta}{2} \Sigma(\mu) - \beta I(\mu_A,\mu) + \text{const.},
\]
where \( \text{const.} \) is the finite constant coming from the partition function \( Z_N \).

We can directly study the optimizing problem (2.3) and find that the optimizer is characterized via solutions of an Euler equation with negative pressure, see [7, Theorem 2.1].

**Theorem 2.3** ([7, Theorem 2.1]). We assume there exists a constant \( \mathcal{R} > 0, \) such that \( \text{supp} \mu_A \subset [-\mathcal{R}, \mathcal{R}] \), \( \mu_B(\{x\}) \leq \mathcal{R} \) and \( \Sigma(\mu_B) \leq \mathcal{R} \). Then \( I(\mu_A,\mu_B) \) is given by
\[
I(\mu_A,\mu_B) = -\frac{1}{2} \inf S(u,\rho) - \frac{1}{2} \left( \Sigma(\mu_A) + \Sigma(\mu_B) \right) + \frac{1}{4} \left( \int x^2 d\mu_A(x) + \int x^2 d\mu_B(x) \right) - \text{const.}
\]
where
\[
S(u,\rho) = \int_0^1 \int_{\mathbb{R}} \left( \frac{x^2}{3} \rho_t^3 + u^2 \rho_t \right) dx dt;
\]
the inf is taken over all the pairs \((u_t, \rho_t)\) such that \( \partial_t \rho_t + \partial_x (\rho_t u_t) = 0 \) in the sense of distributions, \( \rho_t \geq 0 \) almost surely w.r.t. the Lebesgue measure, \( \int \rho_t dx = 1 \), and with initial and terminal data for \( \rho \) given by
\[
\lim_{t \to 0+} \rho_t(x) dx = \mu_A, \quad \lim_{t \to 1-} \rho_t(x) dx = \mu_B.
\]
where convergence holds in the weak sense.

The infimum in (2.6) is reached at a unique probability measure-valued path \( \rho^*_t \) such that for \( t \in (0, 1) \), \( \rho^*_t \) is absolutely continuous with respect to Lebesgue measure. The pair \( (\rho^*, u^*) \) satisfies the Euler equation for isentropic flow described, for \( t \in (0, 1) \), by the equations

\[
\begin{align*}
\partial_t \rho^*_t (x) &= -\partial_x (\rho^*_t (x) u^*_t (x)), \\
\partial_t (\rho^*_t (x) u^*_t (x)) &= -\partial_x \left( \rho^*_t (x) u^*_t (x)^2 - \frac{\pi^2}{3} \rho^*_t (x)^3 \right),
\end{align*}
\]

in the sense of distributions: for all \( \varphi \in C_{\text{c}}^{\infty}([0, 1]) \),

\[
\int_0^1 \int \partial_t \varphi(t, x) \rho^*_t (x) \, dx \, dt + \int_0^1 \int \partial_x \varphi(t, x) u^*_t (x) \rho^*_t (x) \, dx \, dt = 0,
\]

and, for any \( \varphi \in C_{\text{c}}^{\infty}(\Omega) \) with

\[
\Omega := \{(x, t) \in \mathbb{R} \times (0, 1) : \rho^*_t (x) > 0 \},
\]

\[
\int \left( u^*_t (x) \partial_x \varphi(x, t) + \left( u^*_t (x)^2 - \frac{\pi^2}{3} \rho^*_t (x)^2 \right) \partial_x \varphi(x, t) \right) \rho^*_t (x) \, dx \, dt = 0.
\]

The infimum \( (\rho^*, u^*) \) are smooth in the interior of \( \Omega \), which guarantees that (2.8) and (2.9) hold everywhere in the interior of \( \Omega \). Moreover, \( \Omega \) is bounded in \( \mathbb{R} \times [0, 1] \).

From formula (2.2), if we condition on that the eigenvalues of \( X_N(1) \) are given by \( B_N \), i.e. \( X_N(1) = UB_NU^* \), then the law for the eigenvectors \( U \) of \( X_N(1) \) is given by the integrand of the spherical integral:

\[
d\mu_{AN, B_N}(U) = \frac{1}{Z_N} e^{\frac{2}{\pi} \tau_N((B_N U A_N U^*)^{-1})} \, dU.
\]

Moreover, as argued in [?], the law of \( X_N(t) = A_N + H_N(t) \) conditioned on that \( X_N(1) = A_N + H_N(1) = UB_NU^* \) is the Hermitian Brownian bridge between self-adjoint matrices \( A_N, UB_NU^* \):

\[
dX_N(t) = dH_N(t) + \frac{UB_NU^* - X_N}{1 - t} \, dt, \quad X_N(0) = A_N,
\]

where the “joint law” of \( A_N, UB_NU^* \) is given by (2.12). Solving the above equation shows that for each \( t \in [0, 1] \), we can find a GUE (resp. GOE) matrix \( G_N \), independent of \( A_N, UB_NU^* \), such that

\[
X_N(t) = (1 - t)A_N + tUB_NU^* + \sqrt{t(1-t)} G_N.
\]

We can give a meaning to the large \( N \) limit of \( \{X_N(t), t \in [0, 1]\} \) thanks to the notion of non-commutative joint law from free probability.

**Theorem 2.4** ([?; Theorem 2.6]). Assume \( \mu_A, \mu_B \) are compactly supported. Then the space of free Brownian path distributions \( FBB(\mu_A, \mu_B) \) given by

\[
(1 - t)a + tb + \sqrt{t(1-t)}s, \quad t \in [0, 1],
\]

is closed, where \((a, b)\) are free from the semi-circle law \( s \), with joint distribution such that the distribution of \((a, b)\) are \( \mu_A, \mu_B \) respectively. Moreover

\[
J_{\beta}(\mu_A, \mu_B) := \beta \inf \{S_{\mu_A}(\mu) : \mu_1 = \mu_B\} = \beta \inf \{S_{\mu_A}(\mu) : \mu \in FBB(\mu_A, \mu_B)\}.
\]

This theorem gives several a priori properties of the minimizers of \( S_{\mu_A} \), for instance that they are absolutely continuous and with bounded density in \( (0, 1) \), c.f. Proposition 2.5. Putting together the characterizations of Theorems 2.3 and 2.4, we have the following:

11
Proposition 2.5. We assume that the probability measures $\mu_A, \mu_B$ are compactly supported. The pair $(\rho^*, u^*)$ is the unique solution of the variational problem (2.6), and $f(t, x) = u^*_t(x) + \pi \rho^*_t(x)$ satisfies the complex Burgers equation
\[
\partial_t f(t, x) + f(t, x) \partial_x f(t, x) = 0.
\]
Moreover,
1. $(t, x) \to f(t, x)$ is real analytic in each component in the interior of $\Omega$ as defined in (2.10).
2. $\Im[f(t, x)]/\pi$ converges weakly as $t$ goes to $0$ or $1$ to $\mu_A$ and $\mu_B$ respectively.
3. For any $g \in C^1$,
\[
\lim_{t \to 0^+} \int g(x) \Re[f(t, x)] \rho^*_t(x) \, dx = \int g(x) \Re[f(0, x)] \, d\mu_A,
\]
and the same as time $t \to 1$.
4. There exists a finite constant $R > 0$ (depending on the support of $\mu_A, \mu_B$) such that for all $(x, t)$ in the interior of $\Omega$,
\[
|f(t, x)| \leq \frac{R}{\sqrt{t(1-t)}}.
\]
5. For all $t \in (0, 1)$, all $x \in \Omega$ such that $(t, x_0) \in \partial \Omega$,
\[
\rho^*_t(x) \leq \left( \frac{3}{4\pi^2 t^2 (1-t)^2} \right)^{\frac{3}{2}} (x - x_0)^{\frac{3}{2}}.
\]
6. There exist two operators $a, b$ in a non-commutative probability space $(A, \phi)$ with marginal distributions $(\mu_A, \mu_B)$ so that $ta + (1-t)b + \sqrt{t(1-t)}s$ has the law of $\rho^*_t(x) \, dx$ where $s$ is a semicircular variable free from $(a, b)$.
7. For all $t \in [0, 1]$, let $\{x_t\}_{0 \leq t \leq 1}$ be a non-commutative brownian motion independent of $a, b$, and
\[
dx_t = dx_t + \frac{b - x_t}{1-t} dt, \quad x_0 = a, \quad x_1 = b,
\]
then $x_t$ has the law of $(1-t)a + tb + \sqrt{t(1-t)}s$ given by $\rho^*_t$, and
\[
u^*_t = \frac{1}{t-1} \tau(x_t - b|x_t) + H\rho^*_t, \quad \rho^*_t(x) \, dx \text{ a.s.}
\]
where $H\rho = \text{PV} \int (x-y)^{-1} \rho(y) \, dy$ is the Hilbert transform of $\rho$.

Proof. Most of the proof is already contained in [?; Corollary 2.8] and lies in the representation of the solution in terms of a free Brownian bridge stated in the last two points above, i.e. Item 6,7, see Theorem 2.4; namely, it is shown that there exists two non-commutative variables $a, b$ with marginals distributions $\mu_A, \mu_B$ so that $\rho^*_t(x) \, dx$ is the law of $(1-t)a + tb + \sqrt{t(1-t)}s$, with $s$ a semi-circular law free with $a, b$. Items 1,4,5 are then direct consequences of [?]. By the definition of the variational problem (2.6), we have that $\rho^*_t(x) \, dx$ converges weakly towards $\mu_A$ as $t$ goes to 0. Finally, by (2.19) in [?], $u^*$ is given by (2.16). But
\[
\int g(x) \frac{1}{t-1} \tau(x_t - b|x_t) \rho^*_t(x) \, dx = \tau(g(x_t)(x_t - b))
\]
converges as $t$ goes to 0 by continuity of $g$ and $x_t$ whereas since $H\rho^*_t \in L^2(\rho^*_t)$ (as $\rho^*_t \in L^3(dx)$)
\[
\int g(x) H\rho^*_t(x) \rho^*_t(x) \, dx = \frac{1}{2} \int \int \frac{g(x) - g(y)}{x-y} \rho^*_t(x) \rho^*_t(y) \, dx \, dy,
\]
with $(x, y) \to (g(x) - g(y))/(x-y)$ continuous when $g$ is $C^1$, converges as $t$ goes to 0 or 1 by weak continuity of $\rho^*_t(x) \, dx$. The claim (2.15) follows from the above discussion. \qed
2.2 The spherical integral for $L^1$ distributions and convergence of the non-commutative law

Later in the article, we will need to extend the limit of the spherical integral (1.2) to the setting where one measure has bounded support and the other measure has bounded first moment. The following proposition states that the limit function $I(\cdot, \cdot)$ and other quantities appearing in our main theorems are continuous with respect to the Wasserstein distance $d_W (\cdot, \cdot)$ as defined in (1.6).

**Proposition 2.6.** We assume that the probability measures $\mu, \nu$ and $\mu', \nu'$ satisfy that $\text{supp} \mu, \text{supp} \nu' \subset [-R, R]$, and $\nu(|x|), \nu'(|x|) \leq R$ for some constant $R > 0$. Then there exists a finite constant $C_R$ so that for any small $\delta > 0$ such that $d_W (\mu, \mu') \leq \delta$ and $d_W (\nu, \nu') \leq \delta$,

$$|I(\nu, \mu) - I(\nu', \mu')| = C_R o(1),$$

and

$$\left| \int T_{\mu} T_{\nu} dx - \int T_{\mu'} T_{\nu'} dx \right| = C_R o(1),$$

where we can take $C_R = (R + 1)$ and $o(1)$ to be $\nu(|x|1_{|x| \geq \delta^{-1/2}}) + \nu'(|x|1_{|x| \geq \delta^{-1/2}}) + \delta^{1/2}$.

**Proof.** The first point is proven in [7, Lemma 5.1] in the case where $\int x^2 d\nu(x) + \int x^2 d\mu(x) \leq R$ and $\int |x|^2 d\nu'(x) + \int |x|^2 d\mu'(x) \leq R$. However it is straightforward to extend this estimate to our setting up to remark that we approximate $\nu$ in $I(\nu, \mu)$ by $\nu(1_{|x| \geq \delta^{-1/2}})\delta_0 + \nu 1_{|x| < \delta^{-1/2}}$ up to an error of order $\mathcal{R}\nu(|x|1_{|x| \geq \delta^{-1/2}})$. For the second estimate, we have

$$\left| \int T_{\mu} T_{\nu} dx - \int T_{\mu'} T_{\nu'} dx \right|$$

$$\leq \int |T_{\mu} - T_{\mu'}||T_{\nu'}| dx + \int |T_{\mu'}||T_{\mu} - T_{\mu'}| dx$$

$$\leq R |T_{\mu} - T_{\mu'}| dx + \int |T_{\mu} - T_{\mu'}| \delta^{1/2} dx + \int |T_{\nu'}| \delta^{-1/2} dx$$

$$\leq R d_W (\nu, \nu') + \delta^{1/2} d_W (\mu, \mu') + 2 R o(1) = C_R o(1),$$

where in the third line we used that $|T_{\mu}|, |T_{\mu'}| \leq R$, and in the last line we used the definition of Wasserstein distance (1.8).

**Remark 2.7.** We can view $I(\mu_A, \mu_B)$ as a function of $T_{\mu_A}, T_{\mu_B}$, i.e. $I(T_{\mu_A}, T_{\mu_B}) := I(\mu_A, \mu_B)$. Since the spherical integral

$$\lim_{N \to \infty} \frac{1}{\beta N^2} \log \int e^{\frac{\beta}{2} \int \nu_T A_N U B_N U^* dU} = I(T_{\mu_A}, T_{\mu_B})$$

is convex in both $A_N$ and $B_N$, the limit is also convex in $T_{\mu_A}$ and $T_{\mu_B}$.

Using Proposition 2.6, we can extend Theorems 2.1 and 2.3 to measures $\nu$ and $\mu$ such that $\nu(|x|) \leq R$ and $\text{supp} \mu \subset [-R, R]$. In fact, in this case, thanks to Proposition 2.6, $I(\nu, \mu)$ is well defined and continuous. It can thus be extended to the setting that one measure has $L^1$ moment and another measure is compactly supported. Similar to Theorem 2.4, we study this convergence and show that it entails the convergence of the non commutative distribution of $(A_N, U B_N U^*)$ under the tilted measure. Since $A_N$ is unbounded but $B_N$ is bounded, we hereafter shall consider the joint law of unbounded self-adjoint operators (see [7, p. 343] for the definition of the joint law of possibly unbounded self-adjoint operators). If $(A, \varphi)$ denotes the non-commutative probability space in which three self-adjoint
operators $b, s$ bounded and $a$ with finite moment live, their non-commutative probability distributions can be described by the family $\mathcal{F}(a, b, s)$, where $F(a, b, s)$ belongs to the complex vector space of test functions $\mathcal{F}$ given by
\[
\sum_{i} \zeta_i P_i^b(a, s) \frac{1}{z_i^1 - a} P_i^s(b, s) \cdots \frac{1}{z_{k_i}^i - a} P_i^1(b, s).
\]
(2.17)

where the $z^i_j$ belong to $\mathbb{C}\backslash\mathbb{R}$, the $\zeta_i$ are some complex numbers, the $P^i_j$ are polynomials in two non-commutative variables $b, s$ and $k_i$ are some integer numbers. Hence, we can view the non-commutative probability distribution of $(a, b, s)$ as an element of the dual of $\mathcal{F}$. Note that by density, we can approximate uniformly any Stieltjes transform of $(1 - t)a + tb + \sqrt{t(1 - t)}s$ by elements of $\mathcal{F}$. We denote $\tau$ the marginal distribution of $a, b$ under $\varphi$, i.e. for any test function $F(a, b)$ independent of $s$, $\tau(F(a, b)) = \varphi(F(a, b))$.

**Theorem 2.8.** Let $A_N, B_N$ be two sequences of deterministic self-adjoint matrices, such that their spectral measures $\mu_{A_N}$ and $\mu_{B_N}$ converge in Wasserstein distance (1.6) towards $\mu_A$ and $\mu_B$ respectively. We assume that there exists a constant $R > 0$, such that $\mu_{A_N}([x]) \leq R$ and $\text{supp} \mu_{B_N} \subset [-R, R]$. Then
\[
\lim_{N \to \infty} \frac{1}{RN^2} \log \int e^{\frac{i}{N} \text{Tr}(A_N U B_N U^*)} dU = I(\mu_A, \mu_B).
\]
Moreover, for any $F \in \mathcal{F}$ as defined in (2.17), the following limit exists:
\[
\lim_{N \to \infty} \frac{1}{N} \text{Tr}(F(A_N U B_N U^*)) d\mu_{A_N, B_N}(U) = \tau(F(a, b)),
\]
where $d\mu_{A_N, B_N}$ was defined in (2.12). Moreover, for all $t \in [0, 1]$, let $\{s_t\}_{0 \leq t \leq 1}$ be a non-commutative Brownian motion free from $a, b$, and
\[
dx_t = ds_t + \frac{b - x_t}{1 - t}, \quad x_0 = a, \quad x_1 = b,
\]
(2.19)
then $x_t$ has the law of $(1 - t)a + tb + \sqrt{t(1 - t)}s$ given by $\rho_t^s$, for all $t \in (0, 1)$, $\partial_t \rho_t^s + \partial_s (\rho_t^s u_t^s) = 0$ where $u_t^s$ is given by (2.16).

We remark that $x_t$ belongs to $L^1$ for all $t \in [0, 1]$, so that the conditional expectation in (2.16) makes sense. As a consequence of Theorem 2.8, if $G_N$ is a GUE matrix, independent of $U$ with law (2.12), then for all $f$ bounded continuous
\[
\lim_{N \to \infty} \mathbb{E} \left[ \frac{1}{N} \text{Tr} \left( f(t U B_N U^*) + (1 - t)A_N + \sqrt{t(1 - t)}G_N \right) \right] = \int f(x) \rho^a_t(x) dx.
\]

**Proof of Theorem 2.8.** We first show the convergence of the joint law of $(A_N, U B_N U^*)$ under $\mu_{A_N, B_N}$. We denote $A_N = \text{diag}\{a_1, a_2, \ldots, a_N\}$ and $B_N = \text{diag}\{b_1, b_2, \ldots, b_N\}$, and without loss of generality we assume that $a_1 \leq a_2 \leq \cdots \leq a_N$, $b_1 \leq b_2 \leq \cdots \leq b_N$. Let $(H_N(t), t \in [0, 1])$ be the matrix Hermitian (resp. symmetric) Brownian motion and consider the process $X_N(t) = A_N + H_N(t)$. From (2.4), one can see that the distribution of $X_N(1) = A_N + H_N(1)$ conditioning to have eigenvalues given by $B_N$ is the same as the law of $U B_N U^*$ where $U$ has distribution $\mu_{A_N, B_N}$. There is another interpretation of the eigenvalues of $X_N(t)$ as non-intersecting Brownian motions: the law of the eigenvalues of $X_N(t)$ follows a Dyson Brownian bridge between $(a_1, a_2, \ldots, a_N)$ and $(b_1, b_2, \ldots, b_N)$. In other words it is the law of Brownian bridges $w_1(t) \leq w_2(t) \leq \cdots \leq w_N(t)$, where $w_i(t)$ is from $a_i$ to $b_i$, conditioning not to intersect each other. In the following, we first prove the convergence of the empirical measure $\mu^N$ of the eigenvalues of $X_N(t)$ conditioned to have eigenvalues $\{b_1, b_2, \ldots, b_N\}$ at time $t = 1$ by a comparison argument.
Claim 2.9. Let $A_N = \text{diag}\{a_1, a_2, \ldots, a_N\}$ and $B_N = \text{diag}\{b_1, b_2, \ldots, b_N\}$ be as in Theorem 2.8, and $w_1(t), w_2(t), \ldots, w_N(t)$ be the nonintersecting Brownian bridges between them. Then, the empirical measure
\[ \mu_{i}^{N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{w_i(t)}, \]  
(2.20)
converges weakly almost surely.

Proof. There is a monotonicity statement for nonintersecting Brownian bridges [?, Lemmas 2.6 and 2.7]. Given two pairs of boundary data $(a_1 \leq a_2 \leq \cdots \leq a_N)$, $(b_1 \leq b_2 \leq \cdots \leq b_N)$, $(\tilde{a}_1 \leq \tilde{a}_2 \leq \cdots \leq \tilde{b}_N)$ and $(\tilde{b}_1 \leq \tilde{b}_2 \leq \cdots \leq \tilde{b}_N)$. We consider nonintersecting Brownian bridges from $(a_1, a_2, \ldots, a_N)$ and $(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_N)$ to $(b_1, b_2, \ldots, b_N)$ and $(\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_N)$: $w_1(t) \leq w_2(t) \cdots \leq w_N(t)$ and $\tilde{w}_1(t) \leq \tilde{w}_2(t) \cdots \leq \tilde{w}_N(t)$. If $a_i \geq \tilde{a}_i$ and $b_i \geq \tilde{b}_i$ for all $1 \leq i \leq N$, [?, Lemmas 2.6 and 2.7] gives a coupling, such that at any time $0 \leq t \leq 1$, $w_i(t) \geq \tilde{w}_i(t)$ for all $1 \leq i \leq N$. Especially if we denote the empirical particle density of the two nonintersecting Brownian bridges as $\mu_{i}^{N} = (1/N) \sum_{i=1}^{N} \delta_{w_i(t)}$ and $\tilde{\mu}_{i}^{N} = (1/N) \sum_{i=1}^{N} \delta_{\tilde{w}_i(t)}$, then
\[ h^{N}(x, t) := \mu_{i}^{N}([-\infty, x]) \leq \tilde{\mu}_{i}^{N}([-\infty, x]) =: \tilde{h}^{N}(x, t), \]  
(2.21)
for any $x \in \mathbb{R}$, almost surely under the coupling. We note it is possible that some $a_i, \tilde{a}_i, b_i, \tilde{b}_i$, are at $\pm \infty$ and $\mu_{i}^{N}, \tilde{\mu}_{i}^{N}$ may have delta mass at $\pm \infty$. The statements in [?, Lemmas 2.6 and 2.7] still hold if some particles are at $\pm \infty$. Combining the discussion above with Theorem 2.4, the empirical density of the boundary data converges to $\mu_A, \mu_B, \tilde{\mu}_A, \tilde{\mu}_B$ respectively with compact support on $(-\infty, \infty)$, possibly some delta mass at $+\infty$, then $\mu_{i}^{N}, \tilde{\mu}_{i}^{N}$ converge weakly to some measure-valued processes $\mu_i, \tilde{\mu}_i$ respectively. Their cumulative densities satisfy
\[ h(x, t) := \mu_{i}([-\infty, x]) \leq \tilde{\mu}_{i}([-\infty, x]) =: \tilde{h}(x, t). \]  
(2.22)

If $\mu_{A_N}, \mu_{B_N}$ are both uniformly compactly supported, the convergence of the empirical particle density of the nonintersecting Brownian bridges follows from Theorem 2.4. If $\mu_{A_N}$ or $\mu_{B_N}$ are not compactly supported, we approximate them with compact ones and use the monotonicity property of the nonintersecting Brownian bridges to show the existence of limit density.

We denote $A_N^{\epsilon+}, B_N^{\epsilon+}$ the new boundary data by moving the first and last $[\epsilon N]$ particles of $A_N, B_N$ to the location $-\infty$, i.e. they are $(-\infty \leq \cdots \leq -\infty \leq a_{\lfloor \epsilon N \rfloor + 1} \leq a_{\lfloor \epsilon N \rfloor + 2} \cdots \leq a_N - \lfloor \epsilon N \rfloor)$ and $(-\infty \leq \cdots \leq b_{\lfloor \epsilon N \rfloor + 1} \leq b_{\lfloor \epsilon N \rfloor + 2} \cdots \leq b_N - \lfloor \epsilon N \rfloor)$. We denote the nonintersecting Brownian bridge between them as $w_N^{\epsilon+}(t) \leq w_N^{\epsilon+}(t) \leq \cdots \leq w_N^{\epsilon+}(t)$. Then $A_N, B_N, A_N^{\epsilon+}, B_N^{\epsilon+}$ satisfy the monotone condition, (2.21) implies
\[ \mu_{i}^{N}([-\infty, x]) \leq \frac{1}{N} \#\{j : w_{i}^{\epsilon+}(t) \leq x\} =: h_{i}^{N}(x, t), \]  
(2.23)
almost surely. The empirical particle densities of $A_N^{\epsilon+}, B_N^{\epsilon+}$ converge to measures in the form of a delta mass at $-\infty$ plus a compactly supported measure. From the discussion above and Theorem 2.4, the limits exist
\[ \lim_{N \to \infty} h_{i}^{N}(x, t) = h_{i}^{\epsilon+}(x, t), \]  
(2.24)
and $\lim_{\epsilon \to -\infty} h_{i}^{\epsilon+}(x, t) = 2\epsilon$, $\lim_{\epsilon \to +\infty} h_{i}^{\epsilon+}(x, t) = 1$. Similarly, we denote $A_N^{-\epsilon}, B_N^{-\epsilon}$ the new boundary data by moving the first and last $[\epsilon N]$ particles of $A_N, B_N$ to the location $+\infty$, and define $h_{i}^{\epsilon-}$ analogously. Then we have
\[ \liminf_{N \to \infty} \mu_{i}^{N}([-\infty, x]) \geq h_{i}^{\epsilon-}(x, t), \]  
(2.25)
almost surely. Combining those estimates (2.23), (2.24) and (2.25) together, we get
\[ h^e_t(x, t) \leq \lim \inf_{N \to \infty} \mu^N_t([-\infty, x]) \leq \lim \sup_{N \to \infty} \mu^N_t([-\infty, x]) \leq h^e_+(x, t), \]
almost surely. From our construction, \( h^e_t(x, t) \) is simply a shift of \( h^e_+(x, t) \), i.e. \( h^e_t(x, t) = h^e_+(x, t) - 2\varepsilon \).
Moreover, thanks to the monotonicity property of nonintersecting Brownian bridges, \( h^e_t(x, t) \) is nondecreasing in \( \varepsilon \) and \( h^e_+(x, t) \) is nonincreasing in \( \varepsilon \). Thus the limits exist
\[ h(x, t) := \lim_{\varepsilon \to 0} h^e_t(x, t) = \lim_{\varepsilon \to 0} h^e_+(x, t), \]
and \( h(x, t) \) gives the limiting cumulative density of \( \mu^N_t \). This finishes the proof of Claim 2.9.

The law of \( X_N(t) = A_N + H_N(t) \) conditioned on that \( X_N(1) = A_N + H_N(1) = UB_N U^* \) is the Hermitian Brownian bridge between self-adjoint matrices \( A_N, UB_N U^* \):
\[ dX_N(t) = dH_N(t) + \frac{UB_N U^* - X_N(t)}{1 - t} dt, \quad X_N(0) = A_N, \]
where the “joint law” of \( A_N, UB_N U^* \) is given by (2.12). Solving the above equation shows that for each \( t \in [0, 1] \), we can find a GUE (resp. GOE) matrix \( G_N \), independent of \( A_N, UB_N U^* \), such that
\[ X_N(t) = tUB_N U^* + (1 - t)A_N + \sqrt{t(1 - t)} G_N. \] (2.26)
Therefore we have proven in Claim 2.9 that if \( G_N \) is a GUE matrix, independent of \( U \) with law \( \mu_{A_N, B_N} \) then for all \( f \) bounded continuous
\[ \lim_{N \to \infty} \mathbb{E} \left[ \frac{1}{N} \operatorname{Tr} \left( f(tUB_N U^* + (1 - t)A_N + \sqrt{t(1 - t)} G_N) \right) \right] = \int f(x) \rho_1^*(x) dx. \] (2.27)
We next show that this is enough to characterize the limit points of the non-commutative law
\[ \tau_N(F) = \int \frac{1}{N} \operatorname{Tr}(F(A_N, UB_N U^*)) d\mu_{A_N, B_N}(U), \] (2.28)
for all \( F \in \mathcal{F} \). This non-commutative law is sequentially tight since it belongs to a compact space. We can therefore consider a limit point \( \tau_{A,B} \) and need to show it is unique. We construct \( \tau \) to be the joint law of \( a, b \) under \( \tau_{A,B} \) and a free semi-circular variable \( s \). (2.27) already implies that for all \( t \in [0, 1] \), all \( z \in \mathbb{C} \setminus \mathbb{R} \),
\[ \tau \left( \frac{1}{z - (1 - t)a - tb - \sqrt{t(1 - t)s}} \right) = \int \frac{1}{z - x} \rho_1^*(x) dx. \]
This is enough to deduce the distribution \( \nu_t \) of \((1 - t)a + tb\), since the R-transform formula yields for \( z \) small enough
\[ R_{\nu_t}(z) = R_{\nu_1^*}(z) - t(1 - t)z, \]
where \( R_{\nu_t} \) and \( R_{\nu_1^*} \) are the R-transforms of \( \nu_t \) and \( \rho_1^* \) respectively. This defines uniquely the Stieltjes transform of \( \nu_t \). We then deduce by a change of variable \( s = t/(1 - t) \) that
\[ \tau \left( \frac{1}{z - a - sb} \right) = \int \frac{1 - t}{(1 - t)z - x} d\nu_t(x). \] (2.29)
Since \( \|b\| < \infty \), we can Taylor expand the above expression
\[ \tau \left( \frac{1}{z - a - sb} \right) = \tau \left( \frac{1}{z - a} + \frac{s}{(z - a)^2} b \right) + O(s^2). \] (2.30)
Since \( \tau(\frac{1}{z-a}) = \int d\mu_A(x)/(z-x) \) is known, we conclude from combining (2.29) and (2.30) that we also know
\[
\tau \left( \frac{1}{(z-a)^\tau} b \right) = -\partial_\ell \tau \left( \frac{1}{(z-a)} b \right),
\]
(2.31)

Hence, \( \tau((z-a)^{-1} b) \) is known. In the following we show that this is enough to retrieve the complete joint distribution \( \tau \) thanks to loop equations:

**Claim 2.10.** The non-commutative law \( \tau \) is uniquely determined by \( (\rho_1^i(x))_{i \in (0,1)} \), and hence by \( (\mu_A, \mu_B) \).

**Proof.** We want to show that observables
\[
H_{n_1, \ldots, n_p}(z_1, \ldots, z_p) = \tau \left( \frac{1}{z_1 - a} b^{n_1} \frac{1}{z_2 - a} b^{n_2} \cdots \frac{1}{z_{p-1} - a} b^{n_{p-1}} \right),
\]
are uniquely defined for any choices of \( n_i \geq 0 \) and \( z_i \in \mathbb{C} \backslash \mathbb{R} \). We do that by induction on \( p \) and \( \sum n_i \). We assume we know \( H \) for \( p > P - 1 \) for all \( n_i \in \mathbb{N} \), and for \( p = P \) for \( \sum n_i = m \). Note that our induction hypothesis is fulfilled for \( P = 1 \) and \( \sum n_i = 1 \) by (2.31), and for all \( P > 1 \) with \( \sum n_i = 0 \). To proceed by induction we use the loop equations, see e.g. [??], which imply that
\[
\tau \otimes \tau(\partial F) = \tau(F(ab - ba)),
\]
(2.32)

where if \( F = \frac{1}{z_1 - a} b^{n_1} \frac{1}{z_2 - a} b^{n_2} \cdots \frac{1}{z_{p-1} - a} b^{n_{p-1}} \),
\[
\partial F = \sum_{F_1,F_2} (F_1 b \otimes F_2 - F_1 \otimes b F_2).
\]

Hence, if \( p \leq P - 1 \) or \( p = P \) and \( \sum n_i = m \), we can compute uniquely by the induction hypothesis \( \tau \otimes \tau(\partial F) \). In the right hand side of the loop equation (2.32) we have
\[
\tau(F ((a-z_1)b - b(a-z_1))) = \tau(F(a-z_1)b) - \tau \left( b^{n_1} \frac{1}{z_2 - a} b^{n_2} \cdots \frac{1}{z_{p-1} - a} b^{n_{p-1}} \right).
\]

The last term is known by the induction relation. So the loop equation allows us to compute
\[
\tau(F(a-z_1)b).
\]
(2.33)

If \( n_p = 0 \), we can simplify the above to get
\[
(z_p - z_1) \tau(Fb) + \tau \left( \frac{1}{z_1 - a} b^{n_1} \frac{1}{z_2 - a} b^{n_2} \cdots \frac{1}{z_{p-1} - a} b^{n_{p-1}} \right),
\]
where the second term is known by the induction. Hence, we have been able to compute \( \tau(Fb) \) provided \( z_p \neq z_1 \) but then by a continuity argument, it extends to the case when \( z_p = z_1 \). If \( n_p \geq 1 \), by the same argument as above, we use the loop equation (2.32), with \( F = b^{n_p-\ell} \frac{1}{z_1 - a} b^{n_1} \frac{1}{z_2 - a} b^{n_2} \cdots \frac{1}{z_{p-1} - a} b^{n_{p-1}} \), with \( 0 \leq \ell \leq n_p - 1 \), to compute
\[
\tau \left( \tilde{F} ((a-z_1)b - b(a-z_1)) \right)
\]
\[
= \tau \left( \frac{1}{z_1 - a} b^{n_1} \frac{1}{z_2 - a} b^{n_2} \cdots b^\ell (a-z_1) b^{n_{p-\ell+1}} \right) - \tau \left( \frac{1}{z_1 - a} b^{n_1} \frac{1}{z_2 - a} b^{n_2} \cdots b^{\ell+1} (a-z_1) b^{n_{p-\ell}} \right),
\]
where we used that \( \tau \) is tracial. By a telescopic sum for \( 0 \leq \ell \leq n_p - 1 \), we can compute
\[
\tau \left( \frac{1}{z_1 - a} b^{n_1} \frac{1}{z_2 - a} b^{n_2} \cdots \frac{1}{z_{p-1} - a} (a-z_1) b^{n_{p-1}} \right) - \tau(F(a-z_1)b).
\]

17
In the above expression, we have already computed its second term in (2.33), while we can rewrite the first term as

\[(z_p - z_1)\tau \left( \frac{1}{z_1 - a} b_{n_1}^1 \frac{1}{z_2 - a} b_{n_2}^2 \cdots \frac{1}{z_p - a} b_{n_p}^p \right) - \tau \left( \frac{1}{z_1 - a} b_{n_1}^1 \frac{1}{z_2 - a} b_{n_2}^2 \cdots \frac{1}{z_{p-1} - a} b_{n_{p-1}}^{p-1} \right).\]

By our induction hypothesis, we can compute the second term in the above expression. We conclude that we can compute by our induction hypothesis

\[\tau \left( \frac{1}{z_1 - a} b_{n_1}^1 \frac{1}{z_2 - a} b_{n_2}^2 \cdots \frac{1}{z_p - a} b_{n_p}^p \right),\]

when \(z_p \neq z_1\). We then extend the definition for \(z_p = z_1\) by continuity. This yields our induction hypothesis for \(\sum n_i = m + 1\) since \(\tau\) is tracial. This finishes the proof of Claim 2.10.

Claim 2.10 implies the convergence of the non-commutative law \(\tau_N\) as in (2.28). This finishes the proof of claim 2.18. The last point of Theorem 2.8 is a consequence of the fact that free stochastic calculus implies that the distribution of \(x_t\) solution of

\[dx_t = d\sigma_t + \frac{z_t - b}{t-1} dt\]

satisfies the transport equation \(\partial_t \rho_t^* + \partial_z (\rho_t^* u_t^*) = 0\) with the announced \(u_t^*\).

\(\Box\)

Theorem 2.3 gives a quite complicated formula for \(I\). However, we can obtain asymptotic limits which are much easier to handle based on the following proposition. The estimates will be used to study the large deviation rate functions.

**Proposition 2.11.** We assume that the probability measures \(\nu, \mu\) satisfies that \(\nu(|x|) < \infty\) and \(\text{supp } \mu \subset [-R, R]\) for some constant \(R > 0\). Then for any small \(\varepsilon > 0\), there exists a constant \(C(\varepsilon) > 0\) such that

\[\frac{1}{2} \int T_\nu T_\mu dx - O(\varepsilon \nu(|x|)) - C(\varepsilon) \leq I(\nu, \mu) \leq \frac{1}{2} \int T_\nu T_\mu dx.\]  

(2.34)

Here, one can take \(O(\varepsilon) = R(3\varepsilon + \varepsilon^2)\) and \(C(\varepsilon)\) depending only on \(\varepsilon\).

As a consequence, we deduce that if \(L \# \nu\) is the pushforward of \(\nu\) by the homothety of factor \(L:\int f(Lx) d\nu(x) = \int f(x) dL \# \nu(x)\), then

\[\lim_{L \to \infty} \frac{1}{2} \int I(L \# \nu, \mu) = \frac{1}{2} \int T_\nu T_\mu dx.\]

**Proof of Proposition 2.11.** Let \(A_N = \text{diag}(a_1, a_2, \cdots, a_N), B_N = \text{diag}(b_1, b_2, \cdots, b_N)\) be two sequences of deterministic self-adjoint matrices, with \(a_1 \geq a_2 \geq \cdots \geq a_N\), and \(b_1 \geq b_2 \geq \cdots \geq b_N\), such that they are \(N\)-quantiles of the measures \(\nu\) and \(\mu\) respectively. The upper bound in (2.34) follows directly from the spherical integral,

\[I(\nu, \mu) = \lim_{N \to \infty} \frac{1}{\beta N^2} \log \int e^{\beta N \text{Tr}(A_N U B_N U^*)} dU\]

\[\leq \limsup_{N \to \infty} \frac{1}{2N} \sum_{i=1}^N a_i b_i = \frac{1}{2} \int T_\nu T_\mu dx.\]

For the lower bound, we denote \(B_\varepsilon\) the set of unitary matrices when \(\beta = 2\), or orthogonal matrices when \(\beta = 1\), such that

\[B_\varepsilon = \{ U : |U_{ii} - 1| \leq \varepsilon \text{ for } 1 \leq i \leq N\}.

18
On the set $\mathbb{B}_\varepsilon$, we have for all $i$, $|U_{ii} - 1| \leq \varepsilon$ and $\sum_j |U_{ij} - \delta_{ij}|^2 \leq 2\varepsilon$. It follows that

$$\text{Tr}(A_N U B_N U^*) = \text{Tr}(A_N B_N) + \text{Tr}(A_N (U - I_N) B_N (U^* - I_N)) + 2 \Re \text{Tr}(A_N B_N (U^* - I_N))$$

$$\geq \sum_i a_i b_i - \Re(\varepsilon + \varepsilon^2) \sum_i |a_i| - 2 \Re \sum_i |a_i|.$$  (2.35)

Moreover, notice that $U$ is normal with complex eigenvalues $\{z_1, \ldots, z_N\}$ so that

$$B_\varepsilon = \bigcap_{1 \leq i \leq N} \{|z_i - 1| \leq \varepsilon\} = \left\{ \max_{|v| = 1} |\langle v, (U - I) v \rangle| \leq \varepsilon \right\} \subset \mathbb{B}_\varepsilon.$$  

The joint law of the eigenvalues is well known to be given by a Coulomb gas law from which classical large deviation estimates, see [?, ?], show that there exists a constant $C(\varepsilon) > 0$ such that $B_\varepsilon$ holds with probability at least $e^{-C(\varepsilon)N^2}$. As a consequence, we also have $P(B_\varepsilon) \geq e^{-C(\varepsilon)N^2}$. The lower bound in (2.34) follows

$$I(\nu, \mu) = \lim_{N \to \infty} \frac{1}{\beta N^2} \log \int e^\frac{\beta N}{2} \text{Tr}(A_N U B_N U^*) dU$$

$$\geq \frac{1}{2} \int T_\nu T_\mu dx + \liminf_{N \to \infty} \frac{1}{\beta N^2} \log P(\mathbb{B}_\varepsilon) - O(\varepsilon \nu(|x|))$$

$$\geq \frac{1}{2} \int T_\nu T_\mu dx - O(\varepsilon \nu(|x|) - C(\varepsilon)).$$

Proposition 2.12. We assume that the probability measures $\nu, \mu$ satisfies $\nu(|x|) \leq \Re$ and $\sup \mu \subset [-\Re, \Re]$ for some constant $\Re > 0$. Then for any small $\varepsilon > 0$, it holds

$$I(\nu, \mu) = I(\nu^\varepsilon, \mu) + \frac{1}{2} \int_{|\nu'| > 1/\varepsilon} T_\nu T_\mu dx + C_0 o_\varepsilon(1),$$  (2.36)

where $\nu'$ is the restriction of $\nu$ on the interval $|x| \leq 1/\varepsilon$, i.e. $\nu' = \nu 1(|x| \leq 1/\varepsilon) + \delta_0 \int_{|x| > 1/\varepsilon} \nu'$, and the implicit error term depends only on $\varepsilon$ and $\Re$.

Proof. Let $A_N = \text{diag}(a_1, a_2, \ldots, a_N), B_N = \text{diag}(b_1, b_2, \ldots, b_N)$ be two sequences of deterministic self-adjoint matrices, with $a_1 \geq a_2 \geq \cdots \geq a_N$, and $b_1 \geq b_2 \geq \cdots \geq b_N$, such that they are $N$-quantiles of the measures $\nu$ and $\mu$ respectively. We denote the truncated diagonal matrix $A_N^\delta = \text{diag}(a_1 1(|a_1| \leq 1/\varepsilon), a_2 1(|a_2| \leq 1/\varepsilon), \ldots, a_N 1(|a_N| \leq 1/\varepsilon))$, which is obtained by removing large entries of $A_N$. The upper bound in (2.34) follows directly from the spherical integral,

$$I(\nu, \mu) = \lim_{N \to \infty} \frac{1}{\beta N^2} \log \int e^\frac{\beta N}{2} \text{Tr}((A_N - A_N^\delta) U B_N U^* + \text{Tr}(A_N^\delta U B_N U^*)) dU$$

$$\leq \lim_{N \to \infty} \frac{1}{\beta N^2} \log \int e^{\sum_i |a_i| > 1/\varepsilon} a_i b_i + \text{Tr}(A_N^\delta U B_N U^*) dU$$

$$= I(\nu^\varepsilon, \mu) + \frac{1}{2} \int_{|\nu'| > 1/\varepsilon} T_\nu T_\mu dx,$$  (2.37)

where we used that the empirical eigenvalue density of $A_N^\delta$ converges to $\nu^\varepsilon$.

In the following, we prove the lower bound, which is more involved. Let $N_1 = |\{ i : a_i > 1/\varepsilon \}|$, $N_2 = |\{ i : a_i \leq 1/\varepsilon \}|$ and $N_3 = |\{ i : a_i < 1/\varepsilon \}|$. Since by our assumption that $\nu(|x|) \leq \Re$, it follows that $N_1, N_3 \equiv \varepsilon N$. We rewrite $A_N, U$ as block matrices

$$A_N = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{bmatrix}, \quad U = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix}, \quad U_i \in \mathbb{R}^{N_i \times N} \text{ for } i = 1, 2, 3.$$
With these new notations, we rewrite the exponent as
\[ \text{Tr}(A_N U B_N U^*) = \text{Tr}(A_1 U_1 B_N U_1^*) + \text{Tr}(A_2 U_2 B_N U_2^*) + \text{Tr}(A_3 U_3 B_N U_3^*). \]

We denote \( B_δ \) the set of unitary matrices when \( β = 2 \), or orthogonal matrices when \( β = 1 \)
\[ B_δ = \{ U : |U_{ii} - 1| ≤ δ \text{ for } 1 ≤ i ≤ N_1, N - N_3 < i ≤ N \}. \]
Notice that \( P = N_1 + N_3 ≤ 2εRN \). We show that there exists a constant \( C(δ) < ∞ \) such that \( \mathbb{P}(B_δ) ≥ e^{-C(δ)εRN^2} \).
We prove the case for orthogonal matrices, the unitary case can be proven in the same way. We can represent the joint distribution of the vectors \( \{U_i, i ≤ N_1, i > N - N_3\} \) as the array of vectors obtained by applying Gram-Schmidt orthonormalization procedure to independent Gaussian vectors \( (g_1, g_2, \cdots, g_P) \) with \( P = N_1 + N_3 \). We denote the event
\[ B_δ = \left\{ g_i \geq \sqrt{2N/δ}, 1 ≤ i ≤ N_1; g_i(N-P+i) \geq \sqrt{2N/δ}, N_1 + 1 ≤ i ≤ P \right\}. \]
On this event the entries of \( g_1, g_2, \cdots, g_P \) are still independent Gaussian random variables. Moreover, we have \( \mathbb{P}(B_δ) ≥ e^{-C(δ)PN}. \) The Gram-Schmidt orthonormalization procedure sends \( g_i \) to
\[ U_i = \frac{g_i - P_{i-1}g_i}{\|g_i - P_{i-1}g_i\|_2}, \]
where \( P_{i-1} \) is the projection on the span of \( g_1, g_2, \cdots, g_{i-1} \). In the following we show that conditioning on \( B_δ \), with respect to the randomness of \( g_i \), with high probability (larger than \( 1 - e^{-C(δ)εRN} \)):
1. \( \|g_i\|^2 = g_i^2 + N + O(\sqrt{N} \log N), \) for \( 1 ≤ i ≤ N_1; \|g_i\|^2 = g_i^2(N-P+i) + N + O(\sqrt{N} \log N), \) for \( N_1 + 1 ≤ i ≤ P \).
2. \( \|P_{i-1}g_i\|_2 = g_i(\sqrt{\frac{1}{i} + o(\log N)})/\sqrt{N} + \sqrt{\frac{1}{i} + o(\log N)}), \) for \( 1 ≤ i ≤ N_1; \|P_{i-1}g_i\|_2 = g_i(N-P+i)(\sqrt{\frac{1}{i} + o(\log N)})/\sqrt{N} + \sqrt{\frac{1}{i} + o(\log N)}), \) for \( N_1 + 1 ≤ i ≤ P \).
The first item follows easily from the concentration of \( \chi^2 \) distributions. For the second item we prove the case that \( 1 ≤ i ≤ N_1 \), the case for \( N_2 + 1 ≤ i ≤ P \) follows from the same argument. By the triangle inequality
\[ \|P_{i-1}g_i\|_2 ≤ g_i\|P_{i-1}g_i\|_2 + \|P_{i-1}(g_i - g_i e_i)\|_2 \]
For the first projection on the right hand side of (2.40), we can upper bound it by replacing the projection to the span of \( g_1, g_2, \cdots, g_{i-1} \) forgetting the \( 1, 2, \cdots(i-1) \)-th coordinates. In this way the space we project on is the span of \( i-1 \) independent standard Gaussian vectors. The length of the projection is \( g_i(\sqrt{\frac{1}{i} + o(\log N)})/\sqrt{N} \) with high probability. For the second projection on the right hand side of (2.40), we can replace the projection to the span of \( g_1, g_2, \cdots, g_{i-1} \) forgetting the \( i \)-th coordinate. In this way \( g_i - g_i e_i \) is a standard Gaussian vector. The length of its projection to a \( (i-1) \)-dim subspace is \( \sqrt{\frac{1}{i} + o(\log N)} \) with high probability. The second item follows. Combining the arguments above, with high probability
\[ U_{ii} = \frac{g_i - (P_{i-1}g_i)_i}{\|g_i - P_{i-1}g_i\|_2} = \frac{g_i(1 - O(\sqrt{\frac{1}{i}})) + O(\sqrt{\frac{1}{i} \log N})}{\sqrt{1 - O(\sqrt{\frac{1}{i}})g_i^2 + N + O(\sqrt{N} \log N)}} = 1 + O(\sqrt{\frac{1}{i}}). \]
We recall that from our construction of the set \( B_δ \), \( g_i^2 ≥ 2N/δ \). It follows that \( |U_{ii} - 1| ≤ δ \), provided that \( ε \) is small enough. By a union bound, conditioning on the event \( B_δ \), with high probability it holds that \( |U_{ii} - 1| ≤ δ \) for \( 1 ≤ i ≤ N_1, N - N_3 < i ≤ N \). Therefore, \( B_δ \) holds with probability at least \( e^{-C(δ)PN} = e^{-C(δ)εRN} \).
On the set $\mathcal{B}_δ$, we have $|U_{ii} - 1| \leq δ$ and $\sum_{j} |U_{ij} - δ_{ij}|^2 \leq \delta^2$ for $1 \leq i \leq N_1, N - N_3 < i \leq N$. Similarly to (2.35), on the set $\mathcal{B}_δ$, we have

$$\text{Tr}(A_1 U_1 B_N U_1^*) \geq \sum_{i \leq N_1} a_i b_i - R^2(3\delta + \delta^2), \quad \text{and} \quad \text{Tr}(A_3 U_3 B_N U_3^*) \geq \sum_{i > N - N_3} a_i b_i - R^2(3\delta + \delta^2).$$

(2.42)

Since $U_2$ is unitary/orthogonal invariant, with (2.42), we have a lower bound for the spherical integral

$$\frac{1}{\beta N^2} \log \int \frac{d\mathcal{N}}{e} \text{Tr}(A_N U_B U^*) dU \geq \frac{1}{\beta N^2} \log \int_{\mathcal{B}_δ} \frac{d\mathcal{N}}{e} \text{Tr}(A_1 U_1 B_N U_1^*) + \text{Tr}(A_2 U_2 B_N U_2^*) + \text{Tr}(A_3 U_3 B_N U_3^*) dU \geq \frac{1}{2N} \sum_{i > N - N_3} a_i b_i + \frac{1}{\beta N^2} \log \int_{\mathcal{B}_δ} \frac{d\mathcal{N}}{e} \text{Tr}(A_2 W U_2 B_N U_2^*) dW dU_2 - 2R^2(3\delta + \delta^2),$$

(2.43)

where $W$ is an $N_2 \times N_2$ unitary/orthogonal matrix following Haar measure. By our construction, the spectral measure of $A_2$ converges to $\nu(1(|x| \leq 1/\epsilon) / \int_{|x| \leq 1/\epsilon} d\nu$, and

$$dW \left( \nu(1(|x| \leq 1/\epsilon) / \int_{|x| \leq 1/\epsilon} d\nu \right) = O(\epsilon R).$$

Thanks to Cauchy’s Interlacing Theorem, the eigenvalues of $U_2 B_N U_2^*$ and $B_N$ are interlaced. Moreover, we have that $N_2 \geq N - 2\epsilon R N$. The spectral measure of $(N/N_2) U_2 B_N U_2^*$ is close to the spectral measure of $\mu$ in Wasserstein distance as defined in (1.6),

$$dW \left( \mu(N/N_2) U_2 B_N U_2^* \right) = O(\epsilon R).$$

For the integral on the right hand side of (2.43), we can first integrate out $W$, and use Proposition 2.6,

$$\text{Tr}(A_2 W U_2 B_N U_2^*) dW dU_2 = \frac{1}{\beta N^2} \log \int_{\mathcal{B}_δ} \frac{d\mathcal{N}}{e} \text{Tr}(A_2 W U_2 B_N U_2^*) dW dU_2 \geq \frac{1}{\beta N^2} \log \int_{\mathcal{B}_δ} \frac{d\mathcal{N}}{e} \text{Tr}(A_2 W ((N/N_2) U_2 B_N U_2^*) W^*) dW dU_2 \geq \frac{1}{\beta N^2} \log \int_{\mathcal{B}_δ} \nu^{N_2}(I(\nu^ε, \mu) + C_R \alpha(1) + o_N(1)) dU_2 \geq I(\nu^ε, \mu) + C_R \alpha(1) + o_N(1) - C(\delta) R \epsilon$$

(2.44)

Therefore (2.43) and (2.44) together implies that

$$I(\nu, \mu) = \lim_{N \to \infty} \frac{1}{\beta N^2} \log \int e^{\frac{d\mathcal{N}}{e}} \text{Tr}(A_N U_B U^*) dU \geq \frac{1}{2} \int_{|x| > 1/\epsilon} T_ν T_μ dμ + I(\nu^ε, \mu) - C_R \alpha(1),$$

(2.45)

provided we take $ε$ much smaller than $δ$. The estimates (2.37) and (2.45) together conclude the proof of Proposition 2.12.
Proposition 2.13. Given two probability measures $T$, define the multiplicative operator $T: [0, 1] \mapsto \mathbb{R}$:

$$T_N(x) = \sum_{i=1}^{N} a_i 1_{[\frac{i-1}{N}, \frac{i}{N})}(x).$$

From the definition, the empirical eigenvalue distribution $\mu_N = \frac{1}{N} \sum \delta_{\lambda_i}$ of $A_N$ is the push forward measure of $\text{unif}[0, 1]$ by $T_N$. We rearrange $a_1, a_2, \cdots, a_N$ in increasing order: $a_1 \leq a_2 \leq \cdots \leq a_N$, and define the multiplicative operator

$$T_N(x) = \sum_{i=1}^{N} a_i 1_{[\frac{i-1}{N}, \frac{i}{N})}(x).$$

Then $T_N$ is a right continuous nondecreasing function. Moreover, if $F_{A_N}$ is the cumulative density of the empirical eigenvalue distribution $\mu_{A_N}$, then $T_{A_N}$ is the functional inverse of $F_{A_N}$. More generally for any measurable function $T_A: [0, 1] \mapsto \mathbb{R}$, we denote the measure $\mu_A = (T_A)_\# \text{unif}[0, 1]$ the pushforward of $\text{unif}[0, 1]$ by $T_A$. $F_A$ the cumulative density of $\mu_A$ and $T_A$ the functional inverse of $F_A$, which is right continuous and non-decreasing.

A sequence of measures $\mu_{A_N}$ converges weakly to $\mu_A$ if and only if $T_{A_N}$ converges to $T_A$ at all continuous point of $T_A$. And $\mu_{A_N}$ converges in Wasserstein distance (1.6) to $\mu_A$ if and only if $T_{A_N}$ converges to $T_A$ in $L^1$ norm.

Let $A_N, B_N$ be two sequences of deterministic self-adjoint matrices, such that their spectral measures $\mu_{A_N}$ and $\mu_{B_N}$ converge in Wasserstein distance (1.6) towards $\mu_A$ and $\mu_B$ respectively. We assume that there exists a constant $\bar{R} > 0$, such that $\mu_{A_N}(|x|) \leq \bar{R}$ and supp $\mu_{B_N} \subset [-\bar{R}, \bar{R}]$. As a consequence of Theorem 2.8, for any bounded Lipschitz function $f: \mathbb{R} \mapsto \mathbb{R}$,

$$\lim_{N \to \infty} \tau_N(f(A_N)UB_NU^*) = \tau(f(a)b) = \tau(\tau(f(a)b|a)) = \tau(f(a)\tau(b|a)), \quad (2.46)$$

where $U$ follows the law (2.12). The goal of this section is to characterize the derivative of the spherical integral using the non-commutative distribution $\tau$. Indeed, we have

$$\frac{1}{\beta N^2} \ln \int e^{\frac{\beta N}{2} \tau((A_N + \varepsilon C_N)UB_NU^*)} dU \bigg|_{\varepsilon = 0} = \frac{1}{2N} \int \frac{\tau(C_N U B_N U^*) e^{\frac{\beta N}{2} \tau(A_N U B_N U^*)} dU}{\int e^{\frac{\beta N}{2} \tau(A_N U B_N U^*)} dU} \cdot (2.47)$$

Proposition 2.13. Given two probability measures $\mu_A, \mu_B$, such that $\mu_A(|x|) < \infty$ and $\mu_B$ is compactly supported, for any compactly supported and Lipschitz real-valued function $f$, it holds

$$\partial_\varepsilon I(T_A + \varepsilon f(T_A), T_B)|_{\varepsilon = 0} = \frac{1}{2} \int f(x) \tau(b|a)(x) d\mu_A(x). \quad (2.48)$$

If the measure $\mu_A$ has a delta mass at $a$, for any bounded measurable function $\tilde{T}_C$ supported on $\{x : T_A(x) = a\}$, it holds

$$\partial_\varepsilon I(T_A + \varepsilon \tilde{T}_C, T_B)|_{\varepsilon = 0} = \frac{1}{2} \tau(b|a)(a) \int \tilde{T}_C(x) dx. \quad (2.49)$$

We remark that thanks to $(2.19)$ and $(2.16)$, we can express the conditional expectation in terms of the solution $(\rho^*_t, u^*_t)$ of the variational problem $(2.7)$ by $\tau(b|a)(x) = u^*_0(x) - H \rho^*_0(x) + x$. 

22
Proof. Since the spherical integral $I_N(A_N, B_N)$ is convex in both $A_N$ and $B_N$, its limit $I(T_A, T_B)$ is also convex in $T_A$ and $T_B$. Especially for any sequence $C_N$ of deterministic self-adjoint matrices, such that their spectral measures $\mu_{C_N}$ converge weakly towards $\mu_C$, both $I_N(A_N + \varepsilon C_N, B_N)$ and $I(T_A + \varepsilon T_C, T_B)$ are convex in $\varepsilon$. Then for sufficiently small $\varepsilon > 0$,
\[
\partial_\varepsilon I_N(A_N + \varepsilon C_N, B_N) \geq \frac{I_N(A_N + \varepsilon C_N, B_N) - I_N(A_N, B_N)}{\varepsilon} \geq \partial_\varepsilon I_N(A_N + \varepsilon C_N, B_N)|_{\varepsilon=0}.
\]  
(2.50)

Thanks to (2.46) and (2.47), we have that if $C = f(A_N)$, then
\[
\lim_{N \to \infty} \partial_\varepsilon I_N(A_N + \varepsilon f(A_N), B_N)|_{\varepsilon=0} = \frac{1}{2}\tau(f(a)b).
\]  
(2.51)

For the lefthand side of (2.50), we have
\[
\partial_\varepsilon I_N(A_N + \varepsilon f(A_N), B_N) = \frac{1}{2N} \int \frac{\text{Tr}(f(A_N)UB_NU^*)e^{-\frac{\varepsilon N}{2}\text{Tr}((A_N + \varepsilon f(A_N))UB_NU^*)}}{\int e^{\frac{\varepsilon}{2N}\text{Tr}((A_N + \varepsilon f(A_N))UB_NU^*)}dU} dU
\]

\[
= \frac{1}{2N} \int \frac{\text{Tr}(f(A_N)UB_NU^*)e^{-\frac{\varepsilon N}{2}\text{Tr}((A_N + \varepsilon f(A_N))UB_NU^*)}}{\int e^{\frac{\varepsilon}{2N}\text{Tr}((A_N + \varepsilon f(A_N))UB_NU^*)}dU} + O(\varepsilon),
\]
provided $f$ is compactly supported and Lipschitz. By taking the limit $N \to \infty$, we get
\[
\lim_{N \to \infty} \partial_\varepsilon I_N(A_N + \varepsilon f(A_N), B_N) = \frac{1}{2}\tau^\varepsilon(f(a)b) + O(\varepsilon),
\]
where $\tau^\varepsilon$ is the limit of the joint law of $(A_N + \varepsilon f(A_N), UBNU^*)$ with the normalized trace $\text{Tr}()/N$ under the deformed measure $d\mu_{A_N + \varepsilon f(A_N), B_N}(U)$.

The following claim states that $\tau^\varepsilon$ is continuous in $\varepsilon$ close to zero. As a consequence, by taking the limit $N \to \infty$ in (2.50) and combining with (2.51), we get
\[
\frac{1}{2}\tau(f(a)b) + O(1) \geq \frac{I(T_A + \varepsilon f(T_A), T_B) - I(T_A, T_B)}{\varepsilon} \geq \frac{1}{2}\tau(f(a)b).
\]
(2.52)

By a similar argument, we also have the estimate (2.52) for $\varepsilon \leq 0$. The first claim (2.48) follows by taking $\varepsilon$ to zero.

**Claim 2.14.** Under the assumptions of Proposition 2.13, we have
\[
\lim_{\varepsilon \to 0} \tau^\varepsilon(f(a)b) = \tau(f(a)b).
\]
(2.53)

The non-commutative conditional expectation $\tau(b|a)$ is uniquely determined by $(\rho^\varepsilon(x))_{\varepsilon \in [0, 1]}$, and hence by $(\mu_A, \mu_B)$.

Proof. We want to check the continuity of $\tau^\varepsilon(f(a)b)$ in $\varepsilon$. We denote by $(\rho^\varepsilon)^*$ the limit empirical particle density of nonintersecting Brownian bridges with boundary data $A_N + \varepsilon f(A_N)$ and $B_N$. The limit exists thanks to Claim 2.9. We let $\varphi^\varepsilon$ (resp. $\varphi$) denote the joint law of $(a, b, s)$ when $s$ is a semi-circular law free from $(a, b)$ with law $\tau^\varepsilon$ (resp. $\tau$). Then $\rho^\varepsilon_t$ is the law of $ta + (1-t)b + \sqrt{t(1-t)}s$ under $\varphi$, and $(\rho^\varepsilon)^*$ is the law of $ta + (1-t)b + \sqrt{t(1-t)}s$ under $\varphi^\varepsilon$. Since $f$ is bounded, from the proof of Claim 2.9 we have that
\[
\lim_{\varepsilon \to 0} (\rho^\varepsilon)^* = \rho^*_t,
\]
(2.54)
in the weak sense. But we have seen how to compute $\tau^\varepsilon((1/z - a)^{-1}b)$ for $z \in \mathbb{C}\setminus \mathbb{R}$ from $(\rho^\varepsilon)^*$ in (2.31). This is clearly a continuous operation, from which the convergence of $\tau^\varepsilon((z-a)^{-1}b)$ follows. Since $f$ goes to zero at infinity (as it is compactly supported), we can approximate it by linear sums of $(z-x)^{-1}$ and conclude. 

\[\square\]
In the following we deal with the second case, that the measure \( \mu_A \) has a delta mass at \( a \) with \( \mu_A(a) = m \), and \( T_C \) is supported on that \( \{ x : T_A(x) = a \} \). We take a sequence of diagonal matrices \( A_N = \text{diag} \{ a_1, a_2, \cdots , a_N \} \) with non-decreasing diagonal entries, with empirical eigenvalue distribution \( \mu_{A_N} \) converging to \( \mu_A \). Moreover, \( a_i = a \) for \( i \in \lceil \alpha N + 1, (\alpha + m)N \rceil \). We also take the sequence of diagonal matrices \( C_N = \text{diag} \{ c_1, c_2, \cdots , c_N \} \) (not necessarily non-decreasing), with empirical eigenvalue distribution \( \mu_{C_N} \) converging to \( \mu_C \). Moreover, \( c_i = 0 \) for \( i \not\in \lceil \alpha N + 1, (\alpha + m)N \rceil \). We can write \( C_N \) in the block form \( 0_{\alpha N} \oplus C_N \oplus 0_{(1-\alpha-m)N} \) where \( 0_{\alpha N} \) and \( 0_{(1-\alpha-m)N} \) are zero matrices of sizes \( \alpha N \) and \( (1-\alpha-m)N \) respectively, and \( C_N \) is an \( mN \times mN \) diagonal matrix. We take \( P_N \) to be the projection operator onto \( i \in \lceil \alpha N + 1, (\alpha + m)N \rceil \) entries, and \( P_N \) the \( mN \times N \) rectangular matrix consisting of the \( [\alpha N + 1, (\alpha + m)N] \) rows of \( P_N \). It is also a function of \( A_N \): \( P_N = 1_{x=a}(A_N) \). Since \( U \) follows the Haar measure on orthogonal/unitary group, if we multiply \( U \) by the block diagonal matrix \( I_{\alpha N} \oplus V \oplus I_{(1-\alpha-m)N} \) where \( I_{\alpha N} \) and \( I_{(1-\alpha-m)N} \) are identity matrices of sizes \( \alpha N \) and \( (1-\alpha-m)N \) respectively, and \( V \) is an \( mN \times mN \) orthogonal/unitary matrix. Then we have

\[
\int \text{Tr}(C_N U B_N U^*) e^{\frac{AN}{N}} \text{Tr}((A_N + \epsilon C_N) U B_N U^*) dU = \int e^{\frac{AN}{N}} \text{Tr}(A_N U B_N U^*) \int e^{\frac{AN}{N}} \text{Tr}(V C_N V^* P_N U B_N U^* P_N^*) dV dU,
\]

\[
\int e^{\frac{AN}{N}} \text{Tr}((A_N + \epsilon C_N) U B_N U^*) dU = \int e^{\frac{AN}{N}} \text{Tr}(A_N U B_N U^*) \int e^{\frac{AN}{N}} \text{Tr}(V C_N V^* P_N U B_N U^* P_N^*) dV dU.
\]

We can rewrite (2.47) as

\[
\partial_{\epsilon} I_N(A_N + \epsilon C_N, B_N) = \frac{1}{2N} \int \text{Tr}(C_N U B_N U^*) e^{\frac{AN}{N}} \text{Tr}((A_N + \epsilon C_N) U B_N U^*) dU
\]

\[
= \frac{1}{2N} \int e^{\frac{AN}{N}} \text{Tr}(A_N U B_N U^*) \int e^{\frac{AN}{N}} \text{Tr}(V C_N V^* P_N U B_N U^* P_N^*) dV dU.
\]

For the integral over \( V \) conditionally to \( U \), we use the results [?, Theorem 0.1] to find that for \( \epsilon \) small enough, \( N \) large enough,

\[
\frac{1}{2N} \int \text{Tr}(V C_N V^* P_N U B_N U^* P_N^*) dV \int e^{\frac{AN}{N}} \text{Tr}(V C_N V^* P_N U B_N U^* P_N^*) dV dU
\]

\[
= \frac{\text{Tr}(C_N)}{2mN} \frac{\text{Tr}(P_N U B_N U^* P_N)}{N} + o_{\epsilon, N}(1).
\]

We recall that \( P_N \) is a function of \( A_N \): \( P_N = 1_{x=a}(A_N) \). By plugging (2.56) into (2.55), we get that

\[
\partial_{\epsilon} I_N(A_N + \epsilon C_N, B_N) = \frac{\text{Tr}(C_N)}{2mN^2} \int \text{Tr}(1_{x=a}(A_N) U B_N U^*) e^{\frac{AN}{N}} \text{Tr}((A_N + \epsilon C_N) U B_N U^*) dU
\]

\[
= \frac{\text{Tr}(C_N)}{2mN} \frac{\text{Tr}(P_N U B_N U^* P_N)}{N} + o_{\epsilon, N}(1).
\]

By taking the limit \( N \rightarrow \infty \), we deduce

\[
\lim_{N \rightarrow \infty} \partial_{\epsilon} I_N(A_N + \epsilon f(A_N), B_N) = \frac{\text{Tr}(C_N)}{2\mu_A(a)} \frac{\tau^f(1_{x=a}(a)b)}{\mu_A(a)} + o_{\epsilon}(1).
\]

By taking the limit \( N \rightarrow \infty \) in (2.50), thanks to Claim 2.53, (2.57) implies that

\[
\frac{1}{2} \int \frac{\text{Tr}(C_N)}{\mu_A(a)} \frac{\tau(1_{x=a}(a)b)}{\tau^f(1_{x=a}(a)b)} + o_{\epsilon}(1) \geq \frac{1}{\epsilon} \frac{I(T_A + \epsilon T_C, T_B) - I(T_A, T_B)}{\epsilon} \geq \frac{1}{2} \frac{\text{Tr}(C_N)}{\mu_A(a)} \frac{\tau^f(1_{x=a}(a)b)}{\mu_A(a)}.
\]

By a similar argument, we also have the estimate (2.58) for \( \epsilon \leq 0 \). The second claim (2.49) follows by taking \( \epsilon \) to zero. This finishes the proof of Proposition 2.13.

\[
\square
\]
2.4 Continuity at the boundary

In this section, we obtain a more precise description of the solutions \((\rho^*, u^*)\) of the variational problem (2.6) by transforming the complex Burgers equation (2.14) into a Beltrami equation. Due to some technical reason, in this section we assume that the boundary data \(\nu, \mu\) are compactly supported.

Observe that \((\partial_x - i\partial_t) f = \partial_x f + if \partial_t f = (1 + if) \partial_x f = i(f - i) \partial_x f\) and \((\partial_x + i\partial_t) f = \partial_x f - if \partial_t f = (1 - if) \partial_x f = -i(f + i) \partial_x f\). Thus, \((\partial_x - i\partial_t) f = \frac{f-i}{f+i}(\partial_x + i\partial_t) f\). Recall that \(\text{Im}[f] > 0\) for all \((t, x) \in \Omega\) as defined in (2.10), so that \(|(i-f)/(i+f)| < 1\) for all \((t, x) \in \Omega\). Let

\[
\begin{align*}
  z &= x - it, \quad \bar{z} = x + it, \quad (t, x) \in \Omega. \\
  \partial_k f &= \frac{1-f}{i+f} \partial_x f, \quad (\text{Im}[z], \text{Re}[z]) \in \Omega. \
\end{align*}
\]

Then the above shows that (2.14) is equivalent to the Beltrami equation (see, for instance, [?]),

\[
\partial_k f = \frac{1-f}{i+f} \partial_x f, \quad (\text{Im}[z], \text{Re}[z]) \in \Omega.
\]

In general, the Beltrami equation \(\partial_k f = \mu(z) \partial_x f\) is defined on an open set \(\Omega\) on which the measurable function \(\Omega \ni z \mapsto \mu(z) \in \mathbb{C}\) satisfies \(\|\mu\|_{\infty} \leq k < 1\) for a constant \(k\). Under these conditions, it is known that (2.60) has a quasiconformal solution (in particular, a homeomorphism) and that any other solution is obtained by composing it with an analytic map. Since \(|(i-f)/(i+f)| \not\in k\) on \(\Omega\), for any \(k < 1\), it would appear that the general theory for obtaining solutions of Beltrami equations does not apply. However, the regularity of \(\mu = (i-f)/(i+f)\) and the fact that \(\Omega = \bigcup_{\epsilon > 0} \{\text{Im}[f] > \epsilon\}\) allows us easily to conclude that the solutions of (2.60) have essentially all the useful properties of quasiregular maps.

**Theorem 2.15.** Let \(f\) be as in (2.14), with boundary condition \(\nu, \mu\), such that they are compactly supported. Then for \(\nu^{ac}\)-almost all \(x \in \mathbb{R}\), we have \(\lim_{t \to 0} f(t, x) = f(0, x)\), where \(\nu^{ac}\) is the absolutely continuous part of \(\nu\).

**Corollary 2.16.** We assume that the probability measures \(\mu, \mu', \nu\) are compactly supported and neither of them is concentrated in one point. We denote the solutions of the variational problems \(I(\nu, \mu)\) and \(I(\nu, \mu')\) from (2.6) by \(f^{\nu, \mu}(t, x)\) and \(f^{\nu, \mu'}(t, x)\) (as in (2.14)) respectively. If in each connected component of \(\nu\), there is a set of non-zero \(\nu^{ac}\)-measure such that \(f^{\nu, \mu}(0, x) = f^{\nu, \mu'}(0, x)\) for all \(x\) in that set, then we have \(\mu = \mu'\).

We postpone the proof of Theorem 2.15 and Corollary 2.16 to Section 7. As by-products of the proof of Theorem 2.15, we also prove a maximum principle of \(f(t, x)\) and a precise description on the topology of the set \(\Omega\), which might be of independent interest.

3 Large Deviation Principle for \(UBU^*\)

We recall from Theorem 1.1, \(B_N = \text{diag}\{b_1, b_2, \cdots, b_N\}\) is a sequence of deterministic self-adjoint matrices such that the spectral measures \(\mu_{B_N} = (1/N) \sum_{i=1}^N \delta_{b_i}\) of \(B_N\) converge weakly towards \(\mu_B\) as \(N \to \infty\). In this section, we use the spherical integral to study the large deviation principle of the law \(\mathbb{P}_N\) of the empirical measure

\[
\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{(UB_N U^*)_{i1}},
\]

and prove Theorem 1.1.
3.1 Study of the rate function

The classical Schur-Horn theorem [?] states that the diagonal entries of $UB_NU^*$ are in the permutation polytope generated by $(b_1, b_2, \ldots, b_N)$, or equivalently

$$\int_0^1 (T_{\mu} - T_{\mu_B})dx = 0, \quad \int_y^1 (T_{\mu} - T_{\mu_B})dx \leq 0, \quad (3.1)$$

for any $0 \leq y \leq 1$. We recall the functions $H_\mu^D(\cdot)$ and $T_\mu^D(\cdot)$ from Theorem 1.1, $T_\mu^D(\mu) = \sup_{\nu \in M} H_\mu^D(\nu)$. In the following proposition we study these functions, and show the rate function $T_\mu^D(\mu)$ equals $+\infty$ outside the admissible set $A_{\mu_B}$ of probability measures $\mu$ described by the limiting Schur-Horn theorem (3.1):

$$\int_0^1 (T_{\mu} - T_{\mu_B})(x)dx = 0, \quad \int_y^1 (T_{\mu} - T_{\mu_B})(x)dx \leq 0 \quad \forall y \in [0, 1]. \quad (3.2)$$

Proposition 3.1. Under the assumptions of Theorem 1.1, the function $H_\mu^D(\cdot)$ and rate function $T_\mu^D(\cdot)$ as defined in Theorem 1.1 satisfy:

1. For $\mu$ satisfying (3.2), $H_\mu^D(\cdot)$ is upper semi-continuous in weak topology on $\{\nu \in M : \nu(|x|) \leq R\}$ for any $R > 0$. If we view $H_\mu^D(\nu)$ as a function of $T_\nu$, i.e. $H_\mu^D(T_\nu) := H_\mu^D(\nu)$, then it is concave.

2. If $\int_y^1 (T_{\mu}(x) - T_{\mu_B}(x))dx \neq 0$, or there exists some $0 < y < 1$ such that

$$\int_y^1 (T_{\mu}(x) - T_{\mu_B}(x))dx > 0, \quad (3.3)$$

then $T_\mu^D(\mu) = +\infty$.

3. If there exists some small constant $\epsilon > 0$

$$\int_y^1 (T_{\mu}(x) - T_{\mu_B}(x))dx \leq \begin{cases} -cy, & \text{for } 0 \leq y \leq \epsilon, \\ -c, & \text{for } \epsilon < y \leq 1 - \epsilon, \\ -c(1 - y), & \text{for } 1 - \epsilon < y \leq 1, \end{cases} \quad (3.4)$$

then $T_\mu^D(\mu) = H_\mu^D(\nu^*) < \infty$ for some probability measure $\nu^*$ such that $\nu^*(|x|) < \infty$.

4. $T_\mu^D(\cdot)$ is nonnegative and lower semicontinuous on $M([-\infty, \infty])$ (hence it is a good rate function). It vanishes only at the Dirac mass at $\int x d\mu_B$.

5. For any measure $\mu$ in the admissible set $A_{\mu_B}$ as defined in (3.2), there exists a sequence of measures $\mu^\epsilon$ inside the region as given in (3.4), converging to $\mu$ in weak topology and $\lim_{\epsilon \to 0} T_\mu^D(\mu^\epsilon) = T_\mu^D(\mu)$.

Proof. For Item 1, unfortunately, $H_\mu^D(\cdot)$ is not continuous in the weak topology, it is only continuous in the Wasserstein metric. However, $H_\mu^D(\cdot)$ is upper semi-continuous in weak topology on $\{\nu \in M : \nu(|x|) \leq R\}$ for any $R > 0$. Given a probability measure $\nu$ we denote the truncated measure $\nu^\delta = \nu 1(|x| \leq \delta^{-1}) + \delta_0 \int_{|x| > \delta^{-1}} d\nu$.

Claim 3.2. If $\mu$ is a probability measure supported on $[-\infty, \infty]$ which satisfies (3.2), for any probability measure $\nu$ with $\nu(|x|) \leq R$, it holds

$$H_\mu^D(\nu) \leq H_\mu^D(\nu^\delta) + C_R \omega_\delta(1),$$

where the implicit error $\omega_\delta(1)$ is independent of the measure $\nu$.  

26
Proof. We recall the definition of $H^D_\mu (\nu)$ from (1.9) $$H^D_\mu (\nu) = \frac{1}{2} \int T_{\nu}(x)T_{\mu}(x)dx - I(\nu, \mu_B)$$ $$= \frac{1}{2} \int T_{\nu}(x)T_{\mu}(x)dx - \left( \frac{1}{2} \int_{|T_{\nu}|>1/\delta} T_{\nu}(x)T_{\mu_B}(x)dx + I(\nu^\delta, \mu_B) + C_\delta o_\delta(1) \right)$$ $$= \frac{1}{2} \int_{|T_{\nu}|\leq 1/\delta} T_{\nu}(x)T_{\mu}(x)dx - I(\nu^\delta, \mu_B) + \frac{1}{2} \int_{|T_{\nu}|>1/\delta} T_{\nu}(x)(T_{\mu}(x) - T_{\mu_B}(x))dx + C_\delta o_\delta(1)$$ $$\leq \frac{1}{2} \int_{|T_{\nu}|\leq 1/\delta} T_{\nu}(x)T_{\mu}(x)dx - I(\nu^\delta, \mu_B) + C_\delta o_\delta(1)$$ $$= \frac{1}{2} \int T_{\nu}(x)T_{\mu}(x)dx - I(\nu^\delta, \mu_B) + C_\delta o_\delta(1) = H^D_\mu (\nu^\delta) + C_\delta o_\delta(1),$$ where we used Proposition 2.12 in the second line. In the fourth line, we used assumption 3.2 and the fact that $x \rightarrow T_{\nu}(x)1_{T_{\nu}(x)>1/\delta}$ is increasing to show that the last term in the third line is non-positive. Finally, we used that $|T_{\nu}(x)| \leq R$ in the last line.

Let $\{\nu_n\}_{n \geq 1}$ be a sequence of probability measures with $\nu_n(|x|) \leq R$ converging weakly to $\nu$. Take $\delta > 0$ sufficiently small, such that $\nu(|(\delta^{-1},-\delta^{-1})|) = 0$. It is easy to see that $\nu^\delta$ converges to $\nu$ in Wasserstein metric as $\delta \rightarrow 0$. As a consequence, we get

$$H^D_\mu (\nu) = H^D_\mu (\nu^\delta) + o_{\delta, \nu}(1).$$ (3.5)

Moreover, we have that $\nu_n^\delta$ converges to $\nu^\delta$ in Wasserstein distance. Thus it gives

$$\limsup_{n \rightarrow \infty} H^D_\mu (\nu_n^\delta) = H^D_\mu (\nu^\delta).$$ (3.6)

It follows from combining (3.5), Claim 3.2 and (3.6),

$$\limsup_{n \rightarrow \infty} H^D_\mu (\nu_n) \leq \limsup_{n \rightarrow \infty} H^D_\mu (\nu_n^\delta) + C_{R}\nu \circ o_\delta(1)$$

$$= H^D_\mu (\nu^\delta) + C_{R\nu} \circ o_\delta(1) = H^D_\mu (\nu) + C_{R\nu} \circ o_\delta(1) + o_{\delta, \nu}(1),$$

By sending $\delta$ to 0 in the above estimate, we get that

$$\limsup_{n \rightarrow \infty} H^D_\mu (\nu_n) \leq H^D_\mu (\nu),$$

and the upper semicontinuity of $H^D_\mu$ follows.

Both $\int T_{\nu}(x)T_{\mu}(x)dx$ and $- I(T_{\nu}, T_{\mu_B})$ are concave, so is $H^D_\mu (T_{\nu}).$

For Item 2, given any measure $\mu_Y$, we denote its dilation by a factor $L$ as $\mu_{LY} = \mu Y$, then $T_{\mu_{LY}} = LT_{\mu_Y}$. Thanks to Proposition 2.11, for any $\epsilon > 0$, there exists a constant $C(\epsilon)$,

$$\frac{L}{2} \int (T_{\mu} - T_{\mu_B})T_{\mu_Y}dx \leq H^D_\mu (\mu_{LY}) \leq \frac{L}{2} \int (T_{\mu} - T_{\mu_B})T_{\mu_Y}dx + L O(\epsilon)\nu_Y(|x|) + C(\epsilon).$$ (3.7)

If $\int_0^1 (T_{\mu} - T_{\mu_B})dx \neq 0$, we can take $\mu_Y = \delta_1$, then $T_{\mu_Y} = 1_{[0,1]}$, and

$$T^D(\mu) \geq \lim_{L \rightarrow \infty} \max \{H^D_\mu (\mu_{LY}), H^D_\mu (\mu_{-LY}) \} = \lim_{L \rightarrow \infty} \frac{L}{2} \int_0^1 (T_{\mu} - T_{\mu_B})dx = +\infty.$$
If (3.3) holds for some 0 < y < 1, we can take \( \mu_Y = y\delta_0 + (1-y)\delta_{1/(1-y)} \), then \( T_{\mu_Y} = 1_{[y,1)}/(1-y) \), and

\[
T^D(\mu) \geq \lim_{L \to \infty} H^D_{\mu}(\mu_{LY}) = \lim_{L \to \infty} \frac{L}{2(1-y)} \int_y^1 (T_\mu - T_{\mu_B}) \, dx = +\infty.
\]

For Item 3, we remark that the function \( \nu \mapsto H^D_{\mu}(\nu) \) is translation invariant when \( \int_0^1 (T_\mu - T_{\mu_B}) \, dx = 0 \). If \( \nu \) has finite first moment, we can always translate \( \nu \) to make \( \int xd\nu = 0 \). In the rest of the proof, we will restrict ourselves to the set of measures in \( \mathcal{M} \) with mean zero. We will first show that (3.4) implies that there exists a small \( \delta > 0 \) and a large \( L^* > 0 \) such that for any \( \mu_Y \in \mathcal{M} \) such that \( \int |x|d\mu_Y = 1 \), \( \int xd\mu_Y = 0 \) and any \( L \geq L^* \), then \( H^D_{\mu}(\mu_{LY}) \leq -\delta L \). Moreover, note that the set of measures \( \mu_Y \) such that \( T_{\mu_Y} \) is differentiable is dense in \( \mathcal{M} \). Hence, by continuity of \( H^D_{\mu} \), we may assume that \( \mu_Y \) is such that \( T_{\mu_Y} \) is differentiable. Given such a \( \mu_Y \), integration by parts yields

\[
\int_0^1 (T_\mu - T_{\mu_B}) T_{\mu_Y} \, dx = \int_0^1 T'_{\mu_Y}(y) \int_y^1 (T_\mu(x) - T_{\mu_B}(x)) \, dx \, dy.
\]

Since \( T_{\mu_Y} \) is non-decreasing, \( T'_{\mu_Y} \) is non-negative, and we deduce from (3.4) that

\[
\int_0^1 (T_\mu - T_{\mu_B}) T_{\mu_Y} \, dx \leq -\epsilon \int_0^1 (y1_{[0,\epsilon]}(y) + 1_{[\epsilon,1-\epsilon]}(y) + (1-y)1_{[1-\epsilon,1]}(y)) T'_{\mu_Y}(y) \, dy. \quad (3.8)
\]

On the other hand, because \( \int_0^1 |T_{\mu_Y}|(x) \, dx = 1 \) and \( \int_0^1 T_{\mu_Y}(x) \, dx = 0 \), and thanks to the smoothness and the monotonicity of \( T_{\mu_Y} \), we know that there exists \( y_0 \in [0,1] \) such that \( T_{\mu_Y}(y_0) = 0 \) and then by integration by parts

\[
\int_0^1 |T_{\mu_Y}(y)| \, dy = \int_{y_0}^{y_0} yT'_{\mu_Y}(y) \, dy + \int_{y_0}^1 (1-y)T'_{\mu_Y}(y) \, dy = 1.
\]

We deduce that one of the two terms above is greater or equal than 1/2. Hence, since for \( y \in [0,1] \)

\[
y1_{[0,\epsilon]}(y) + 1_{[\epsilon,1-\epsilon]}(y) + (1-y)1_{[1-\epsilon,1]}(y) \geq \max\{y1_{[0,y_0]}(y), (1-y)1_{[y_0,1]}(y)\},
\]

we get from (3.8) the upper bound

\[
\int_0^1 (T_\mu - T_{\mu_B})(x) T_{\mu_Y}(x) \, dx \leq -\epsilon. \quad (3.9)
\]

We use (3.7) and (3.9) to estimate \( H^D_{\mu}(\mu_{LY}) \). As a consequence, if we take \( \epsilon \) much smaller than \( \epsilon \), (3.7) implies that there exists a small \( \delta > 0 \) and a large \( L^* > 0 \) (depending only on \( \epsilon \)) such that for any \( L \geq L^* \), it holds \( H^D_{\mu}(\mu_{LY}) \leq -\delta L \). We conclude that

\[
\sup_{\nu \in \mathcal{M}} H^D_{\mu}(\nu) = \sup_{\mu_Y : \int |x|d\mu_Y \leq L^*} H^D_{\mu}(\mu_Y) < \infty,
\]

and the supremum is achieved at some \( \nu^* \) with \( \int |x|d\nu^* \leq L^* \), since \( \{\mu_Y : \int |x|d\mu_Y \leq L^*\} \) is compact and \( H^D_{\mu} \) is upper semicontinuous.

For Item 4, since \((\mu, \nu) \mapsto H^D_{\mu}(\nu) \) is continuous in \( \mu \), \( \mathcal{I}^D(\mu) = \sup_{\nu \in \mathcal{M}} H^D_{\mu}(\nu) \) is lower semicontinuous. Moreover \( \mathcal{I}^D(\mu) \geq H^D_{\mu}(\delta_0) = 0 \), so \( \mathcal{I}^D(\cdot) \) is nonnegative.

If \( \mathcal{I}^D(\mu) = 0 \), then \( \int xd\mu = \int xd\mu_B \) and \( H^D_{\mu}(\nu) \leq 0 \) for all probability measures \( \nu \in \mathcal{M} \),

\[
\frac{1}{2} \int T_\mu T_\nu \, dx \leq I(\nu, \mu_B). \quad (3.10)
\]
We denote $\nu_\epsilon = \epsilon \# \nu$ the pushforward of $\nu$ by the homothety of factor $\epsilon$, and then $T_{\nu_\epsilon} = \epsilon T_\nu$. [7, Theorem 0.1] implies that for $\epsilon > 0$ small enough

$$I(\nu_\epsilon, \mu_B) = \frac{\epsilon}{2} \int x d\mu_B \int x d\nu + O(\epsilon^2).$$

Hence, we deduce from (3.10) by replacing $\nu$ by $\nu_\epsilon$ and sending $\epsilon$ to zero that

$$\int T_\mu(x) T_\nu(x) dx \leq \int x d\mu_B \int x d\nu = \int x d\mu_B \int T_\nu dx,$$

or equivalently for any probability measure $\nu \in \mathcal{M}$

$$\int \left( T_\mu(x) - \int x d\mu_B \right) T_\nu(x) dx \leq 0.$$

Taking $T_\nu = 1_{\{x: T_\mu(x) \leq \int x d\mu_B\}}$, we deduce that $T_\mu(x) \leq \int x d\mu_B$ almost surely. On the other hand, $\int T_\mu dx = \int x d\mu_B$, and therefore $T_\mu = \int x d\mu_B$ almost surely. We conclude that if $T(\mu) = 0$ then $T_\mu = \int x d\mu_B$ almost surely and $\mu$ is the delta mass at $\int x d\mu_B$.

Finally for the second point of Item 5, we pick $\mu \in \mathcal{A}_{\mu_B}$ and construct $\mu^\epsilon$ satisfying (3.4) converging to $\mu$ when $\epsilon$ goes to zero. If $\mu$ is not a delta mass, we have for small enough $\epsilon > 0$, $T_\mu(1 - \epsilon) > T_\mu(\epsilon) + 2\epsilon$.

We take

$$T_{\mu^\epsilon}(y) = \begin{cases} T_\mu(y) + \epsilon & \text{for } y \in [0, \epsilon], \\ T_\mu(y) + \epsilon & \text{for } y \in [\epsilon, \epsilon_1], \\ T_\mu(y) & \text{for } y \in [\epsilon_1, \epsilon_2], \\ T(1 - \epsilon) - \epsilon & \text{for } y \in [\epsilon_2, 1 - \epsilon], \\ T_\mu(y) - \epsilon & \text{for } y \in [1 - \epsilon, 1], \end{cases}$$

where

$$\epsilon_1 = \sup \{ x > \epsilon : T_\mu(\epsilon) + \epsilon \geq T_\mu(x) \}, \quad \epsilon_2 = \sup \{ x : T_\mu(x) \leq T_\mu(1 - \epsilon) - \epsilon \}.$$

Then we get $T_{\mu^\epsilon}$ by shifting $T_{\mu^\epsilon}$ such that its first moment is the same as $T_\mu$:

$$T_{\mu^\epsilon}(y) = T_{\mu^\epsilon}(y) - \Delta, \quad \Delta := \int_0^1 (T_{\mu^\epsilon}(y) - T_{\mu^\epsilon}(y)) dy.$$

We can check $\Delta = \int_{\epsilon_1}^{\epsilon_1} (T_\mu(\epsilon) + \epsilon - T_\mu(y)) dy + \int_{\epsilon_2}^{1 - \epsilon} (T_\mu(1 - \epsilon) - \epsilon - T_\mu(y)) dy$. On the interval $[\epsilon, \epsilon_1]$, $T_\mu$ is non decreasing, it holds that on this interval $0 \leq T_\mu(\epsilon) + \epsilon - T_\mu(y) \leq \epsilon$. Therefore $0 \leq \int_{\epsilon}^{\epsilon_1} (T_\mu(\epsilon) + \epsilon - T_\mu(y)) dy \leq \epsilon(\epsilon_1 - \epsilon)$. Similarly we have $-\epsilon(1 - \epsilon - \epsilon_2) \leq \int_{\epsilon_2}^{1 - \epsilon} (T_\mu(1 - \epsilon) - \epsilon - T_\mu(y)) dy \leq 0$. As a consequence, we have $|\Delta| \leq \epsilon - \epsilon^2$, and also that $\mu^\epsilon$ goes to $\mu$ as $\epsilon$ goes to zero.

We claim that $\mu^\epsilon$ satisfies (3.4). We denote for all $y \in [0, 1]$,

$$\varphi(y) := \int_y^1 (T_\mu(x) - T_{\mu^\epsilon}(x)) dx \leq \int_y^1 (T_{\mu_B}(x) - T_{\mu^\epsilon}(x)) dx.$$

From the construction $\varphi(0) = \varphi(1) = 0$, and $\varphi(y)$ first decreases then increases on $[0, 1]$. Moreover, for $y \in [0, \epsilon]$,

$$\varphi(y) = -\int_0^y (T_\mu(x) - T_{\mu^\epsilon}(x)) dx = -\int_0^y [(T_\mu(x) - T_{\mu^\epsilon}(x)) + (T_{\mu^\epsilon}(x) - T_{\mu}(x))] dx = (\epsilon - \Delta)y.$$

For $y \in [1 - \epsilon, 1]$, we have

$$\varphi(y) = \int_0^1 [(T_\mu(x) - T_{\mu^\epsilon}(x)) + (T_{\mu^\epsilon}(x) - T_{\mu}(x))] dx = (\epsilon + \Delta)(1 - y).$$

29
And for $y \in [\varepsilon, 1 - \varepsilon]$,

$$
\varphi(y) = \int_{y}^{1} (T_{\mu}(x) - T_{\nu}(x))\,dx \geq \varepsilon^2 - |\Delta|\varepsilon.
$$

Therefore, $\mu^\varepsilon$ satisfies (3.4) with $c = \varepsilon^2 - |\Delta|\varepsilon \geq \varepsilon^3 > 0$. We finally prove that $\mathcal{I}^D(\mu^\varepsilon)$ goes to $\mathcal{I}^D(\mu)$ as $\varepsilon$ goes to zero. By lower semi-continuity of $\mathcal{I}^D$, we already know that

$$
\mathcal{I}^D(\mu) \leq \liminf_{\varepsilon \to 0} \mathcal{I}^D(\mu^\varepsilon). \tag{3.12}
$$

For the converse bound, note that for all $\nu \in \mathcal{M}$, integration by parts and (3.11) imply that

$$
\int T_{\nu}(T_{\mu} - T_{\nu}) \,dx = \int T_{\nu}'(y) \int_{y}^{1} (T_{\mu}(x) - T_{\nu}(x))\,dx\,dy \geq 0,
$$

which results in

$$
H^D_{\mu}(\nu) = \int T_{\nu}(T_{\mu}(x) - I(\nu, \mu)B) \geq \int T_{\nu}(T_{\mu}(x) - I(\nu, \mu)B) \,dx = H^D_{\mu}(\nu).
$$

As a consequence, we have

$$
\mathcal{I}^D(\mu) \geq \mathcal{I}^D(\mu^\varepsilon),
$$

and therefore

$$
\mathcal{I}^D(\mu) \geq \limsup_{\varepsilon \to 0} \mathcal{I}^D(\mu^\varepsilon). \tag{3.13}
$$

The claim follows from combining (3.12) and (3.13).

\[\square\]

### 3.2 Large deviation upper bound

In this section we prove the large deviation upper bound in Theorem 1.1. We first notice that if $\Delta$ is the simplex $\{y_N \in \mathbb{R}^N : y_1 \geq y_2 \geq \cdots \geq y_N\}$, since the law of $U$ is permutation invariant as well as $\mu_N$,

$$
P_N(\mu_N \in \cdot) = N! P_N(\{\mu_N \in \cdot\} \cap \{(UB_NU^*)_{1 \leq i \leq N} \in \Delta\}).
$$

We estimate the probability of a small $\delta$-neighborhood $\mathcal{B}_\delta(\mu)$ of $\mu$, by tilting the measure as follows:

$$
P_N(\mathcal{B}_\delta(\mu)) = N! \mathbb{E}\left[1(\{\mu_N \in \mathcal{B}_\delta(\mu)\} \cap \{(UB_NU^*)_{1 \leq i \leq N} \in \Delta\})\frac{\exp((\beta/2)N \text{Tr}(Y_NUB_NU^*)}{\exp((\beta/2)N \text{Tr}(Y_NUB_NU^*)}ight]. \tag{3.14}
$$

where $Y_N = \text{diag}\{y_1, y_2, \cdots, y_N\}$ is a sequence of diagonal matrices, with $y_N \in \Delta$ and its spectral measure converging in Wasserstein distance (1.6) towards $\mu_Y \in \mathcal{M}$ (we can take $y_1, y_2, \cdots, y_N$ the $N$-quantiles of the measure $\mu_Y$). We notice that when $\mu_N$ is in the neighborhood $\mathcal{B}_\delta(\mu)$, and the diagonal entries of $(UB_NU^*)$ and $Y_N$ are both in $\Delta$, the integrand of the spherical integral is approximately

$$
\exp((\beta/2)N \text{Tr}(Y_NUB_NU^*)) = \exp\left\{\frac{\beta N^2}{2} \left(\int T_{\nu} dx + o_3(1)\right)\right\}, \tag{3.15}
$$

$$
\begin{align*}
\text{exp}(\beta N^2/2) &\sim o_3(1) \\
\text{as } N \to \infty.
\end{align*}
$$
The estimates (3.14) and (3.15) give a large deviation upper bound for the random measure $\mu_N$ as follows

$$\mathbb{P}_N(\mathcal{B}_B(\mu)) = N! \mathbb{E} \left[ \left\{ (\mu_N \in \mathcal{B}_B(\mu)) \cap \left\{ ((UB_NU^*)_{ii})_{1 \leq i \leq N} \in \Delta \right\} \right\} \right] \exp \left\{ (\beta/2)N \text{Tr}(YNUB_NU^*) \right\}$$

$$\leq N! \exp \left\{ -\beta N^2 \left( \frac{1}{2} \int T_{\mu Y_N} T_{\mu} dx + o_1(1) \right) \right\} \mathbb{E} \left[ (\mu_N \in \mathcal{B}_B(\mu)) \exp \left\{ (\beta/2)N \text{Tr}(YNUB_NU^*) \right\} \right]$$

$$= N! \exp \left\{ -\beta N^2 \left( \frac{1}{2} \int T_{\mu Y} T_{\mu} dx + o_1(1) + o_N(1) \right) \right\} \mathbb{E} \left[ \exp \left\{ (\beta/2)N \text{Tr}(YNUB_NU^*) \right\} \right]$$

$$= N! \exp \left\{ -\beta N^2 \left( \frac{1}{2} \int T_{\mu Y} T_{\mu} dx - I(\mu_Y, \mu_B) + o_1(1) + o_N(1) \right) \right\}.$$  (3.16)

It follows by taking the large $N$ limit, then $\delta$ going to zero and taking the infimum on the right hand side of (3.16), we get the following large deviation upper bound

$$\limsup_{\delta \to 0} \limsup_{N \to 0} \frac{1}{\beta N^2} \log \mathbb{P}_N(\mathcal{B}_B(\mu)) \leq - \sup_{\mu_Y \in \mathcal{M}} H^D_\mu(\mu_Y) = - I^D(\mu).$$  (3.17)

### 3.3 Large deviation lower bound

In this section we derive the large deviation lower bound for the empirical measure of the diagonal entries of $UB_NU^*$, which matches the upper bound (3.17). The large deviation lower bound follows from combining the following Propositions 3.3 and 3.4.

**Proposition 3.3.** We assume the assumptions of Theorem 1.1. For any probability measure $\mu_Y \in \mathcal{M}$, there exists a unique $\mu$ supported on $[-\mathcal{A}, \mathcal{A}]$ such that

$$\mu_Y \in \arg \sup_{\nu \in \mathcal{M}} H^D_\nu(\nu), \quad H^D_\mu(\nu) = \frac{1}{2} \int T_\nu T_\mu dx - I(\nu, \mu_B).$$  (3.18)

Here, $T_\mu$ is uniquely determined by $T_Y$ by

$$T_\mu = \tau(b|y) \circ T_Y.$$

Here, $\tau(b|y)$ is the conditional expectation of $b$ knowing $y$ under the non-commutative distribution $\tau$ uniquely associated to $(\mu_B, \mu_Y)$ as in Theorem 2.8.

Observe that the above shows that $\tau(b|y) \circ T_Y$ is non-decreasing for any $\mu_Y$: this is coherent with the fact that we can always reorder the $(UB_NU^*)_{ii}$ up to neglactable factors $N$! and that the density of the tilted measure is maximal when the $A_{ii}$ and $(UB_NU^*)_{ii}$ are both increasing.

**Proposition 3.4.** We assume the assumptions of Theorem 1.1. For any probability measure $\mu_Y \in \mathcal{M}$, let $\mu$ be the unique measure supported on $[-\mathcal{A}, \mathcal{A}]$ so that

$$\mu_Y \in \arg \sup_{\nu \in \mathcal{M}} H^D_\nu(\nu), \quad H^D_\mu(\nu) = \frac{1}{2} \int T_\nu T_\mu dx - I(\nu, \mu_B).$$

Then we have

$$\liminf_{\delta \to 0} \liminf_{N \to \infty} \frac{1}{\beta N^2} \log \mathbb{P}_N(\mathcal{B}_B(\mu)) \geq -H^D_\mu(\mu_Y) = - I^D(\mu).$$  (3.19)

**Proof of Theorem 1.1.** Item 1 of Theorem 1.1 follows from Proposition 3.1. For Item 2, the large deviation upper bound follows from (3.17). If $\mu$ does not satisfy $\int T_\mu(x) - T_{\mu_B}(x) dx \neq 0$ or the limiting Schur-Horn inequalities (1.11), then both sides of (1.12) are $-\infty$. There is nothing to prove. In the
following we first prove (1.12) when \( \mu \) satisfies \( \int_0^1 (T_\nu(x) - T_{\mu_B}(x))dx = 0 \) and the strong limiting Schur-Horn inequalities (3.4) with some \( \epsilon > 0 \). In this case, thanks to Item 3 in Proposition 3.1, there exists a probability measure \( \mu_Y \) such that \( T^D(\mu) = H^D_2(\mu_Y) < \infty \) and \( \mu_Y \in \mathcal{M} \). Then Propositions 3.3 and 3.4 imply that \( \mu \) is uniquely determined by \( \mu_Y \) and the large deviation lower bound holds. This gives the full large deviation principle when the strong limiting Schur-Horn inequalities (3.4) hold. Next we extend it to the boundary case by a continuity argument. Thanks to Item 5 in Proposition 3.1, for any measure \( \mu \) inside the admissible set but not satisfying (3.4), there exists a sequence of measures \( \mu^\epsilon \) inside the region as given in (3.4), converging to \( \mu \). In this case, thanks to Item 3 in Proposition 3.1, there exists a positive constant \( c(\delta) > 0 \).

Proposition 3.5. We assume the assumptions of Theorem 1.1. Let \( Y_N = \text{diag}\{y_1, y_2, \ldots, y_N\} \) be a sequence of diagonal matrices whose spectral measures converge in Wasserstein distance (1.6) towards \( \mu_Y \in \mathcal{M} \) and a probability measure \( \mu \) supported on \([-\alpha, \alpha] \), such that

\[
\sup_{\nu \in \mathcal{M}} \left\{ \frac{1}{2} \int T_\nu T_\mu dx - I(\nu, \mu_B) \right\} > \frac{1}{2} \int T_{\mu_Y} T_\mu dx - I(\mu_Y, \mu_B),
\]

then there exists a small \( \delta > 0 \), and positive constant \( c(\delta) > 0 \) such that

\[
\mathbb{E}[\mathbf{1}(\mu_N \in \mathcal{B}_\delta(\mu)) \exp\{((\beta/2)N \text{Tr}(Y_N U B_N U^*))\}] \leq e^{-c(\delta)N^2} \mathbb{E}[\exp\{((\beta/2)N \text{Tr}(Y_N U B_N U^*))\}].
\]

Proof of Proposition 3.5. Under the assumption (3.21), for sufficiently small \( \epsilon > 0 \), there exists a measure \( \nu \in \mathcal{M} \) such that

\[
\frac{1}{2} \int T_\nu T_\mu dx - I(\nu, \mu_B) \geq \frac{1}{2} \int T_{\mu_Y} T_\mu dx - I(\mu_Y, \mu_B) + \epsilon.
\]

The large deviation lower bound follows by first sending \( \delta \) to zero and then \( \epsilon \) to zero in the right hand side of (3.20). This finishes the proof of Theorem 1.1.

The proofs of both Propositions 3.3 and 3.4 rely on the following probability estimate.

\[
\liminf_{N \to \infty} \frac{1}{\beta N^2} \log \mathbb{P}(\mu_N \in \mathcal{B}_\delta(\mu)) \geq \liminf_{N \to \infty} \frac{1}{\beta N^2} \log \mathbb{P}(\mu_N \in \mathcal{B}_{\delta/2}(\mu^\epsilon)) = T^D(\mu^\epsilon) + o_\delta(1).
\]

We divide \( \exp\{\beta N^2(I(\mu_Y, \mu_B) + o_N(1))\} \) on both sides of (3.22), and take a small \( \delta > 0 \) which will be chosen later,

\[
\exp\left\{ -\beta N^2(I(\mu_Y, \mu_B) + o_N(1)) \right\} \mathbb{E}[\mathbf{1}(\mu_N \in \mathcal{B}_\delta(\mu)) \exp\{((\beta/2)N \text{Tr}(Y_N U B_N U^*))\}]
\]

\[
= \exp\left\{ -\beta N^2 \left( I(\mu_Y, \mu_B) - \frac{1}{2} \int T_{\mu_Y} T_\mu dx + o_\delta(1) + o_N(1) \right) \right\} \mathbb{E}[\mathbf{1}(\mu_N \in \mathcal{B}_\delta(\mu))]
\]

\[
\leq \exp\left\{ -\beta N^2 \left( I(\mu_Y, \mu_B) - \frac{1}{2} \int T_{\mu_Y} T_\mu dx - I(\nu, \mu_B) + \frac{1}{2} \int T_\nu T_\mu dx + o_\delta(1) + o_N(1) \right) \right\}
\]

\[
\leq \exp\left\{ -\beta N^2 \left( \epsilon + o_\delta(1) + o_N(1) \right) \right\},
\]

where in the first inequality we used the large deviation upper bound (3.17), and (3.23) in the last inequality. The claim follows provided we take \( \delta \) sufficiently small and \( N \) large.
Proof of Proposition 3.3. We first prove the existence of such \( \mu \) by contradiction. If there is no such \( \mu \), i.e. for any measure \( \mu \) supported on \([-\mathbb{R}, \mathbb{R}]\), we have

\[
\mu_Y \notin \text{arg sup}_{\nu \in \mathcal{M}} \left\{ \frac{1}{2} \int T_\nu T_\mu dx - I(\nu, \mu_B) \right\}.
\]

It follows that for any measure \( \mu \) supported on \([-\mathbb{R}, \mathbb{R}]\), it holds

\[
\sup_{\nu \in \mathcal{M}} \left\{ \frac{1}{2} \int T_\nu T_\mu dx - I(\nu, \mu_B) \right\} > \frac{1}{2} \int T_\mu T_\mu dx - I(\mu_Y, \mu_B).
\]

Then Proposition 3.5 implies that there exists a small \( \delta \), and positive constant \( c(\delta) > 0 \) such that

\[
\mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \exp\{((\beta)/2)N \text{Tr}(Y_NUB_NU^*)\} \right] \leq e^{-c(\delta)N^2} \mathbb{E} \left[ \exp\{((\beta)/2)N \text{Tr}(Y_NUB_NU^*)\} \right].
\]

Since the space of probability measures supported on \([-\mathbb{R}, \mathbb{R}]\) is compact, we get a finite open cover \( \cup \mathcal{B}_\delta(\mu_i) \) of the set of probability measures supported on \([-\mathbb{R}, \mathbb{R}]\),

\[
\mathbb{E} \left[ \exp\{((\beta)/2)N \text{Tr}(Y_NUB_NU^*)\} \right] = \mathbb{E} \left[ \sum_i 1(\mu_N \in \mathcal{B}_\delta(\mu_i)) \exp\{((\beta)/2)N \text{Tr}(Y_NUB_NU^*)\} \right] 
\leq \sum_i e^{-c(\delta_i)N^2} \mathbb{E} \left[ \exp\{((\beta)/2)N \text{Tr}(Y_NUB_NU^*)\} \right] \leq \mathbb{E} \left[ \exp\{((\beta)/2)N \text{Tr}(Y_NUB_NU^*)\} \right] < \mathbb{E} \left[ \exp\{((\beta)/2)N \text{Tr}(Y_NUB_NU^*)\} \right],
\]

for sufficiently large \( N \). This gives a contradiction.

In the following we prove the uniqueness of such measure \( \mu \) satisfying (3.18). Since \( \mu_Y \) is one of the maximizer, then for any \( \epsilon > 0 \),

\[
\frac{1}{2} \int T_Y T_\mu dx - I(T_Y, T_B) \geq \frac{1}{2} \int (T_Y + \epsilon \tilde{T}_C)_{\#} \left( \text{uni}[0,1] \right) T_\mu dx - I(T_Y + \epsilon \tilde{T}_C, T_B) \\
\geq \frac{1}{2} \int (T_Y + \epsilon \tilde{T}_C) T_\mu dx - I(T_Y + \epsilon \tilde{T}_C, T_B).
\]

By rearranging the above expression, and sending \( \epsilon \) to 0, we have

\[
\partial_t I(T_Y + \epsilon \tilde{T}_C, T_B) \bigg|_{\epsilon = 0} \geq \frac{1}{2} \int \tilde{T}_C T_\mu dx.
\] (3.24)

We will choose \( \tilde{T}_C \) in either the case (2.48) or (2.49). We notice that in both cases if we replace \( \tilde{T}_C \) by \(-\tilde{T}_C\), both sides of (3.24) change the sign. Therefore, we conclude that

\[
\partial_t I(T_Y + \epsilon \tilde{T}_C, T_B) \bigg|_{\epsilon = 0} = \frac{1}{2} \int \tilde{T}_C T_\mu dx.
\]

Now if we choose \( \tilde{T}_C \) in (2.49), i.e. \( \tilde{T}_C \) supported on that \( \{x : T_Y(x) = a\} \), we have

\[
\partial_t I(T_Y + \epsilon \tilde{T}_C, T_B) \bigg|_{\epsilon = 0} = \frac{1}{2} \int T_C(x) dx \tau(\mathcal{B}|y)(a) = \frac{1}{2} \int \tilde{T}_C T_\mu dx,
\]

We conclude that \( T_\mu(x) = \tau(\mathcal{B}|y)(a) = \tau(\mathcal{B}|y) \circ T_Y(x) \) on \( \{x : T_Y(x) = a\} \). Especially, on the intervals \( T_Y \) is a constant, we have \( T_\mu(x) = \tau(\mathcal{B}|y) \circ T_Y(x) \). Next we take \( T_C = f(T_Y) \) as in (2.48),

\[
\partial_t I(T_Y + \epsilon f(T_Y), T_B) \bigg|_{\epsilon = 0} = \frac{1}{2} \int f(x) \tau(\mathcal{B}|y)(x) d\mu_Y \\
= \frac{1}{2} \int f(T_Y) \tau(\mathcal{B}|y) \circ T_Y(x) dx = \frac{1}{2} \int f(T_Y) T_\mu dx.
\] (3.25)
Because the complement of the open ball $B$ is compact, we get a finite open cover.

As a consequence the assumption in Proposition 3.5 holds.

**Proof of Proposition 3.4.** Thanks to the uniqueness of $\mu$, we have that for any $\mu' \neq \mu$ supported on $[-R,R]$

$$\mu_Y \notin \arg \sup_{\nu \in \mathcal{M}} \left\{ \frac{1}{2} \int T_\nu T_\mu' \, dx - I(\nu,\mu_B) \right\}.$$ 

As a consequence the assumption in Proposition 3.5 holds

$$\sup_{\nu \in \mathcal{M}} \left\{ \frac{1}{2} \int T_\nu T_\mu' \, dx - I(\nu,\mu_B) \right\} > \frac{1}{2} \int T_{\mu_Y} T_\mu' \, dx - I(\mu_Y,\mu_B).$$ 

It follows from Proposition 3.5, there exists a small $\delta > 0$, and positive constant $c(\delta) > 0$ such that

$$\mathbb{E}\left[ 1(\nu \in \mathbb{B}_\delta(\mu')) \exp\{(\beta/2)N \text{Tr}(Y_N U_B N U^*)\} \right] \leq e^{-c(\delta)N^2} \mathbb{E}\left[ \exp\{(\beta/2)N \text{Tr}(Y_N U_B N U^*)\} \right].$$ 

Because the complement of the open ball $\mathbb{B}_\delta(\mu)$ in the space of probability measures supported in $[-R,R]$ is compact, we get a finite open cover $\cup \mathbb{B}_\delta(\mu_i)$,

$$\mathbb{E}\left[ 1(\nu \in \mathbb{B}_\delta(\mu)) \exp\{(\beta/2)N \text{Tr}(Y_N U_B N U^*)\} \right] \geq \mathbb{E}\left[ \exp\{(\beta/2)N \text{Tr}(Y_N U_B N U^*)\} \right] - \sum_i \mathbb{E}\left[ 1(\mu_N \in \mathbb{B}_\delta(\mu_i)) \exp\{(\beta/2)N \text{Tr}(Y_N U_B N U^*)\} \right]

= \left( 1 - \sum_i e^{-c(\delta_i)N^2} \right) \mathbb{E}\left[ \exp\{(\beta/2)N \text{Tr}(Y_N U_B N U^*)\} \right]$$

$$\geq \left( 1 - \sum_i e^{-c(\delta_i)N^2} \right) \mathbb{E}\left[ \exp\{(\beta/2)N \text{Tr}(Y_N U_B N U^*)\} \right] \left( 1 - \sum_i o_N(1) \right).$$

The large deviation lower bound at $\mu$ follows from (3.14) and the estimate (3.26)

$$\mathbb{P}_N(\mathbb{B}_\delta(\mu)) = N! \exp\left\{ -\beta N^2 \left( \frac{1}{2} \int T_{\mu_Y} T_\mu \, dx + o_4(1) \right) \right\} \mathbb{E}\left[ 1(\mu_N \in \mathbb{B}_\delta(\mu)) \exp\{(\beta/2)N \text{Tr}(Y_N U_B N U^*)\} \right]

\geq \exp\left\{ -\beta N^2 \left( \frac{1}{2} \int T_{\mu_Y} T_\mu \, dx + o_4(1) + o_N(1) \right) \right\} \exp\left\{ \beta N^2 (I(\mu_Y,\mu_B) + o_N(1)) \right\}

= \exp\left\{ -\beta N^2 \left( \frac{1}{2} \int T_{\mu_Y} T_\mu \, dx - I(\mu_Y,\mu_B) + o_4(1) + o_N(1) \right) \right\}.$$

\[ \square \]

4 **Large Deviation Estimates for $A + U_B U^*$**

We recall from Theorem 1.3, $A_N, B_N$ are two $N \times N$ diagonal matrices with eigenvalues $a_1 \geq a_2 \geq \cdots \geq a_N$ and $b_1 \geq b_2, \cdots \geq b_N$. In this section, we use the spherical integral to study the large deviation of the law $\mathbb{P}_N$ of the empirical eigenvalue distribution of $A_N + U_B N U^*$,

$$\mu_N = \frac{1}{N} \sum_{i=1}^N \Lambda_i(A_N + U_B N U^*),$$

34
and prove Theorem 1.3. Besides the relations \( \text{Tr} A_N + \text{Tr} B_N = \text{Tr}(A_N + UB_N U^*) \), and the Ky Fan inequalities,

\[
\sum_{i=1}^k a_i + \sum_{i=1}^k b_i \geq \sum_{i=1}^k \lambda_i (A_N + UB_N U^*), \quad 1 \leq i \leq N, \tag{4.1}
\]

Horn [?] had conjectured the form of a set of necessary and sufficient inequalities to be satisfied for the eigenvalues of \( A_N + UB_N U^* \). After contributions by several authors, see in particular [?], these conjectures were proven by Knutson and Tao [?, ?].

### 4.1 Study of the rate function

In the following proposition we study the rate function \( I^{A+B}(\cdot, \cdot) \) from Theorem 1.3. Clearly, it is a good rate function again by Proposition 2.6. Unfortunately, it does not capture the admissible set for the possible eigenvalues given by Horn’s problem. However it contains the information about the constraints given by the Ky Fan inequalities, i.e. it equals \( +\infty \) outside the region described by the limiting Ky Fan inequalities:

\[
\int_0^1 (T_{\mu} - T_{\mu_A} - T_{\mu_B}) \, dx = 0, \quad \int_y^1 (T_{\mu} - T_{\mu_A} - T_{\mu_B}) \, dx \leq 0, \quad \forall y \in [0, 1]. \tag{4.2}
\]

**Proposition 4.1.** Under the assumptions of Theorem 1.3, the function \( H_{\mu}^{A+B}(\cdot) \) and rate function \( I^{A+B}(\cdot) \) as defined in Theorem 1.3 satisfy:

1. For measure \( \mu \) satisfies (4.2), \( H_{\mu}^{A+B}(\cdot) \) is upper semi-continuous in weak topology on \( \{ \nu \in \mathcal{M} : \nu(|x|) \leq R \} \) for any \( R > 0 \).
2. If \( \int_0^1 (T_{\mu} - T_{\mu_A} - T_{\mu_B}) \, dx \neq 0 \), or there exists some \( 0 < y < 1 \) such that

\[
\int_y^1 (T_{\mu} - T_{\mu_A} - T_{\mu_B}) \, dx > 0, \tag{4.3}
\]

then \( I^{A+B}(\mu) = +\infty \).
3. If there exists some small constant \( c > 0 \)

\[
\int_y^1 (T_{\mu} - T_{\mu_A} - T_{\mu_B}) \, dx \leq \begin{cases} -cy, & \text{for } 0 \leq y \leq c, \\ -c, & \text{for } c \leq y \leq 1 - c, \\ -c(1 - y), & \text{for } 1 - c \leq y \leq 1, \end{cases} \tag{4.4}
\]

then \( I^{A+B}(\mu) = H_{\mu}^{A+B}(\nu^*) < \infty \) for some probability measure \( \nu^* \) such that \( \nu^*([x]) < \infty \).
4. The rate function \( I^{A+B}(\cdot) \) is nonnegative and lower semicontinuous on \( \mathcal{M}([-2\delta, 2\delta]) \).

**Proof.** For Item 1, we first prove

**Claim 4.2.** Under the assumption (4.2), for any probability measure \( \nu \) with \( \nu(|x|) \leq R \) and probability measure \( \mu \) supported on \([-R, R] \), then setting \( \nu^* = \nu 1(|x| \leq \delta^{-1}) + \delta_0 \int_{|x| > \delta^{-1}} \, d\nu \), we have

\[
H_{\mu}^{A+B}(\nu) \leq H_{\mu}^{A+B}(\nu^*) + C_\delta \omega_3(1), \tag{4.5}
\]

where the implicit error \( \omega_3(1) \) is independent of the measure \( \nu \).
Proof. We recall the definition of \( H_{µ}^{A+B}(ν) \) from (1.13) and (2.36):

\[
H_{µ}^{A+B}(ν) = I(ν, µ) - I(ν, µ_A) - I(ν, µ_B)
\]

\[
= \frac{1}{2} \int_{|x| > 1/δ} T_ν(T_µ - T_{µ_A} - T_{µ_B})dx + (I(µ^δ, µ) - I(ν^δ, µ_A) - I(ν^δ, µ_B)) + C_δ o_δ(1) \tag{4.6}
\]

\[
≤ I(µ^δ, µ) - I(ν^δ, µ_A) - I(ν^δ, µ_B)) + C_δ o_δ(1) = H_{µ}^{A+B}(ν^δ) + C_δ o_δ(1),
\]

where we used Proposition 2.12 in the second line, and Assumption 4.2 in the last line (with the remark that \( x \to 1/|T_ν(x)| > 1/δT_ν(x) \) is non-decreasing).

Let \( \{ν_n\}_{n≥1} \) be a sequence of probability measures with \( ν_n(|x|) \) \( ≤ \) \( R \), converging weakly to \( ν \). Take \( δ > 0 \) sufficiently small, such that \( ν(\{-δ^{-1}, -δ^{-1}\}) = 0 \). Then, \( ν^δ \) converges to \( ν \) in Wasserstein metric as \( δ \to 0 \). Moreover (4.6) shows

\[
H_{µ}^{A+B}(ν) = H_{µ}^{A+B}(ν^δ) + o_δ(1). \tag{4.7}
\]

Moreover, \( ν^δ \) converges to \( ν^δ \) in Wasserstein distance, so that

\[
\limsup_{n→∞} H_{µ}^{A+B}(ν^δ_n) = H_{µ}^{A+B}(ν^δ). \tag{4.8}
\]

It follows from combining (4.7), Claim 4.2 and (4.8),

\[
\limsup_{n→∞} H_{µ}^{A+B}(ν_n) ≤ \limsup_{n→∞} H_{µ}^{A+B}(ν^δ_n) + C_{R, ν} o_δ(1)
\]

\[
= H_{µ}^{A+B}(ν^δ) + C_{R, ν} o_δ(1) = H_{µ}^{A+B}(ν) + C_{R, ν} o_δ(1) + o_δ(1).
\]

By sending \( δ \) to 0 in the above estimate, the upper semicontinuity of \( H_{µ}^{A+B} \) follows as

\[
\limsup_{n→∞} H_{µ}^{A+B}(ν_n) ≤ H_{µ}^{A+B}(ν).
\]

For Item 2, given any measure \( µ_Y \), we denote its dilation by a factor \( L \) as \( µ_{LY} = Lµ_Y \), then \( T_{µ_{LY}} = LT_{µ_Y} \). Thanks to Proposition 2.11, for any ε > 0, there exists a constant \( C(ε) \) such that

\[
H_{µ}^{A+B}(µ_{LY}) = \frac{L}{2} \int (T_µ - T_{µ_A} - T_{µ_B})T_{µ_Y}dx + L O(ε) µ_Y(|x|) + C(ε). \tag{4.9}
\]

If \( \int_0^1 (T_µ - T_{µ_A} - T_{µ_B}) dx \neq 0 \), we can take \( µ_Y = δ_1 \), then \( T_{µ_Y} = 1_{[0,1]} \) and

\[
\mathcal{T}^{A+B}(µ) ≥ \lim_{L→∞} \max \{H_{µ}^{A+B}(µ_{LY}), H_{µ}^{A+B}(µ_{-LY})\}
\]

\[
≥ \lim_{L→∞} \frac{L}{2} \int (T_µ - T_{µ_A} - T_{µ_B})T_{µ_Y}dx + L O(ε) µ_Y(|x|) + C(ε) = ∞.
\]

If (4.3) holds for some \( 0 < y < 1 \), we can take \( µ_Y = yδ_0 + (1-y)δ_{1/(1-y)} \), then \( T_{µ_Y} = 1_{[y,1]/(1-y)} \), and

\[
\mathcal{T}^{A+B}(µ) ≥ \lim_{L→∞} H_{µ}^{A+B}(µ_{LY})
\]

\[
≥ \lim_{L→∞} \frac{L}{2(1-y)} \int_y^1 (T_µ - T_{µ_A} - T_{µ_B})dx + L O(ε) µ_Y(|x|) + C(ε) = ∞,
\]

provided we take ε small enough.

For Item 3, to prove that (4.4) implies that there exists \( ν^* \) \( ∈ \) \( M \) such that \( H_{µ}^{A+B}(ν^*) = \sup_{ν∈M} H_{µ}^{A+B}(ν) \), we need to define another functional

\[
\hat{H}_{µ}^{A+B}(ν) = \frac{1}{2} \int T_νT_µdx - I(ν, µ_A) - I(ν, µ_B),
\]

36
which is an upper bound of $H_{\mu}^{A+B}(\nu)$, i.e., $H_{\mu}^{A+B}(\nu) \leq \tilde{H}_\mu^{A+B}(\nu)$, thanks to Proposition 2.11. We remark that the functions $H_{\mu}^{A+B}(\nu)$ and $\tilde{H}_\mu^{A+B}(\nu)$ are both translation invariant. If $\nu$ has finite first moment and $\int_0^1 (T_\mu - T_{\mu A} - T_{\mu B}) \, dx = 0$, we can always translate $\nu$ to make $\int x \, d\nu = 0$. In the rest of the proof, we will restrict ourselves to the set of measures in $\mathcal{M}$ with mean zero. Moreover, since $I(\mu_{LY}, \mu_A)$ and $I(\mu_{LY}, \mu_B)$ are both convex in $\mu$, the function $\tilde{H}_\mu^{A+B}(\mu_{LY})$ is concave in $\mu$ for any $\mu_Y$. We first prove that under (4.4), there exists a small $\delta > 0$ and a large $L^* > 0$ such that for any $\mu_Y \in \mathcal{M}$ with $\int |x| \, d\mu_Y = 1$, $\int xd\mu_Y = 0$ and any $L \geq L^*$, then $\tilde{H}_\mu^{A+B}(\mu_{LY}) \leq -\delta L$. Given such $a_{\mu_Y}$, there exists some $y_0 \in (0, 1)$ such that $T_{\mu_Y}(y_0) = 0$ and

$$\int_0^1 y_1_{[0, y_0]}(y)T_{\mu_Y}(y) \, dy = \frac{L}{2}, \quad \int_0^1 (1 - y)1_{[y_0, 1]}(y)T_{\mu_Y}(y) \, dy = \frac{L}{2}.$$  

(4.10)

Thanks to Proposition 2.11, we have the following estimates for $\tilde{H}_\mu^{A+B}(\mu_{LY})$,

$$\tilde{H}_\mu^{A+B}(\mu_{LY}) = \frac{L}{2} \int (T_\mu - T_{\mu A} - T_{\mu B}) \, dx + L O(\varepsilon) \mu_Y(|x|) + C(\varepsilon).$$  

(4.11)

Integration by parts yields

$$\int_0^1 (T_\mu - T_{\mu A} - T_{\mu B}) \, dx = \int_0^1 T_{\mu A}^T(y) \int_0^1 (T_\mu(x) - T_{\mu A}(x) - T_{\mu B}(x)) \, dx \, dy$$

$$\leq -c \int_0^1 (y_1_{[0, y]}(y) + 1_{[1 - y]}(y) + (1 - y)1_{[1 - y]}(y))T_{\mu_Y}^T(y) \, dy$$

$$\leq -c \min \left\{ \int_0^1 (y_1_{[0, y]}(y)T_{\mu_Y}^T(y) \, dy, \int_0^1 (1 - y)1_{[y_0, 1]}(y))T_{\mu_Y}^T(y) \, dy \right\} = -c/2,$$

(4.12)

where we used Assumption (4.4), $T_{\mu_Y}$ is non-decreasing, and (4.10). Therefore, if we take $\varepsilon$ much smaller than $c$, (4.11) and (4.12) imply that there exists a small $\delta > 0$ and a large $L^* > 0$ (depending only on $c$) such that for any $L \geq L^*$, it holds $\tilde{H}_\mu^{A+B}(\mu_{LY}) \leq -\delta L$.

Using Proposition 2.11 again, for arbitrarily small $\varepsilon > 0$, we have

$$0 \leq \tilde{H}_\mu^{A+B}(\mu_{LY}) - \tilde{H}_\mu^{A+B}(\mu_{LY}) \leq L O(\varepsilon) \mu_Y(|x|) + C(\varepsilon),$$

Therefore by taking $\varepsilon$ small and $L^*$ large enough, for $L \geq L^*$, $\int |x| \, d\mu_Y = 1$ and $\int xd\mu_Y = 0$ we have

$$H_{\mu}^{A+B}(\mu_{LY}) \leq -\delta L/2.$$

We conclude that

$$\sup_{\nu} H_{\mu}^{A+B}(\nu) = \sup_{\nu, \int |x| \, d\nu \leq L^*} H_{\mu}^{A+B}(\nu) < \infty,$$

and the supremum is achieved at some $\nu^*$ with $\int |x| \, d\nu^* \leq L^*$, since $\{\mu_Y : \int |x| \, d\mu_Y \leq L^*\}$ is compact and $H_{\mu}^{A+B}$ is upper semicontinuous, thanks to Item 1.

For Item 4, since $(\mu, \nu) \mapsto H_{\mu}^{A+B}(\nu)$ is continuous in $\mu$, $Z_{\mu}^{A+B}(\mu) = \sup_{\nu \in \mathcal{M}} H_{\mu}^{A+B}(\nu)$ is lower semicontinuous. Moreover $Z_{\mu}^{A+B}(\mu) \geq H_{\mu}^{A+B}(\delta_0) = 0$, so $Z_{\mu}^{A+B}(\cdot)$ is nonnegative.

4.2 Large deviation upper bound

In this section we prove the large deviation upper bound in Theorem 1.3. We take $Y_N = \text{diag}\{y_1, y_2, \cdots, y_N\}$ a sequence of diagonal matrices, whose spectral measures converge in Wasserstein distance (1.6) towards
where we used Proposition 2.6 that the functional $I$ is continuous in Wasserstein distance. Then we have

$$
P_N(\mathcal{B}_\delta(\mu)) = e^{-\beta N^2 (I(\mu_{\nu}, \mu) + o_\delta(1) + o_N(1))} \mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V (A_N + U B_N U^*) V^*)} dV \right],
$$

(4.13)

where we used Proposition 2.6 that the functional $I(\cdot, \cdot)$ is continuous in Wasserstein distance. Then we have

$$
P_N(\mathcal{B}_\delta(\mu)) = e^{-\beta N^2 (I(\mu_{\nu}, \mu) + o_\delta(1) + o_N(1))} \mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V (A_N + U B_N U^*) V^*)} dV \right]
\leq e^{-\beta N^2 (I(\mu_{\nu}, \mu) + o_\delta(1) + o_N(1))} \mathbb{E} \left[ \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V (A_N + U B_N U^*) V^*)} dV \right] \Rightarrow \text{exp}\{-\beta N^2 (I(\mu_{\nu}, \mu) - I(\mu_{\nu}, \mu_A) - I(\mu_{\nu}, \mu_B) + o_\delta(1) + o_N(1))\}.
$$

(4.14)

It follows by taking the large $N$ limit, then $\delta$ going to zero and taking the infimum on the right hand side of (4.14), we get the following large deviation upper bound

$$
\limsup_{\delta \to 0} \limsup_{N \to \infty} \frac{1}{\beta N^2} \log \mathbb{P}_N(\mathcal{B}_\delta(\mu)) \leq - \sup_{\nu \in \mathcal{M}} H^A+B_\mu(\nu) = -\mathcal{I}^{A+B}(\mu),
$$

(4.15)

where $\mathcal{M}$ is the set of probability measures with bounded first moment.

### 4.3 Large deviation lower bound

In this section we derive the large deviation lower bound for the empirical eigenvalue distribution of $A_N + U B_N U^*$, which matches the upper bound (3.17) on a large set of measures. The large deviation lower bound follows from combining the following Propositions 4.3 and 4.4.

**Proposition 4.3.** We assume the assumptions of Theorem 1.3. Let $\mu_{\nu}$ be compactly supported and has absolutely continuous part in each of its connected components. Then there exists a unique $\mu$ supported on $[-2R, 2R]$ such that

$$
\mu_{\nu} \in \arg \sup_{\nu \in \mathcal{M}} H^A+B_\mu(\nu), \quad H^A+B_\mu(\nu) = I(\nu, \mu) - I(\nu, \mu_A) - I(\nu, \mu_B).
$$

**Proposition 4.4.** We assume the assumptions of Theorem 1.3. If $\mu$ is the unique probability measure supported on $[-2R, 2R]$ with

$$
\mu_{\nu} \in \arg \sup_{\nu \in \mathcal{M}} H^A+B_\mu(\nu), \quad H^A+B_\mu(\nu) = I(\nu, \mu) - I(\nu, \mu_A) - I(\nu, \mu_B),
$$

then we have

$$
\liminf_{\delta \to 0} \liminf_{N \to \infty} \frac{1}{\beta N^2} \log \mathbb{P}_N(\mu_N \in \mathcal{B}_\delta(\mu)) \geq -H^A+B_\mu(\mu_{\nu}) = -\mathcal{I}^{A+B}(\mu).
$$

(4.16)

The proofs of both Propositions 4.3 and 4.4 rely on the following probability estimate.
Proposition 4.5. We assume the assumptions in Theorem 1.3. Let $Y_N = \text{diag}\{y_1, y_2, \cdots, y_N\}$ be a sequence of diagonal matrices whose spectral measures converge in Wasserstein distance (1.6) towards $\mu_Y \in \mathcal{M}$ and a probability measure $\mu$ supported on $[-2R, 2R]$, such that

$$\sup_{\nu \in \mathcal{M}} \{I(\nu, \mu) - I(\nu, \mu_A) - I(\nu, \mu_B)\} > I(\mu_Y, \mu) - I(\mu_Y, \mu_A) - I(\mu_Y, \mu_B),$$

(4.17)

then there exists a small $\delta > 0$, and positive constant $c(\delta) > 0$ such that

$$\mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V (A_N + UB_N U^*) V^*)} dV \right] \leq e^{-c(\delta) N^2} \mathbb{E} \left[ \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V (A_N + UB_N U^*) V^*)} dV \right].$$

(4.18)

Proof of Proposition 4.5. Under Assumption (4.17), for sufficiently small $\varepsilon > 0$, there exists a measure $\nu \in \mathcal{M}$ such that

$$I(\nu, \mu) - I(\nu, \mu_A) - I(\nu, \mu_B) \geq I(\mu_Y, \mu) - I(\mu_Y, \mu_A) - I(\mu_Y, \mu_B) + \varepsilon.$$

(4.19)

The right hand side of (4.18) is

$$\mathbb{E} \left[ \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V (A_N + UB_N U^*) V^*)} dV \right] = \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V A_N V^*)} dV \int e^{\frac{\beta N}{2} \text{Tr}(Y_N U B_N U^*)} dU = \exp\{\beta N^2(I(\mu_Y, \mu_A) + I(\mu_Y, \mu_B) + o_N(1))\}.$$  

We divide $\exp\{\beta N^2(I(\mu_Y, \mu_A) + I(\mu_Y, \mu_B) + o_N(1))\}$ on both sides of (4.18), and take a small $\delta > 0$, which will be chosen later,

$$e^{-\beta N^2(I(\mu_Y, \mu_A) + I(\mu_Y, \mu_B) + o_N(1))} \mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V (A_N + UB_N U^*) V^*)} dV \right]$$

$$= e^{-\beta N^2(I(\mu_Y, \mu_A) + I(\mu_Y, \mu_B) + o_N(1))} \mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \right]$$

$$\leq e^{-\beta N^2(\varepsilon + o_N(1))} \mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \right]$$

$$\leq e^{-\beta N^2(\varepsilon + o_N(1))},$$

where in the first equality we used the large deviation upper bound (4.15), and (4.19) in the last inequality. The claim follows provided we take $\delta$ sufficiently small. \(\square\)

Proof of Proposition 4.3. We prove the existence of such $\mu$ by contradiction. If there is no such $\mu$, for any measure $\mu$ supported on $[-2R, 2R]$, we have

$$\mu_Y \notin \arg \sup_{\nu \in \mathcal{M}} \{I(\nu, \mu) - I(\nu, \mu_A) - I(\nu, \mu_B)\}.$$

It follows that for any measure $\mu$ supported on $[-2R, 2R]$,

$$\sup_{\nu \in \mathcal{M}} \{I(\nu, \mu) - I(\nu, \mu_A) - I(\nu, \mu_B)\} > I(\mu_Y, \mu) - I(\mu_Y, \mu_A) - I(\mu_Y, \mu_B).$$

Hence Proposition 4.5 implies that there exists a small $\delta > 0$, and positive constant $c(\delta) > 0$ such that

$$\mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V (A_N + UB_N U^*) V^*)} dV \right] \leq e^{-c(\delta) N^2} \mathbb{E} \left[ \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V (A_N + UB_N U^*) V^*)} dV \right].$$
Since the space of probability measures supported on $[-2R,2R]$ is compact, we get a finite open cover $\cup B_{\delta_i}(\mu_i)$ of the set of probability measures supported on $[-2R,2R]$,

$$E \left[ \int e^{\frac{\delta N}{2}} \text{Tr}(Y_N V(A_N + UB_N U^*)V^*) dV \right] = E \left[ \sum_i 1(\mu_N \in B_{\delta_i}(\mu_i)) \int e^{\frac{\delta N}{2}} \text{Tr}(Y_N V(A_N + UB_N U^*)V^*) dV \right]$$

$$\leq \sum_i e^{-c(\delta_i)N^2} E \left[ \int e^{\frac{\delta N}{2}} \text{Tr}(Y_N V(A_N + UB_N U^*)V^*) dV \right]$$

$$< E \left[ \int e^{\frac{\delta N}{2}} \text{Tr}(Y_N V(A_N + UB_N U^*)V^*) dV \right],$$

where the last inequality holds for $N$ large enough. This gives a contradiction.

The proof of the uniqueness of $\mu$ in Proposition 4.3 is where we need the regularity of the measure $\mu_Y$. If $\mu_Y$ is a maximizer of $H_{\mu}^{A+B}()$, then

$$0 = \partial_k H_{\mu}^{A+B}(TY + \varepsilon \tilde{T}_C) \big|_{\varepsilon = 0}$$

$$= \partial_k I(TY + \varepsilon \tilde{T}_C, T_\mu) \big|_{\varepsilon = 0} - \partial_k I(TY + \varepsilon \tilde{T}_C, T_\mu) \big|_{\varepsilon = 0} - \partial_k I(TY + \varepsilon \tilde{T}_C, T_B) \big|_{\varepsilon = 0}.$$

We take $\tilde{T}_C = f(T_Y)$ as in (2.48),

$$0 = \partial_k I(T_Y + \varepsilon f(T_Y), T_\mu) \big|_{\varepsilon = 0} - \partial_k I(T_Y + \varepsilon f(T_Y), T_\mu) \big|_{\varepsilon = 0} - \partial_k I(T_Y + \varepsilon f(T_Y), T_B) \big|_{\varepsilon = 0}$$

$$= \int f(x)(\tau(\mu|y)(x) - \tau(a|y)(x) - \tau(b|y)(x)) d\mu_Y(x).$$

(4.20)

Since we can take $f$ any bounded Lipschitz function, we conclude from (4.20) that $\mu_Y$-almost surely for any $x \in \mathbb{R},$

$$\tau(\mu|y)(x) = \tau(a|y)(x) + \tau(b|y)(x).$$

(4.21)

We denote the solutions of the variational problem $I(\mu_Y, \mu), I(\mu_Y, \mu_A), I(\mu_Y, \mu_B)$ as given in (2.7) by $f^{\mu_Y \rightarrow \mu}(t, x), f^{\mu_Y \rightarrow \mu_A}(t, x), f^{\mu_Y \rightarrow \mu_B}(t, x)$ respectively. Then Item 7 in Proposition 2.5 gives that

$$\tau(\mu|y)(x) = \text{Re}[f^{\mu_Y \rightarrow \mu}(0, x)] + H_{\mu_Y} - x,$$

$$\tau(a|y)(x) = \text{Re}[f^{\mu_Y \rightarrow \mu_A}(0, x)] + H_{\mu_Y} - x,$$

$$\tau(b|y)(x) = \text{Re}[f^{\mu_Y \rightarrow \mu_B}(0, x)] + H_{\mu_Y} - x.$$

Therefore (4.21) implies that

$$\text{Re}[f^{\mu_Y \rightarrow \mu}(0, x)] = \text{Re}[f^{\mu_Y \rightarrow \mu_A}(0, x)] + \text{Re}[f^{\mu_Y \rightarrow \mu_B}(0, x)] + H_{\mu_Y} - x,$$

for $\mu_Y$-almost surely all $x \in \mathbb{R}$. We also notice that $f^{\mu_Y \rightarrow \mu}(0, x), f^{\mu_Y \rightarrow \mu_A}(0, x), f^{\mu_Y \rightarrow \mu_B}(0, x)$ have the same imaginary parts (equal to $\pi d\mu_Y/dx$). By our assumption that $\mu_Y$ is compactly supported and has absolutely continuous part in each of its connected components, we conclude from Corollary 2.16 that $\mu$ is uniquely determined by $\mu_Y$ and $f^{\mu_Y \rightarrow \mu}(0, x)$. This finishes the proof of Proposition 4.3.

$\square$

Proof of Proposition 4.4. Thanks to the uniqueness of $\mu$, we have that for any $\mu' \neq \mu$ supported on $[-2R,2R]$

$$\mu_Y \notin \text{arg sup}_{\nu \in M} \{ I(\nu, \mu') - I(\nu, \mu_A) - I(\nu, \mu_B) \}.$$
As a consequence, the assumption in Proposition 4.5 holds

\[ \sup_{\nu \in M} \{ I(\nu, \mu') - I(\nu, \mu_A) - I(\nu, \mu_B) \} > I(\mu_Y, \mu') - I(\mu_Y, \mu_A) - I(\mu_Y, \mu_B). \]

It follows from Proposition 4.5, there exists a small \( \delta > 0 \), and positive constant \( c(\delta) > 0 \) such that

\[ \mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu')) \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V(A_N + U B_N U^*) V^*)} dV \right] \leq e^{-c(\delta)N^2} \mathbb{E} \left[ \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V(A_N + U B_N U^*) V^*)} dV \right]. \]

Again, because the space of probability measures supported on \([-2\tilde{\mathcal{R}}, 2\tilde{\mathcal{R}}]\) after removing the open set \( \mathcal{B}_\delta(\mu) \) is compact, we get a finite open cover \( \bigcup \mathcal{B}_i(\mu_i) \).

\[ \mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V(A_N + U B_N U^*) V^*)} dV \right] \geq \mathbb{E} \left[ \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V(A_N + U B_N U^*) V^*)} dV \right] - \mathbb{E} \left[ \sum_i 1(\mu_N \in \mathcal{B}_i(\mu_i)) \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V(A_N + U B_N U^*) V^*)} dV \right] \]

\[ \geq \left( 1 - \sum_i e^{-c(\delta_i)N^2} \right) \mathbb{E} \left[ \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V(A_N + U B_N U^*) V^*)} dV \right] = \left( 1 - \sum_i e^{-c(\delta_i)N^2} \right) e^{\beta N^2(I(\mu_Y, \mu_A) + I(\mu_Y, \mu_B) + o_N(1))}. \]

The large deviation lower bound at \( \mu \) follows from (4.13) and the estimate (4.22) \( \mathbb{P}_N(\mathcal{B}_\delta(\mu)) = e^{-\beta N^2(I(\mu_Y, \mu_A) + o_N(1))} \mathbb{E} \left[ 1(\mu_N \in \mathcal{B}_\delta(\mu)) \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V(A_N + U B_N U^*) V^*)} dV \right] \geq e^{-\beta N^2(I(\mu_Y, \mu_A) + o_N(1))} e^{\beta N^2(I(\mu_Y, \mu_A) + I(\mu_Y, \mu_B) + o_N(1))} = e^{-\beta N^2(I(\mu_Y, \mu_A) + I(\mu_Y, \mu_B) + o_N(1))}. \)

□

A drawback of our large deviation bound is that we do not know how to prove that the rate function \( T^{A+B} \) has a unique minimizer \( \mu_A \boxplus \mu_B \) (and whether this is true). To circumvent this fact we can improve our large deviation upper bound as follows. From [4], there exists \( \varepsilon > 0 \) such that for any sequence of Hermitian matrices \( Y_N \) such that \( \| Y_N \|_\infty < \varepsilon \) and with spectral measure converging towards \( \mu_Y \), we know that the following limit exists:

\[ \lim_{N \to \infty} \frac{1}{\beta N^2} \log \mathbb{E} \left[ \int e^{\frac{\beta N}{2} \text{Tr}(Y_N V(A_N + U B_N U^*) V^*)} dV \right] = -I(\mu_Y, \mu_A, \mu_B). \]

**Theorem 4.6.** Let \( A_N, B_N \) be a sequence of deterministic self-adjoint matrices, such that their spectral measures \( \hat{\mu}_{A_N}, \hat{\mu}_{B_N} \) converge weakly towards \( \mu_A, \mu_B \) respectively, and there exists a constant \( \tilde{\mathcal{R}} > 0 \), such that \( \text{supp} \hat{\mu}_{A_N}, \text{supp} \hat{\mu}_{B_N} \subset [-\tilde{\mathcal{R}}, \tilde{\mathcal{R}}] \), then the empirical eigenvalue distribution \( \mu_N \) of \( A_N + U B_N U^* \) satisfies a large deviation upper bound with rate function

\[ \hat{T}^{A+B}(\mu) = \max \{ I^{A+B}(\mu), I^{A+B}_-(\mu) \}, \]

where

\[ I^{A+B}_-(\mu) = \sup \{ I(\mu_Y, \mu_A, \mu_B) - I(\mu_Y, \mu) \} \]

where the supremum is taken over probability measures \( \mu_Y \) with support in \([-\varepsilon/\tilde{\mathcal{R}}, \varepsilon/\tilde{\mathcal{R}}]\). Moreover, \( \hat{T}^{A+B}(\cdot) \) is a good rate function and vanishes only at \( \mu_A \boxplus \mu_B \).
Proof. The same reasoning as for the proof of Theorem 1.3 shows that

$$\mathbb{P}_N(\mathcal{B}_d(\mu)) = E \left[ 1(\mu_N \in \mathcal{B}_d(\mu)) \right] = \mathbb{E} \left[ \frac{\exp \left\{ (\beta/2)N \text{Tr}(Y_N V(A_N + U B_N U^*)V^*) \right\} }{\exp \left\{ (\beta/2)N \text{Tr}(Y_N V(A_N + U B_N U^*)V^*) \right\} } \right]$$

which gives the large deviation upper bound. Hence, the only thing to show is that $\tilde{I}^{A+B}(\cdot)$ is non negative and vanishes only at $\mu_A \boxplus \mu_B$. Moreover, if $\mu_Y$ has the distribution of $\varepsilon Y$ with $Y$ uniformly bounded with law $\bar{Y}$ and $\varepsilon > 0$ small enough, by [?], we see that $I(\mu_Y, \mu_A, \mu_B)$ is an absolutely converging series in $\varepsilon$ whose coefficients only depends on the moments of $\bar{Y}, \mu_A, \mu_B$. As a consequence, it is a continuous function of these compactly supported measures. It clearly follows that $\tilde{I}^{A+B}(\cdot)$ is a good rate function. To show that it vanishes only at $\mu_A \boxplus \mu_B$, we use that from [?], p. 38, it is proven that if we take $\nu_\varepsilon = (1 - \tau) \delta_0 + \tau \delta_{\theta}$, with $\tau, \theta$ small enough, then

$$I(\nu_\varepsilon, \mu) = \tau \int_0^\theta R_\mu(t) dt + O(\tau^2),$$

where $R_\mu(\cdot)$ is the $R$-transform of the measure $\mu$. Hence,

$$R^{A+B}_\mu(\nu_\varepsilon) = \tau \int_0^\theta (R_\mu(t) - R_{\mu_A}(t) - R_{\mu_B}(t)) dt + O(\tau^2),$$

which implies that

$$\tilde{I}^{A+B}(\mu) \geq \tau \int_0^\theta (R_\mu(t) - R_{\mu_A}(t) - R_{\mu_B}(t)) dt + O(\tau^2).$$

This implies that if $\tilde{I}^{A+B}(\mu) = 0$ then for sufficiently small $\theta$

$$\int_0^\theta (R_\mu(t) - R_{\mu_A}(t) - R_{\mu_B}(t)) dt \leq 0,$$

which is still verified by many probability measures. Next, we use the symmetrization to show that $\tilde{I}^{A+B} = 0$ implies that the equality holds in (4.23). In fact for $\theta, \tau$ small enough

$$I(\nu_\varepsilon, \mu_A, \mu_B) = \tau \int_0^\theta (R_{\mu_A}(t) + R_{\mu_B}(t)) dt + O(\tau^2).$$

Hence for sufficiently small $\tau, \theta$ we have

$$\tilde{I}^{A+B}(\mu) \geq \tau \left| \int_0^\theta (R_\mu(t) - R_{\mu_A}(t) - R_{\mu_B}(t)) dt \right| + O(\tau^2).$$

By sending $\tau$ to zero, we have that $\tilde{I}^{A+B}(\mu) = 0$ implies that $R_\mu(\theta) = R_{\mu_A}(\theta) + R_{\mu_B}(\theta)$ for all $\theta$ small enough, which further implies that $\mu = \mu_A \boxplus \mu_B$. We conclude that $\tilde{I}^{A+B}(\cdot)$ vanishes only at $\mu_A \boxplus \mu_B$. 

\qed
5 Large Deviation Principle for Kostka numbers

In this section, we use the spherical integral to derive the large deviation estimates of the Kostka numbers and prove Theorem 1.5. From the definition (1.3) of Kostka numbers,

\[ K_{\lambda_N \eta_N} \leq \frac{S_{\lambda_N}(e^{Y_N})}{m_{\eta_N}(e^{Y_N})}, \quad e^{Y_N} = (e^{y_1}, e^{y_2}, \ldots, e^{y_N}), \quad (5.1) \]

where \( Y_N = \text{diag}\{y_1, y_2, \ldots, y_N\} \) is a sequence of diagonal matrices, with \( y_1 \geq y_2 \geq \cdots \geq y_N \) and spectral measure converging in Wasserstein distance (1.6) towards \( \mu_Y \) (we can take \( y_1, y_2, \ldots, y_N \) the \( N \)-quantiles of \( \mu_Y \)). For the monomial symmetric function

\[ m_{\eta_N}(x_N) = \sum_{a_N \sim \eta_N} x_1^{a_1} x_2^{a_2} \cdots x_N^{a_N}, \]

where \( a_N = (a_1, a_2, \ldots, a_N) \sim \eta_N = (\eta_1 \geq \eta_2 \geq \cdots \geq \eta_N) \), if the parts of \( a_N \) is a rearrangement of the parts of \( \eta_N \). We easily see that if \( x_N = e^{Y_N}, Y_N = \text{diag}\{y_1, y_2, \ldots, y_N\} \) with \( y_1 \geq y_2 \geq \cdots \geq y_N \), then

\[ e^{\eta_1 y_1 + \eta_2 y_2 + \cdots + \eta_N y_N} \leq m_{\eta_N}(x_N) = m_{\eta_N}(e^{Y_N}) \leq N! e^{\eta_1 y_1 + \eta_2 y_2 + \cdots + \eta_N y_N}. \quad (5.2) \]

We recall from (1.17) that \( m[\eta_N] = \frac{1}{N^2} \sum \delta\left(\frac{n N - i}{N}\right) \) where \( i \rightarrow \eta_i + N - i \) is non-decreasing. Hence, if \( m[\eta_N] \) goes to \( (T_{\mu})_#(\text{unif}[0, 1]) \), \( \frac{1}{N} \sum \delta\left(\frac{n}{N}\right) \) goes to \( (T_{\mu} - x)_#(\text{unif}[0, 1]) \). This implies, with (5.2), that

\[ \frac{1}{N^2} \log m_{\eta_N}(e^{Y_N}) = \int (T_{\mu} - x)T_{\nu_Y} \, dx + o_2(1) + o_N(1), \quad (5.3) \]

if \( m[\eta_N] \)

\[ m_{\eta_N}(\mu) \]

for some probability measure \( \mu \) such that \( T_{\mu}(x) \geq x \). The large deviation upper bound follows from combining the asymptotics of Schur symmetric polynomials (1.19), (5.1), and (5.3),

\[ \lim_{\delta \rightarrow 0} \sup_{N \rightarrow \infty} \frac{1}{N^2} \log \sup_{m[\eta_N] \in \mathcal{B}_N(\mu)} K_{\lambda_N \eta_N} \leq -H^K_{\mu}(\mu_Y), \quad \text{for } H^K_{\mu}(\nu_Y) = \int (T_{\mu} - x)T_{\nu_Y} \, dx - J(\nu_Y, m_{\lambda}), \quad (5.4) \]

where the functional \( J(\cdot, \cdot) \) is defined in (1.19). Taking the infimum over \( \mu_Y \in \mathcal{M} \) on the right hand side of (5.4) finishes the proof of the large deviation upper bound in Theorem 1.5. It is known that the Kostka number \( K_{\lambda_N \mu_N} \) is positive if and only if \( \lambda_N \) and \( \eta_N \) are of the same size, and \( \lambda_N \) is larger than \( \eta_N \) in dominance order:

\[ \lambda_1 + \lambda_2 + \cdots + \lambda_i \geq \eta_1 + \eta_2 + \cdots + \eta_i, \quad 1 \leq i \leq N. \quad (5.5) \]

We recall from Theorem 1.5 that \( Z^K(\mu) = \sup_{\nu \in \mathcal{M}} H^K_\mu(\nu) \). It turns out that the rate function \( Z^K(\mu) \) equals \(+\infty\) outside the admissible region \( \mathcal{A}_{m_{\lambda}} \) described by the limit of (5.5):

\[ \int_0^1 (T_{\mu} - T_{m_{\lambda}})(x) \, dx = 0, \quad \int_0^1 (T_{\mu} - T_{m_{\lambda}})(x) \, dx \leq 0 \quad \forall y \in [0, 1]. \quad (5.6) \]

In fact, from the expression (1.19) of \( J(\mu_Y, m_{\lambda}) \), we have

\[ J(\mu_Y, m_{\lambda}) = 2I(\mu_Y, m_{\lambda}) - L \int xT_{\nu_Y} \, dx + O(\log L), \quad (5.7) \]

as \( L \rightarrow \infty \). Using the estimate (5.7) as input, by a similar argument as in Proposition 3.1, we have
Proposition 5.1. Under the assumptions and notations of Theorem 1.5, the function $H^K_\mu(\cdot)$ and rate function $I^K(\cdot)$ satisfy:

1. For $\mu$ satisfies (3.2), $H^K_\mu(\cdot)$ is upper semi-continuous in weak topology on $\{\nu \in \mathcal{M} : \nu(|x|) \leq R\}$ for any $R > 0$.
2. If $\int_0^1 (T_\mu(x) - T_{m_\lambda}(x))dx \neq 0$, or there exists some $0 < y < 1$ such that
\[ \int_y^1 (T_\mu(x) - T_{m_\lambda}(x))dx > 0, \]
then $I^K(\mu) = +\infty$.
3. If there exists some small constant $\epsilon > 0$
\[ \int_y^1 (T_\mu(x) - T_{m_\lambda}(x))dx \leq \begin{cases} -cy, & \text{for } 0 \leq y \leq \epsilon, \\ -c, & \text{for } \epsilon \leq y \leq 1 - \epsilon, \\ -c(1-y), & \text{for } 1 - \epsilon \leq y \leq 1. \end{cases} \tag{5.8} \]
then $I^K(\mu) = H^K_\mu(\nu^*) < \infty$ for some probability measure $\nu^*$ such that $\nu^*(|x|) < \infty$.
4. The rate function $I^K(\cdot)$ is lower semicontinuous on $\mathcal{M}^b([0, R])$ and achieves its minimum value only at the uniform measure $\text{unif}[\int xdm_\lambda - 1/2, \int xdm_\lambda + 1/2]$. 
5. For any measure $\mu$ in the admissible set $\mathcal{A}_{m_\lambda}$ as defined in (5.6), there exists a sequence of measures $\mu^\varepsilon$ inside the region as given in (5.8), converging to $\mu$ in weak topology and $\lim_{\varepsilon \to 0} I^D(\mu^\varepsilon) = I^D(\mu)$.

5.1 Large deviation lower bound

In this section we prove the large deviation lower bound in Theorem 1.5. It follows from combining the following Propositions 5.2 and 5.3, and noticing that the number of partitions with at most $N$ rows in $\mathcal{B}_3(\mu)$ is at most $\exp(O(N \log N))$.

**Proposition 5.2.** We assume the assumptions of Theorem 1.5. For any probability measure $\mu_\nu \in \mathcal{M}$, there exists a unique $\mu \in \mathcal{M}^b([0, R])$ such that
\[ \mu_\nu \in \arg \sup_{\nu \in \mathcal{M}} H^K_\mu(\nu), \quad H^K_\mu(\nu) = \int T_\nu(T_\mu - x)dx - J(\nu, m_\lambda). \tag{5.9} \]
and $T_\nu$ is uniquely determined by $T_\nu$ by
\[ T_\nu(x) = \tau(m_\lambda | y) \circ T_\nu(x) + x + \int \left( \frac{1}{T_\nu(x) - T_\nu(y) - T_\nu(z)} \right) dy. \]
Here, $\tau(m_\lambda | y)$ is the conditional expectation of $m_\lambda$ knowing $y$ under the non-commutative distribution $\tau$ uniquely associated to $(m_\lambda, \mu_\nu)$ as in Theorem 2.8.

**Proposition 5.3.** We assume the assumptions of Theorem 1.5. For any probability measure $\mu_\nu \in \mathcal{M}$, and $\mu$ be the unique measure in $\mathcal{M}^b([0, R])$ so that
\[ \mu_\nu \in \arg \sup_{\nu \in \mathcal{M}} H^K_\mu(\nu), \quad H^K_\mu(\nu) = \frac{1}{2} \int T_\nu(T_\mu - x)dx - J(\nu, m_\lambda). \]
Then we have
\[ \frac{1}{N^2} \log \sup_{\eta_N \in \mathcal{B}_3(\mu)} K_{\lambda_N \eta_N} \geq - \left( H^K_\mu(\mu_\nu) + o_\nu(1) + o_N(1) \right). \tag{5.10} \]
Proof of Theorem 1.5. Item 1 of Theorem 1.5 follows from Proposition 5.1. For Item 2, the large deviation upper bound follows from (5.4). If $\mu$ does not satisfy $\int_0^1 (T_{\mu}(x) - T_{\mu_\mu}(x))dx \neq 0$ or the limiting Schur-Horn inequalities (1.22), then both sides of (1.23) are $-\infty$. There is nothing to prove. In the following we first prove (1.23) when $\mu$ satisfies $\int_0^1 (T_{\mu}(x) - T_{\mu_\mu}(x))dx = 0$ and the strong limiting Schur-Horn inequalities (5.8) with some $c > 0$. In this case, thanks to Item 3 in Proposition 5.1, there exists a probability measure $\mu_Y$ such that $T^K(\mu) = H^\mu_Y(\mu_Y) < \infty$ and $\mu_Y \in M$. Then Propositions 5.2 and 5.3 imply that $\mu$ is uniquely determined by $\mu_Y$ and the large deviation lower bound holds. This gives the full large deviation principle when the strong limiting Schur-Horn inequalities (5.8) hold. Next we extend it to the boundary case by a continuity argument. Thanks to Item 5 in Proposition 5.1, for any measure $\mu$ inside the admissible set (5.6) but not satisfying (5.8), there exists a sequence of measures $\mu^\varepsilon$ inside the region as given in (5.8), converging to $\mu$ in weak topology and $\lim_{\varepsilon \to 0} T^K(\mu^\varepsilon) = T^K(\mu)$. Then for any $\delta > 0$, there exists sufficiently small $\varepsilon > 0$

$$\liminf_{N \to \infty} \frac{1}{N^2} \log \sup_{m([\eta_N] \in \mathcal{B}_k(\mu))} K_{\lambda_N \eta_N} \geq \liminf_{N \to \infty} \frac{1}{N^2} \log \sup_{m([\eta_N] \in \mathcal{B}_k(\mu^\varepsilon))} K_{\lambda_N \eta_N} = T^K(\mu^\varepsilon) + o_\delta(1). \quad (5.11)$$

The large deviation lower bound follows by first sending $\varepsilon$ and then $\delta$ to zero in (5.11). This finishes the proof of Theorem 1.5.

The proofs of both Propositions 5.2 and 5.3 rely on the following probability estimate.

**Proposition 5.4.** We assume the assumptions of Theorem 1.5. Let $Y_N = \text{diag}\{y_1, y_2, \ldots, y_N\}$ be a sequence of diagonal matrices, whose spectral measures converge in Wasserstein distance (1.6) towards $\mu_Y$. For any $\mu$ with support on $[0, \infty]$, if

$$\sup_{\nu \in \mathcal{M}} \left\{ \int (T_{\mu} - x)T_{\nu}dx - J(\nu, m_{\lambda}) \right\} > \int (T_{\mu} - x)T_{\mu_Y}dx - J(\mu_Y, m_{\lambda}). \quad (5.12)$$

Then there exists a small $\delta > 0$, and positive constant $c(\delta) > 0$ such that

$$\sum_{\eta_N : m([\eta_N] \in \mathcal{B}_k(\mu))} K_{\lambda_N \eta_N} m_{\eta_N}(e^{Y_N}) \leq e^{-c(\delta)N^2} S_{\lambda_N}(e^{Y_N}). \quad (5.13)$$

**Proof of Proposition 5.4.** Under Assumption (5.12), for sufficiently small $\varepsilon > 0$, there exists a measure $\nu \in \mathcal{M}$ such that

$$\int (T_{\mu} - x)T_{\nu}dx - J(\nu, m_{\lambda}) \geq \int (T_{\mu} - x)T_{\mu_Y}dx - J(\mu_Y, m_{\lambda}) + \varepsilon. \quad (5.14)$$

We divide by $S_{\lambda_N}(e^{Y_N})$ on both sides of (5.13), and use the estimates (5.3) and (1.19),

$$\frac{1}{S_{\lambda_N}(e^{Y_N})} \sum_{\eta_N : m([\eta_N] \in \mathcal{B}_k(\mu))} K_{\lambda_N \eta_N} m_{\eta_N}(e^{Y_N})$$

$$= \exp \left\{ -N^2 \left( J(\mu_Y, m_{\lambda}) - \int (T_{\mu} - x)T_{\mu_Y}dx + o_\delta(1) + o_N(1) \right) \right\} \sum_{\eta_N : m([\eta_N] \in \mathcal{B}_k(\mu))} K_{\lambda_N \eta_N}$$

$$\leq \exp \left\{ -N^2 \left( J(\mu_Y, m_{\lambda}) - \int (T_{\mu} - x)T_{\mu_Y}dx - J(\nu, m_{\lambda}) + \int (T_{\mu} - x)T_{\nu}dx + o_\delta(1) + o_N(1) \right) \right\}$$

$$\leq \exp \left\{ -N^2 (\varepsilon + o_\delta(1) + o_N(1)) \right\},$$

where in the first inequality we used the large deviation upper bound (5.4), and (5.14) in the last inequality. The claim follows provided we take $\delta$ sufficiently small and $N$ large. \qed
Proof of Proposition 5.2. We first prove the existence of such \( \mu \) by contradiction. If there is no such \( \mu \), i.e. for any measure \( \mu \) supported on \([0, \bar{\theta}]\), we have

\[
\mu_Y \notin \arg \sup_{\nu \in \mathcal{M}} \left\{ \int (T_{\mu} - x)T_{\nu}dx - J(\nu, m_\lambda) \right\}.
\]

Then it follows from Proposition 5.4 that there exists a small \( \delta > 0 \), and positive constant \( c(\delta) > 0 \) such that

\[
\sum_{\eta_N : m(\eta_N) \notin \mathcal{B}_\delta (\mu)} K_{\lambda_N} m_{\eta_N}(e^{Y_N}) \leq e^{-c(\delta)N^2} S_{\lambda_N}(e^{Y_N}).
\]

Since the space of probability measure supported on \([0, \bar{\theta}]\) is compact, we get a finite open cover \( \cup \mathcal{B}_\delta (\mu) \) of the set of probability measure supported on \([0, \bar{\theta}]\), we get a contradiction since for \( N \) large enough

\[
S_{\lambda_N}(e^{Y_N}) = \sum_{\eta_N : m(\eta_N) \notin \mathcal{B}_\delta (\mu)} K_{\lambda_N} m_{\eta_N}(e^{Y_N}) \leq \sum_{\eta_N} e^{-c(\delta)N^2} S_{\lambda_N}(e^{Y_N}) < S_{\lambda_N}(e^{Y_N}).
\]

In the following we prove the uniqueness of such measure \( \mu \) satisfying (5.9). We note that if \( \mu \in \mathcal{M}^p ([0, \bar{\theta}]) \), then \( T_\mu (x) - x \) is monotonic increasing. Since \( \mu_Y \) is one of the maximizer, then for any \( \varepsilon > 0 \),

\[
\int T_Y(T_{\mu} - x)dx - J(T_Y, T_{m_\lambda}) \geq \int (T_Y + \varepsilon \bar{T}_C)(\text{unif}[0, 1])(T_{\mu} - x)dx - J(T_Y + \varepsilon \bar{T}_C, T_{m_\lambda})
\]

\[
\geq \int (T_Y + \varepsilon \bar{T}_C)(T_{\mu} - x)dx - J(T_Y + \varepsilon \bar{T}_C, T_{m_\lambda}).
\]

By rearranging the above expression, and sending \( \varepsilon \) to 0, we have

\[
\partial_\varepsilon J(T_Y + \varepsilon \bar{T}_C, T_{m_\lambda}) \bigg|_{\varepsilon = 0} \geq \int \bar{T}_C(T_{\mu} - x)dx.
\]

(5.15)

We recall that \( J \) as in (1.19) is given by \( I \) and some explicit integrals. We compute the derivative of \( J \),

\[
\partial_\varepsilon J(T_Y + \varepsilon \bar{T}_C, T_{m_\lambda}) \bigg|_{\varepsilon = 0} = 2\partial_\varepsilon I(T_Y + \varepsilon \bar{T}_C, T_{m_\lambda}) \bigg|_{\varepsilon = 0} - \frac{1}{2} \int \bar{T}_C dx,
\]

\[
+ \frac{1}{2} \int \left( \frac{1}{T_Y(x) - T_Y(y)} - \frac{1}{1 - e^{T_Y(y) - T_Y(x)}} \right) (\bar{T}_C(x) - \bar{T}_C(y)) dxdy.
\]

We will choose \( \bar{T}_C \) in either the case (2.48) or (2.49). We notice that in both cases if we replace \( \bar{T}_C \) by \(-\bar{T}_C\), both sides of (5.15) change the sign. Therefore, we conclude that

\[
\partial_\varepsilon J(T_Y + \varepsilon \bar{T}_C, T_{m_\lambda}) \bigg|_{\varepsilon = 0} = \int \bar{T}_C(T_{\mu} - x)dx.
\]

Now if we choose \( \bar{T}_C \) in (2.49), i.e. \( \bar{T}_C \) supported on that \( \{ x : T_Y(x) = a \} \), we have

\[
2\partial_\varepsilon I(T_Y + \varepsilon \bar{T}_C, T_{m_\lambda}) \bigg|_{\varepsilon = 0} - \frac{1}{2} \int T_C dx = \int T_C(x)dx\tau(m_\lambda y)(a) - \frac{1}{2} \int T_C dx = \int \bar{T}_C(T_{\mu} - x)dx,
\]

We conclude that \( T_{\mu}(x) = \tau(m_\lambda y)(a) + x - 1/2 = \tau(m_\lambda y) \circ T_Y(x) + x - 1/2 \) on \( \{ x : T_Y(x) = a \} \).

Especially, on the intervals where \( T_Y \) is a constant, we have

\[
T_{\mu}(x) = \tau(m_\lambda y) \circ T_Y(x) + x - 1/2.
\]

(5.16)
Next we take $\tilde{T}_C = f(T_Y)$ as in (2.48),
\[
\partial_t J(T_Y + \varepsilon f(T_Y), T_{m\lambda})|_{\varepsilon = 0} = \int f(x) \tau(m_{\lambda}|y)(x) d\nu + \int \int \left( \frac{1}{T_Y(x) - T_Y(y)} - \frac{1}{1 - e^{T_Y(y) - T_Y(x)}} \right) dy f(T_Y(x)) dx
\]
\[
= \int f(T_Y) \tau(m_{\lambda}|y) \circ T_Y(x) dx + \int \int \left( \frac{1}{T_Y(x) - T_Y(y)} - \frac{1}{1 - e^{T_Y(y) - T_Y(x)}} \right) dy f(T_Y(x)) dx
\]  
(5.17)

On the intervals where $T_Y$ is increasing, (5.17) implies that
\[
T_{\mu}(x) = \tau(m_{\lambda}|y) \circ T_Y(x) + x + \int \left( \frac{1}{T_Y(x) - T_Y(y)} - \frac{1}{1 - e^{T_Y(y) - T_Y(x)}} \right) dy.
\]  
(5.18)

Therefore, we conclude from (5.16) and (5.18) that (5.18) holds almost surely on $[0, 1]$, which uniquely determines $\mu$. This finishes the proof of Proposition 5.2.

\[\square\]

**Proof of Proposition 5.3.** Thanks to the uniqueness of $\mu$, we have that for any $\mu' \neq \mu$ in $\mathcal{M}^b([0, \mathcal{R}])$
\[
\mu_Y \not\in \arg \sup_{\nu \in \mathcal{M}} \left\{ \int (T_{\mu'} - x) T_{\nu} dx - J(\nu, m_{\lambda}) \right\}.
\]

As a consequence the assumption in Proposition 5.4 holds,
\[
\sup_{\nu \in \mathcal{M}} \left\{ \int (T_{\mu'} - x) T_{\nu} dx - J(\nu, m_{\lambda}) \right\} > \int (T_{\mu} - x) T_{\nu} dx - J(\mu_Y, m_{\lambda}).
\]

and there exists a small $\delta > 0$, and positive constant $c(\delta) > 0$ such that
\[
\sum_{\eta_{N} : m_{\eta_{N}} \in \mathbb{B}_{\delta} (\mu')} K_{\lambda N} m_{\eta_{N}} (e_{Y_{N}}) \leq e^{-c(\delta) N^2} S_{\lambda N} (e_{Y_{N}}).
\]

The space of probability measures $\mathcal{M}^b([0, \mathcal{R}])$, removing the open $\mathbb{B}_{\delta} (\mu)$ is compact, we get a finite open cover $\cup \mathbb{B}_{\delta} (\mu)_{1}$,
\[
\sum_{\eta_{N} : m_{\eta_{N}} \in \mathbb{B}_{\delta} (\mu)} K_{\lambda N} m_{\eta_{N}} (e_{Y_{N}}) \geq S_{\lambda N} (e_{Y_{N}}) - \sum_{i} \sum_{\eta_{N} : m_{\eta_{N}} \in \mathbb{B}_{\delta} (\mu_{i})} K_{\lambda N} m_{\eta_{N}} (e_{Y_{N}})
\]
\[
\geq \left( 1 - \sum_{i} e^{-c(\delta_{i}) N^2} \right) S_{\lambda N} (e_{Y_{N}}) = \left( 1 - \sum_{i} e^{-c(\delta_{i}) N^2} \right) \exp \{ N^2 (J(\mu_Y, \mu_{\lambda}) + o_N (1)) \}.
\]  
(5.19)

The large deviation lower bound at $\mu$ follows from the estimate (5.19) and (5.3)
\[
\sum_{\eta_{N} : m_{\eta_{N}} \in \mathbb{B}_{\delta} (\mu)} K_{\lambda N} m_{\eta_{N}}
\]
\[
= \exp \left\{ -N^2 \left( \int (T_{\mu} - x) T_{\nu} dx + o_{\delta} (1) + o_N (1) \right) \right\} \sum_{\eta_{N} : m_{\eta_{N}} \in \mathbb{B}_{\delta} (\mu)} K_{\lambda N} m_{\eta_{N}} (e_{Y_{N}})
\]
\[
\geq \exp \left\{ -N^2 \left( \int (T_{\mu} - x) T_{\nu} dx + o_{\delta} (1) + o_N (1) \right) \right\} \exp \{ N^2 (J(\mu_Y, \mu_{\lambda}) + o_N (1)) \}
\]
\[
= \exp \left\{ -N^2 \left( \int (T_{\mu} - x) T_{\nu} dx - J(\mu_Y, m_{\lambda}) + o_{\delta} (1) + o_N (1) \right) \right\}.
\]  

\[\square\]
Large Deviation Estimates for Littlewood-Richardson Coefficients

In this section, we use the spherical integral to study the large deviation of the Littlewood-Richardson coefficients and prove Theorem 1.7. From the definition (1.4) of Littlewood-Richardson coefficients,

\[ c^{\kappa_N}_{\lambda_N \eta_N} \leq \frac{S_{\lambda_N}(e^{\nu_N}) S_{\eta_N}(e^{\nu_N})}{S_{\kappa_N}(e^{\nu_N})}. \]  

(6.1)

The large deviation upper bound follows from combining the upper bound (6.1) and the asymptotics of Schur symmetric polynomials (1.19),

\[ \frac{1}{N^2} \log \sup_{\kappa_N : m(\kappa_N) \in B_{\delta}(\mu)} c^{\kappa_N}_{\lambda_N \eta_N} \leq -H_{LR}(\mu) + \alpha_1(1) + o_1(1), \]

(6.2)

In the following proposition we study the rate function \( I_{LR}(\cdot) \) from Theorem 1.7. Clearly, it is a good rate function again by Proposition 2.6. Unfortunately, it does not capture the admissible set for the possible eigenvalues given by Horn’s problem. However it contains the information about the constraints given by the Ky Fan type inequalities, i.e. it equals \(+\infty\) outside the region described by the Ky Fan type inequalities:

\[ \int_0^1 (T_{\mu} - T_{\mu_\lambda} - T_{\mu_\eta}) \, dx = 0, \quad \int_y^1 (T_{\mu} - T_{\mu_\lambda} - T_{\mu_\eta}) \, dx \leq 0, \quad \forall y \in [0, 1]. \]

(6.3)

**Proposition 6.1.** Under the assumptions of Theorem 1.7, the function \( H^{LR}_{\mu}(\cdot) \) and rate function \( I^{LR}(\cdot) \) as defined in Theorem 1.7 satisfy:

1. For measure \( \mu \) satisfies (6.3), \( H^{LR}_{\mu}(\cdot) \) is upper semi-continuous in weak topology on \( \{ \nu \in M : \nu(|x|) \leq \Re \} \) for any \( \Re > 0 \).
2. If \( \int_0^1 (T_{\mu} - T_{\mu_\lambda} - T_{\mu_\eta}) \, dx \neq 0 \), or there exists some \( 0 < y < 1 \) such that

\[ \int_y^1 (T_{\mu} - T_{\mu_\lambda} - T_{\mu_\eta}) \, dx > 0, \]

(6.4)

then \( I^{LR}(\mu) = +\infty \).
3. If there exists some small constant \( \epsilon > 0 \)

\[ \int_y^1 (T_{\mu} - T_{\mu_\lambda} - T_{\mu_\eta}) \, dx \leq \begin{cases} -\epsilon y, & \text{for } 0 \leq y \leq \epsilon, \\ -\epsilon, & \text{for } \epsilon \leq y \leq 1 - \epsilon, \\ -\epsilon(1 - y), & \text{for } 1 - \epsilon \leq y \leq 1, \end{cases} \]

(6.5)

then \( I^{LR}(\mu) = H^{LR}_{\mu}(\nu^*) < \infty \) for some probability measure \( \nu^* \) such that \( \nu^*(|x|) < \infty \).
4. The rate function \( I^{LR}(\cdot) \) is nonnegative and lower semicontinuous on \( M^b([0, \Re]) \).

**6.1 Large deviation lower bound**

In this section we prove the large deviation lower bound in Theorem 1.7. It follows from combining the following Propositions 6.2 and 6.3, and noticing the number of partitions with at most \( N \) rows in \( B_\delta(\mu) \) is at most \( \exp\{O(N \log N)\} \).
**Proposition 6.2.** We assume the assumptions of Theorem 1.7. Let $\mu_Y$ be compactly supported and with an absolutely continuous part in each of its connected components. Then there exists a unique $\mu \in \mathcal{M}^b([0,2\mathbb{R}])$ such that
\[
\mu_Y \in \arg \sup_{\nu \in \mathcal{M}} H_{\mu_Y}^{LR}(\nu), \quad H_{\mu_Y}^{LR}(\nu) = J(\nu, \mu) - J(\nu, m_\lambda) - J(\nu, m_\eta).
\]

**Proposition 6.3.** We assume the assumptions of Theorem 1.7. If $\mu \in \mathcal{M}^b([0,2\mathbb{R}])$ is the unique measure with
\[
\mu_Y \in \arg \sup_{\nu \in \mathcal{M}} H_{\mu_Y}^{LR}(\nu), \quad H_{\mu_Y}^{LR}(\nu) = J(\nu, \mu) - J(\nu, m_\lambda) - J(\nu, m_\eta).
\]
Then we have
\[
\lim \inf \lim \inf_{\delta \to 0} \lim_{N \to \infty} \frac{1}{N^2} \log \sum_{\kappa_N : m[\kappa_N] \in B(\mu)} c^{\kappa_N}_{\lambda_N, \eta_N} \geq -H_{\mu_Y}^{LR}(\mu_Y) = -\mathcal{Z}^{LR}(\mu). \quad (6.6)
\]

The proofs of both Propositions 6.2 and 6.3 relies on the following probability estimate.

**Proposition 6.4.** We assume the assumptions of Theorem 1.7. Let $Y_N = \text{diag}\{y_1, y_2, \ldots, y_N\}$ be a sequence of diagonal matrices, whose spectral measures converge in Wasserstein distance (1.6) towards $\mu_Y$. For any $\mu \in \mathcal{M}^b([0,2\mathbb{R}])$, such that
\[
\sup_{\nu \in \mathcal{M}} \{J(\nu, \mu) - J(\nu, m_\lambda) - J(\nu, m_\eta)\} > J(\mu_Y, \mu) - J(\mu_Y, m_\lambda) - J(\mu_Y, m_\eta). \quad (6.7)
\]
Then there exists a small $\delta > 0$, and positive constant $c(\delta) > 0$ such that
\[
\sum_{\kappa_N : m[\kappa_N] \in B(\mu)} c^{\kappa_N}_{\lambda_N, \eta_N} S_{\kappa_N}(e^{Y_N}) \leq e^{-c(\delta)N^2} S_{\lambda_N}(e^{Y_N}) S_{\eta_N}(e^{Y_N}). \quad (6.8)
\]

**Proof of Proposition 6.4.** Under Assumption (6.7), for sufficiently small $\varepsilon > 0$, there exists a measure $\nu \in \mathcal{M}$ such that
\[
J(\nu, \mu) - J(\nu, m_\lambda) - J(\nu, m_\eta) > J(\mu_Y, \mu) - J(\mu_Y, m_\lambda) - J(\mu_Y, m_\eta) + \varepsilon. \quad (6.9)
\]
We divide $S_{\lambda_N}(e^{Y_N}) S_{\eta_N}(e^{Y_N})$ on both sides of (6.8), and use the asymptotics of Schur symmetric polynomials (1.19),
\[
\frac{1}{(S_{\lambda_N}(e^{Y_N}) S_{\eta_N}(e^{Y_N}))} \sum_{\kappa_N : m[\kappa_N] \in B(\mu)} c^{\kappa_N}_{\lambda_N, \eta_N} S_{\kappa_N}(e^{Y_N})
\leq e^{-N^2(J(\mu_Y, \mu) - J(\mu_Y, m_\lambda) - J(\mu_Y, m_\eta) + o_2(1) + o_N(1))} \sum_{\kappa_N : m[\kappa_N] \in B(\mu)} c^{\kappa_N}_{\lambda_N, \eta_N}
\leq e^{-N^2(J(\nu, \mu) - J(\nu, m_\lambda) - J(\nu, m_\eta) + o_2(1) + o_N(1))} \leq e^{-N^2(\varepsilon + o(1) + o_N(1))},
\]
where in the first equality we used the large deviation upper bound (6.2), and (6.9) in the last inequality. The claim follows provided we take $\delta$ sufficiently small.

**Proof of Proposition 6.2.** We prove the existence of such $\mu$ by contradiction. If there is no such $\mu$, i.e. for any measure $\mu \in \mathcal{M}^b([0,2\mathbb{R}])$, we have
\[
\mu_Y \not\in \arg \sup_{\nu \in \mathcal{M}} \{J(\nu, \mu) - J(\nu, m_\lambda) - J(\nu, m_\eta)\}.
\]
Then it follows from proposition 6.4, there exists a small $\delta > 0$, and positive constant $c(\delta) > 0$ such that

$$\sum_{\kappa_N: m|\kappa_N| \in B_{\delta}(\mu)} c_{\kappa_N \eta_N}^N S_{\kappa_N}(e^{YN}) \leq e^{-c(\delta)N^2} S_{\lambda_N}(e^{YN}) S_{\eta_N}(e^{YN}).$$

Since the space of probability measure supported on $[0, 2\delta]$ with density bounded by 1 is compact, we get a finite open cover $\bigcup_{\delta_i} (\mu_i)$ of $\mathcal{M}^+(\{[0, 2\delta]\})$, we get the contradiction

$$S_{\lambda_N}(e^{YN}) S_{\eta_N}(e^{YN}) = \sum_i \sum_{\kappa_N: m|\kappa_N| \in B_{\delta_i}(\mu_i)} c_{\kappa_N \eta_N}^N S_{\kappa_N}(e^{YN}) \leq \sum_i e^{-c(\delta_i)N^2} S_{\lambda_N}(e^{YN}) S_{\eta_N}(e^{YN}) < S_{\lambda_N}(e^{YN}) S_{\eta_N}(e^{YN}).$$

The proof of the uniqueness of Proposition 6.2 is where we need the regularity of the measure $\mu_Y$. If $\mu_Y$ is a maximizer of $H^L_{\mu}(\cdot)$, then

$$0 = \partial_{\epsilon} H^L_{\mu}(T_Y + \epsilon \tilde{T}_C) \bigg|_{\epsilon = 0} = \partial_{\epsilon} J(T_Y + \epsilon \tilde{T}_C, T_{\mu}) \bigg|_{\epsilon = 0} - \partial_{\epsilon} J(T_Y + \epsilon \tilde{T}_C, T_{m_\lambda}) \bigg|_{\epsilon = 0} - \partial_{\epsilon} J(T_Y + \epsilon \tilde{T}_C, T_{m_\eta}) \bigg|_{\epsilon = 0},$$

(6.10)

We take $\tilde{T}_C = f(T_Y)$ and recall from (5.17), we have

$$\partial_{\epsilon} J(T_Y + \epsilon f(T_Y), T_{\mu}) \bigg|_{\epsilon = 0} = \int f(x) \tau(\mu_1(y)(x))d\mu_Y + \int \int \left( \frac{1}{x - y} - \frac{1}{1 - e^{y - x}} \right) d\mu_Y(y) f(x)d\mu_Y(x).$$

(6.11)

Using (6.11), we can simplify (6.10) as

$$0 = \partial_{\epsilon} J(T_Y + \epsilon \tilde{T}_C, T_{\mu}) \bigg|_{\epsilon = 0} - \partial_{\epsilon} J(T_Y + \epsilon \tilde{T}_C, T_{m_\lambda}) \bigg|_{\epsilon = 0} - \partial_{\epsilon} J(T_Y + \epsilon \tilde{T}_C, T_{m_\eta}) \bigg|_{\epsilon = 0}$$

$$= \int f(x)(\tau(m_\lambda y)(x) - \tau(m\lambda y)(x) - \tau(m_\eta y)(x))d\mu_Y(x)$$

$$- \int \int \left( \frac{1}{x - y} - \frac{1}{1 - e^{y - x}} \right) d\mu_Y(y) f(x)d\mu_Y(x).$$

Since we can take $f$ any bounded Lipschitz function, we conclude from (4.20) that $\mu_Y$-almost surely for any $x \in \mathbb{R}$,

$$\tau(m_\lambda y)(x) = \tau(m_\lambda y)(x) + \tau(m_\eta y)(x) + \int \left( \frac{1}{x - y} + \frac{1}{1 - e^{y - x}} \right) d\mu_Y(y).$$

(6.12)

We denote the solutions of the variational problem $I(\mu_Y, \mu), I(\mu_Y, m_\lambda), I(\mu_Y, m_\eta)$ as given in (2.7) by $f^{\mu_Y \to \mu}(t, x), f^{\mu_Y \to m_\lambda}(t, x), f^{\mu_Y \to m_\eta}(t, x)$ respectively. Then Item 7 in Proposition 2.5 gives that

$$\tau(\mu_1(y)(x)) = \text{Re}[f^{\mu_Y \to \mu}(0, x)] + H\mu_Y - x,$$

$$\tau(m_\lambda y)(x) = \text{Re}[f^{\mu_Y \to m_\lambda}(0, x)] + H\mu_Y - x,$$

$$\tau(m_\eta y)(x) = \text{Re}[f^{\mu_Y \to m_\eta}(0, x)] + H\mu_Y - x.$$

Therefore (6.12) implies that the real part of $f^{\mu_Y \to \mu}(0, x)$ is given by

$$\text{Re}[f^{\mu_Y \to \mu}(0, x)] = \text{Re}[f^{\mu_Y \to m_\lambda}(0, x)] + \text{Re}[f^{\mu_Y \to m_\eta}(0, x)]$$

$$+ H\mu_Y - x + \int \left( \frac{1}{x - y} + \frac{1}{1 - e^{y - x}} \right) d\mu_Y(y),$$

50
for \( \mu_Y \)-almost surely all \( x \in \mathbb{R} \). The imaginary part of \( f^{\nu_{x_{1}}}^{\mu} \) is given by the measure \( \nu \). By our assumption that \( \mu_Y \) is compactly supported and has absolutely continuous part in each of its connected components, we conclude from Corollary 2.16 that \( \mu \) is uniquely determined by \( \mu_Y \) and \( f^{\mu_Y_{x_{1}}}^{\mu}(0, x) \). This finishes the proof of Proposition 6.2.

**Proof of Proposition 6.3.** Thanks to the uniqueness of \( \mu \), we have that for any \( \mu' \neq \mu \) in \( \mathcal{M}((0, 2K]) \)

\[
\mu_Y \not\in \text{arg sup} \left\{ J(\nu; \mu') - J(\nu; \mu) - J(\nu; \mu_\eta) \right\}.
\]

As a consequence the assumption in Proposition 6.4 holds,

\[
\text{sup} \left\{ J(\nu; \mu') - J(\nu; \mu) - J(\nu; \mu_\eta) \right\} > J(\mu_Y; \mu') - J(\nu; \mu) - J(\nu; \mu_\eta),
\]

and there exists a small \( \delta > 0 \), and positive constant \( c(\delta) > 0 \) such that

\[
\sum_{\kappa_N: m[\kappa_N] \in \mathbb{B}_\delta(\mu')} c^{\kappa_N}_{\lambda_N Y} S_{\kappa_N}(e^{Y_N}) \leq e^{-c(\delta)N^2} S_{\lambda_N}(e^{Y_N}) S_{\eta_N}(e^{Y_N}). \tag{6.13}
\]

Again the space of probability measure supported on \([0, 2\delta]\) with density bounded by 1 removing the open ball \( \mathbb{B}(\mu) \) is compact, we get a finite open cover \( \bigcup \mathbb{B}_\delta(\mu_i) \),

\[
\sum_{\kappa_N: m[\kappa_N] \in \mathbb{B}_\delta(\mu)} c^{\kappa_N}_{\lambda_N Y} S_{\kappa_N}(e^{Y_N}) \geq \sum_{\kappa_N: m[\kappa_N] \in \mathbb{B}_\delta(\mu)} e^{-c(\delta)N^2} S_{\lambda_N}(e^{Y_N}) S_{\eta_N}(e^{Y_N}) \tag{6.14}
\]

where we used (6.13) in the last line. The large deviation lower bound at \( \mu \) follows from the estimates (6.14) and (1.19),

\[
\sum_{\kappa_N: m[\kappa_N] \in \mathbb{B}_\delta(\mu)} c^{\kappa_N}_{\lambda_N Y} S_{\kappa_N}(e^{Y_N}) \geq e^{-N^2(J(\mu_Y; \mu) + o_N(1))} e^{N^2(J(\mu_Y; \mu) + J(\mu_Y; \mu_\eta) + o_N(1))}
\]

\[
\geq e^{-N^2(J(\mu_Y; \mu) - J(\mu_Y; \mu_\eta) + o_N(1))}.
\]

\[\square\]

### 7 Properties of \( f(t, x) \)

We recall that given two compactly supported probability measures \( \mu, \nu, (\rho^*, u^*) \) is the unique solution of the variational problem (2.6), and \( f(t, x) = u_t^*(x) + \pi \rho_t^*(x) \) satisfies the Burgers equation (2.14). We recall the set \( \Omega \) from (2.10),

\[
\Omega = \{ (t, x) \in (0, 1) \times \mathbb{R} : \rho_t^*(x) > 0 \}. \tag{7.1}
\]

In this section we study the continuity properties of \( f(t, x) \) at boundaries, and prove Theorem 2.15 and Corollary 2.16 showing that \( f(0, .) \) uniquely determines \( f(1, .) \). We start by studying the set \( \Omega \) and showing its connected components are simply connected. For the rest of this section, we assume that neither \( \mu \) nor \( \nu \) are Dirac masses. If supplementary hypotheses are required in order for certain results to hold, we will state them at the required moment.
We introduce the following notation: given an open set $\Omega$, a function $\omega: \Omega \to \mathbb{C}$, and a subset $D \subseteq \Omega$, we denote by $\mathcal{C}_D(\omega, D)$ the set of all points $w \in \mathbb{C} \cup \{\infty\}$ for which there exists a sequence $\{w_k\}_{k \in \mathbb{N}} \subseteq \Omega$ such that $\lim_{k \to \infty} \sup_{w \neq w_k} \text{dist}(w, D) = 0$ and $\lim_{k \to \infty} \omega(w_k) = w$. When there is no risk of confusion, we suppress the domain $\Omega$ from the notation and denote $\mathcal{C}_D(\omega, D)$ simply by $\mathcal{C}(\omega, D)$. Also, if $D = \{x\}$ is a singleton, we write $\mathcal{C}(\omega, x)$ rather than $\mathcal{C}(\omega, \{x\})$. From Proposition 2.5, we deduce

**Proposition 7.1.** We assume that the probability measures $\mu, \nu$ are compactly supported. The pair $(\rho^*, u^*)$ is the unique solution of the variational problem (2.6), and $f(t, x) = u^*_t(x) + \pi i \rho^*_t(x)$ satisfies the Burgers equation (2.14). For any $t \in (0, 1)$, the function $x \mapsto \text{Im}[f(t, x)]$ extends continuously to all points $(t, x) \in \partial \Omega$. Moreover, for any $m \in (0, 1)$ and any connected subset $D \subseteq \partial \Omega \cap (0, m) \times \mathbb{R}$, the function $g(t, x) = x - t f(t, x)$ satisfies the condition $\mathcal{C}(g, D) \subseteq \mathbb{R}$, and $\mathcal{C}(g, D)$ is bounded.

Our main tool is a change of variable, transforming (2.14) into a Beltrami equation. We recall from (2.60) that $f(z)$ satisfies the following Beltrami equation

$$\partial_z f = \frac{1 - f}{1 + f} \partial_z f, \quad z = x - it, \quad (-\text{Im}[z], \text{Re}[z]) \in \Omega. \quad (7.2)$$

We also define the functions $g(t, x)$ and $h(t, x)$

$$g(t, x) = x - tf(t, x), \quad h(t, x) = x + (1 - t)f(t, x), \quad (t, x) \in \Omega. \quad (7.3)$$

Then $(\partial_x - i\partial_t)g = 1 - t\partial_x f + if + it\partial_t f = 1 + if - t\partial_x f - itf\partial_x f = (1 + if)(1 - t\partial_x f)$ and $(\partial_x + i\partial_t)g = 1 - t\partial_x f - if + itf\partial_x f = (1 - if)(1 - t\partial_x f)$. Thus,

$$\partial_z g = (\partial_x - i\partial_t)g = \frac{(1 + if)(1 - t\partial_x f)}{(1 - if)(1 - t\partial_x f)}(\partial_x + i\partial_t)g = \frac{1 - f}{1 + f} \partial_z f.$$ 

That is, $f$ and $g$ satisfy the same Beltrami equation (7.2). Since $h = g + f$, it follows easily that $h$ also satisfies the Beltrami equation (7.2).

Let us enumerate some of the properties that are imposed on $f, g, h$ by the fact that they satisfy (7.2). We refer to [?] for details.

1. $f, g, h$ are either open, in the sense that the image of any open subset of $\Omega$ via one of these functions is an open set, or constant;
2. None of $f, g, h$ have a local maximum or minimum in $\Omega$. The lack of a minimum imposes also the inclusions $f(\Omega), h(\Omega), -g(\Omega) \subseteq \mathbb{C}_+;$
3. $\partial f(D) \subseteq f(\partial D)$ for any open set $D \neq \emptyset \subseteq \Omega$, and the same for $g, h$. If $f$ does not extend continuously to $\partial D$, then $f(\partial D)$ should be understood as the set of limit points of $f$ along sequences in $D$ converging to points in $\partial D;$
4. The branching points of $f$ are the critical points $\{(t, x) \in (0, 1) \times \mathbb{R}: \partial_x f(t, x) = 0 = \partial_z f(t, x)\}$; at all other points, $f$ is a local homeomorphism. The same holds for $g$ and $h$.

**Lemma 7.2.** Any connected component of $\Omega$ as defined in (7.1) is simply connected.

**Proof.** Recall that, by definition, $\Omega = \{(t, x) \in (0, 1) \times \mathbb{R}: \text{Im}[f_t(x)] > 0\}$. Assume towards contradiction that a smooth simple closed curve $\gamma$ in $\Omega$ is not homotopy equivalent to a point in $\Omega$. Let $C$ be the bounded component of $\mathbb{C} \setminus \gamma$ within $\Omega$. Then $f(\gamma) \subseteq \mathbb{C}_+$ is a compact set (hence bounded and bounded away from $\mathbb{R}$) and $f(\partial C \setminus \gamma) \subseteq \mathbb{R}$ is a bounded set according to Item 4 in Proposition 2.5. Thus, $\partial f(C) \subseteq f(\gamma) \cup f(\partial C \setminus \gamma)$ is a bounded set in $\mathbb{R} \cup (\mathbb{C}_+ + i\varepsilon)$ for some $\varepsilon > 0$. This implies that $f(C)$ cannot be bounded since the real part of its boundary is not connected to the rest of its boundary, contradicting Item 4 in Proposition 2.5.

As an easy consequence of the conservation of mass, we have
Lemma 7.3. If $\Omega_0$ is a connected component of $\Omega$, then as a function on $[0, 1]$, $\int_{\Omega_0 \cap \{(t, x)\} \times \mathbb{R}} \text{Im}[f(t, x)] \, dx \in (0, 1]$ is constant.

Thanks to the weak continuity of $t \mapsto \text{Im} f(t, \cdot)$, an interesting phenomenon is illustrated by the above lemmas, and particularly the above conservation of mass: for $\Psi$ which is not connected (and hence simply connected), it is necessary (but not sufficient) that both $\text{supp} \mu$ and $\text{supp} \nu$ have more than one connected component, and there exist strict subsets $C^\mu_1, \ldots, C^\mu_n$ of the set of connected components of $\text{supp} \mu$ and $\{C^\nu_1, \ldots, C^\nu_m\}$ of that of $\text{supp} \nu$ such that $\mu(C^\mu_1 \cup \cdots \cup C^\mu_n) = \nu(C^\nu_1 \cup \cdots \cup C^\nu_m)$, for some $1 \leq \ell, m \in \mathbb{N} \cup \{\infty\}$.

The great advantage of $g$ over $f$ is that one can guarantee that $g$ is a homeomorphism close to $(0, 0) \times \mathbb{R}$. We will make this statement precise in Proposition 7.5. According to Item 4 in Proposition 2.5, we have $|g(t, x) - x| = |tf(t, x)| < \sqrt{\tau / \sqrt{1 - t}}$, so that $\lim_{n \to \infty} g(t, x_n) = x$ whenever $\Omega \ni (t, x_n) \rightarrow (0, 0)$ as $n \to \infty$. Thus,

Lemma 7.4. If $\Omega_0$ is a connected component of $\Omega$, then either $\partial \Omega_0 \cap \{(0, 0) \times \mathbb{R}\}$ is one point, and then $\nu$ is an isolated atom at that point, or there exist $-\infty < a < b = \min \partial \Omega_0 \cap \{(0, 0) \times \mathbb{R}\} < \max \partial \Omega_0 \cap \{(0, 0) \times \mathbb{R}\} = b < +\infty$ such that $\mathbb{R} \ni g(\partial \Omega_0 \cap \{(0, 0) \times \mathbb{R}\}) \supseteq [a, b]$.

Proof. The first statement of the lemma is obvious. Thus, assume that $a < b$. By the definition of the topological boundary, there exists a sequence $\{(t_n, x_n)\}_{n \in \mathbb{N}} \subset \Omega_0$ such that $t_n \to 0$ and $x_n \to a$ as $n \to \infty$. As seen just above the statement of the lemma, it follows that $|g(t_n, x_n) - a| < |x_n - a| + \sqrt{\tau / \sqrt{1 - t_n}}$, and so $\lim_{n \to \infty} g(t_n, x_n) = a$. A similar statement holds for $b$ and a sequence $(t'_n, x'_n)$. For any $n \in \mathbb{N}$, we may draw a path $q_n$ starting at $(t_n, x_n)$ and continuing left along $(t_n, \{x\}) \times \mathbb{R}$ until it hits $\partial \Omega_0$, and a path $q'_n$ starting at $(t'_n, x'_n)$ and continuing right along $(t'_n, \{x\}) \times \mathbb{R}$ until it hits $\partial \Omega_0$. We may consider a simple path $p_n$ starting at $(t_n, x_n)$ and ending at $(t'_n, x'_n)$, and completely included in $\Omega_0 \setminus (q_n \cup q'_n)$. Clearly $g(q_n \cup p_n \cup q'_n)$ (the image via $g$ of the concatenation of the three paths) is included in $\mathbb{C}^-$, except for the images of the beginning and of the end of $q_n \cup p_n \cup q'_n$, which are mapped inside $\mathbb{R}$. The left endpoint (beginning) of this path is, by construction, at some point $(\xi_n, t_n)$ for a $\xi_n < x_n$; a similar statement - with the obvious modifications - holds for the right endpoint (end) of $q_n \cup p_n \cup q'_n$. It follows from the above that $g$ maps the beginning of this path into a real number which is no larger than $a + |x_n - a| + \sqrt{\tau / \sqrt{1 - t_n}}$ and the end point into a real number which is no smaller than $b - |x'_n - b| - \sqrt{\tau / \sqrt{1 - t_n}}$. Recalling that $f(\partial \Omega \cap \{(0, 1) \times \mathbb{R}\}) \subseteq \mathbb{R}$ and that $g(0, x) = x$ for any $(0, x) \in \partial \Omega$, we obtain that the segment $[a + |x_n - a| + \sqrt{\tau / \sqrt{1 - t_n}}, b - |x'_n - b| - \sqrt{\tau / \sqrt{1 - t_n}}]$ is included in $g(\partial \Omega_0 \cap \{(0, 1) \times \mathbb{R}\})$ for all $n \in \mathbb{N}$. Thus, by letting $n \to \infty$, we conclude that $g(\partial \Omega_0 \cap \{(0, 1) \times \mathbb{R}\}) \supseteq [a, b]$. This finishes the proof of Lemma 7.4.

Let us notice that we will assume that $\nu$ has an absolutely continuous part in each connected component, and therefore by the above lemma $\partial \Omega_0 \cap \{(0, 0) \times \mathbb{R}\}$ cannot be reduced to a point. The next proposition addresses the other case of Lemma 7.4.

Proposition 7.5. Consider a connected component $\Omega_0$ of $\Omega$ and points $a < b$ as in Lemma 7.4. Then there exist a domain $K \subset \mathbb{C}^-$ such that

- $[a, b] \subset \partial K$;
- for any $w \in (a, b)$ and $0 < \varepsilon < \text{dist}(w, \{a, b\})/2$ there exists $v > 0$ such that $\{x - iy; w - \varepsilon < x < w + \varepsilon, 0 < y < v\} \subset K$. Moreover, as $w$ tends to a (respectively $b$), $v$ is at least $O(|w - a|)$ (respectively $O(|b - w|)$);
- $K \subset g(\Omega_0)$;

and an analytic function $\Phi: K \rightarrow \mathbb{C}^+$ such that $\Phi \circ g = f$. The map $g$ maps a simply connected open set $O \subset \Omega_0$ satisfying the conditions that $\Omega_0 \cap \bar{O}$ contains the connected component of $\partial \Omega_0 \setminus \{(0, a), (0, b)\}$ containing points from the set $\{(t, x) \in \partial \Omega_0; a \leq x \leq b, t = \min \{s \in (0, 1); (s, x) \in \Omega_0\}\}$ and $\bar{O} \cap \{(1) \times \mathbb{R}\} = \emptyset$, bijectively onto $K$. 53
This proposition is key to the proof of Theorem 2.15 and Corollary 2.16 as we will show that under our hypotheses, \( \Phi(x) = f(x, 0) \) and \( \Phi(x - f(x, 1)) = f(x, 1) \), providing a one to one relation between \( f(x, 0) \) and \( f(x, 1) \).

**Proof.** While we believe there should be a direct argument guaranteeing the injectivity of \( g \) close to the “lower” boundary, we are not aware of it, and will show this indirectly.

We introduce the following notations, besides the ones introduced in Lemma 7.4:

- \(-\infty < a' = \min \partial \Omega_0 \cap (\{1\} \times \mathbb{R}) \leq \max \partial \Omega_0 \cap (\{1\} \times \mathbb{R}) = b' < +\infty\).
- \( \partial_1 \Omega_0 \) is the part of \( \partial \Omega_0 \) between \((0, a)\) and \((0, b)\), and away from \(\{1\} \times \mathbb{R}\): \( \partial_1 \Omega_0 \) is the closure of the connected component of \( \partial \Omega_0 \setminus \{(0, a), (0, b)\} \) containing points from
  \(\{(t, x) \in \partial \Omega_0 : a \leq x \leq b, t = \min\{s \in [0, 1] : (s, x) \in \Omega_0\}\}\).
- \( \partial_1 \Omega_0 \) is the analogue of \( \partial_1 \Omega_0 \) for \( a', b' \): the closure of the connected component of \( \partial_\Omega_0 \setminus \{(1, a'), (1, b')\} \) containing points from
  \(\{(t, x) \in \partial \Omega_0 : a' \leq x \leq b', t = \max\{s \in [0, 1] : (s, x) \in \Omega_0\}\}\).
- Finally, \( \partial_2 \Omega_0, \partial_3 \Omega_0 \) are the closures of the two connected components of \( \partial \Omega_0 \setminus (\partial_1 \Omega_0 \cup \partial_2 \Omega_0) \): \( \partial_2 \Omega_0 \) contains points from
  \(\{(t, x) \in \partial \Omega_0 : t \in [0, 1], x = \max\{r \in \mathbb{R} : (t, r) \in \Omega_0\}\}\), and \( \partial_3 \Omega_0 \) defined the same way, but with max replacing min.

As a consequence of Lemma 7.4 (or, rather, its proof), \( g \) maps \( \partial_1 \Omega_0 \) onto \([a, b]\), and \( h \) maps \( \partial_3 \Omega_0 \) onto \([a', b']\) (we did not exclude here the possibility that one, or both, of these intervals reduces to a point). All of \( f, g, h \) map \( \partial_2 \Omega_0 \cup \partial_3 \Omega_0 \) to \( \mathbb{R} \). We do not exclude here the possibility that either, or both, of \( \partial_2 \Omega_0, \partial_3 \Omega_0 \) is/are reduced to a point.

Note that since \( \mu(z) = (i - f(z))/(i + f(z)) \) is smooth (in fact real analytic) on \( \Omega_0 \), Equation (7.2) has a solution, namely \( f \) (in fact it has several solutions, \( g, h \) being two others), and \( |\mu(z)| \) can take the value one only on the boundary \( \partial \Omega_0 \), we are guaranteed the existence of a differentiable (as a function of two variables) homeomorphism \( W : \Omega_0 \to W(\Omega_0) \subseteq \mathbb{C} \) that is a solution of (7.2). Then there exist analytic functions \( F, H : W(\Omega_0) \to \mathbb{C}^+ \), \( G : W(\Omega_0) \to \mathbb{C}^- \) such that \( f = F \circ W, g = G \circ W, h = H \circ W \) (the so-called Stoilov factorization - see [2]). Since neither of \( f, g, h \) is constant, this has as an immediate consequence that \( W(\Omega_0) \subseteq \mathbb{C} \). In fact, more can be said: since \( W \) is a homeomorphism and \( \Omega_0 \) is simply connected (Lemma 7.2), so must be \( W(\Omega_0) \). In particular, by composing \( W \) to the left with a conformal mapping if necessary, we may assume without loss of generality that

\[ W(\Omega_0) = \mathbb{C}^+. \]

---

1 Stoilov’s result states that if \( F \) is a continuous “light open” mapping on a complex domain, then it factorizes \( F = f \circ W \), with \( W \) homeomorphic and \( f \) analytic: “A function which is continuous, light and open is called an interior mapping. It is a famous theorem of Stoilov that an interior mapping \( f \) of a plane domain is of the form \( f = \varphi \circ h \), where \( h \) is homeomorphic and \( \varphi \) analytic.” (Quote from Olli Lehto, [?].) “Light” means that the preimage of each point is discrete in the domain of the function, and “open” means that it sends open sets to open sets. It is known that a solution of a Beltrami equation with \( ||\mu||_\infty \leq k < 1 \) is necessarily light and open - see for instance [?, p.76]. The function \( f \) is the solution of (7.2) on the set \( \{(t, x) : |\mu(t, x)| < k\} \) too, so that on each of these sets it is an interior mapping. Of course, \( \Omega = \cup_{0 < k < 1} \{(t, x) : |\mu(t, x)| < k\} \), increasing union, so that \( f \) being discrete on each set means \( f \) is discrete on \( \Omega \), and \( f \) being open on each set means \( f \) is open on \( \Omega \). Now \( f = F \circ W \) with \( F \) analytic means that \( locally W = F^{1-1} \circ f \), so \( W \) is a solution of the same (7.2) except possibly on a discrete set. But solutions to (7.2) are in any case considered a.e. - in the sense that \( W \) is only required to be differentiable a.e.
This makes $F, -G, H$ non-constant self-maps of $\mathbb{C}^+$. As seen before, the homeomorphism $W : \partial \Omega \to \mathbb{R} \cup \{\infty\}$ sends boundary to boundary, i.e. $\partial \Omega$ into $\mathbb{R} \cup \{\infty\}$. As seen in Lemma 7.4, $g$ sends $\partial \Omega$ onto the segment $[a, b]$. The homeomorphism $W$ sends this same $\partial \Omega$ onto a connected subset of $\mathbb{R} \cup \{\infty\}$, that is, either an interval (possibly unbounded), or the complement of a bounded, open interval. As $W$ is a homeomorphism and the piece $\partial \Omega$ of $\partial \Omega$ under consideration is a strict subset of $\partial \Omega$ (with a complement containing at least an open arc in $(0, 1] \times \mathbb{R}$), $W$ cannot map it onto all of $\mathbb{R} \cup \{\infty\}$ (indeed, if that were the case, the Stoilow factorization for $g$ would provide an analytic function $G : \mathbb{C}^+ \to \mathbb{C}^-$ that sends all of $\mathbb{R} \cup \{\infty\}$ onto $[a, b]$, which is absurd). Since it is by necessity closed, this image has a complement that is either an open, nonempty interval, or the union of two unbounded open intervals.

By pre-composing $W$ with a map of the type $z \mapsto \frac{1}{\pi z}$ for a $d \in \mathbb{R}$ outside this range, we may assume without loss of generality that $W$ maps $\partial \Omega$ onto a compact interval and $W^{-1}(\infty) \in \partial \Omega$. Denote this interval by $[\alpha, \beta]$. The same argument, with $g$ replaced by $h$, allows us to conclude that $W$ maps $\partial \Omega$ onto a closed connected strict subset of $\mathbb{R} \cup \{\infty\}$.

As $g$ maps $\partial \Omega$ onto $[a, b]$ and $W$ maps $\partial \Omega$ onto $[\alpha, \beta]$ (note that $W$ must send $a$ to $\beta$ and $b$ to $\alpha$), the Stoilow factorization $g = G \circ W$ guarantees that $G([\alpha, \beta]) = [a, b]$. A priori this must be understood in the sense of limits at the boundary. However, we claim that $G$ must reflect analytically through the interval $[\alpha, \beta]$, which it maps bijectively onto $[a, b]$. Indeed, since $G$ maps $\mathbb{C}^+$ into $\mathbb{C}^-$, it has a Nevanlinna representation

$$G(z) = p - gq + \int_{\mathbb{R}} \frac{1 + sz}{z - s} \, d\rho(s), \quad z \in \mathbb{C} \setminus \text{supp } \rho,$$

for some $p \in \mathbb{R}, q \in [0, +\infty)$, and positive finite Borel measure $\rho$ on $\mathbb{R}$. If $(\alpha, \beta) \cap \text{supp } \rho = \emptyset$, then $G$ is analytic, and takes real values, on $(\alpha, \beta)$. In particular, if $\{t_n, x_n\} \subset \Omega$ is such that $(t_n, x_n) \to \partial \Omega$ and $W(t_n, x_n) \to \gamma \in (\alpha, \beta)$, then $g(t_n, x_n) = G(W(t_n, x_n)) \to G(\gamma) \in (a, b)$. The function $G$ is known (and easily seen) to be strictly decreasing on intervals in the complement of supp $\rho$. Thus necessarily $G([\alpha, \beta]) = [a, b]$, with $b = \lim_{s \uparrow \alpha} G(z)$. Now assume towards contradiction that $(\alpha, \beta) \cap \text{supp } \rho \neq \emptyset$. Then there exists at least one point $\gamma \in (\alpha, \beta)$ where the nontangential limit of $G$ exists, and the nontangential limit of $\text{Im } G$ belongs to $[\infty, 0)$. As $W$ is a homeomorphism, its functional inverse $W^{-1}$ is a well-defined continuous bijective map from $\mathbb{C}^+$ onto $\Omega$. In particular, $W^{-1}(\gamma + i(0, 1]) \subset \Omega$ is a simple path that approaches $\partial \Omega$. For any sequence $\{y_n\}_{n \in \mathbb{N}}, y_n \searrow 0$, such that $\{t_n, x_n\} = W^{-1}(\gamma + iy_n)] \subset \Omega$ converges (necessarily to a point in the boundary of $\Omega$), we have

$$\lim_{n \to \infty} \text{Im } g(t_n, x_n) = \lim_{n \to \infty} \text{Im } G(W(t_n, x_n)) = \lim_{n \to \infty} \text{Im } G(W^{-1}(\gamma + iy_n))) = \lim_{n \to \infty} \text{Im } G(z_n) = \lim_{n \to \infty} \text{Im } G(z) \in [-\infty, 0).$$

As seen just before Lemma 7.4, this forces $t_n \to 1$. However, points in $(\alpha, \beta)$ are necessarily limits of sequences $W(t_n, x_n)$ with $t_n$ converging to a number in $[0, 1 - \varepsilon]$ for some $\varepsilon > 0$ (see also the proof of Lemma 7.4). This is a contradiction. Thus, $(\alpha, \beta) \cap \text{supp } \rho = \emptyset$, as claimed.

We write the Nevanlinna representation as

$$G(z) = p - gq + \int_{(\alpha, \beta) \cup (\beta, \infty)} \frac{1 + sz}{z - s} \, d\rho(s), \quad z \in \mathbb{C} \setminus \text{supp } \rho;$$

(normally one would integrate on the closed intervals, but $G([\alpha, \beta]) = [a, b] \subset \mathbb{R}$ implies $\rho(\{\alpha\}) = \rho(\{\beta\}) = 0$). Its derivative is

$$G'(z) = -q - \int_{(\alpha, \beta) \cup (\beta, \infty)} \frac{1 + s^2}{(z - s)^2} \, d\rho(s).$$

Thus, we have $G'(x) < 0$ for any $x \in \mathbb{R} \setminus \text{supp } \rho$, and in particular on $(\alpha, \beta)$. We shall next find a convenient domain in $\mathbb{C}$ containing $(\alpha, \beta)$ on which we can guarantee that $-\text{Re } G'$ is greater than zero. If
For any $s \in (-\infty, \alpha)$, if $|y| \leq x - \alpha < s - x \implies \int_{(-\infty, \alpha) \cup (\beta, \infty)} \frac{(x-s)^2-y^2}{((x-s)^2+y^2)^2} (1+s^2) \, \mathrm{d}\rho(s) \geq 0$, and for any $s \in (\beta, +\infty)$, if $|y| \leq \beta - x < s - x \implies \int_{(\beta, +\infty)} \frac{(x-s)^2-y^2}{((x-s)^2+y^2)^2} (1+s^2) \, \mathrm{d}\rho(s) \geq 0$, with at least one of the two integrals being strictly positive. Thus, $-\Re G'(x+iy) > 0$ on $D = \{x+iy: \alpha \leq x \leq \beta, |y| \leq \min\{x-\alpha, \beta-x\}\}$. If there exists some $0 < \eta < +\infty$ such that $\rho(\alpha-\eta, \alpha) = 0$ and/or $\rho(\beta, \beta+\eta) = 0$ then we may increase the size of $D$ the obvious way.

It is obvious that $D$ is convex. Since $-\Re G'$ is strictly positive on $D$, it follows that $G$ is injective on $D$ and $G(\partial D)$ is a simple closed curve in $\mathbb{C}$, symmetric with respect to $\mathbb{R}$. We have $G(\alpha) = 0, G(\beta) = a$.

Note that $0 > \Re G'(x) = \lim_{y \to 0} \frac{\mathrm{Im} G(x+iy) - \mathrm{Im} G(x)}{y} = \lim_{y \to 0} \frac{\mathrm{Im} G(x+iy)}{y}$, so that $G'(x)$ being bounded away from zero provides a lower bound for the vertical thickness of $G(D)$ at any given $x \in (\alpha, \beta)$. More specifically, $G$ is conformal on $D$ so that, by Koebe’s distortion theorem (see, for instance, Kari Astala, Tadeusz Iwaniec, and Gaven Martin [? , Theorem 2.10.6]), we have

$$\frac{|G'(x)|}{4} \text{dist}(x, \partial D) \leq \text{dist}(G(x), \partial G(D)) \leq |G'(x)| \text{dist}(x, \partial D).$$

The shape of our domain $D$ guarantees that $\text{dist}(x, \partial D) = \min\{x-\alpha, \beta-x\}/\sqrt{2}$, thus allowing us to conclude with $K = G(D \cap \mathbb{C}^+)$. (We note that in our context, this provides a universal lower bound $\min\{|G'(v)|: \alpha \leq v \leq \beta\} \leq 4\sqrt{2} \max\{|y|: G(x)+it \in G(D) \forall t \in [0, y]\}$.

Denote $D^c = D \cap \mathbb{C}^c$. The relation $g = G \circ W$ and the fact that $W: \Omega_0 \to \mathbb{C}^+$ is a homeomorphism provides us with a set $O = W^{-1}(D^c)$ as claimed in our proposition.

Recall that $g = G \circ W, f = F \circ W$. Trying to find a map $\Phi$ such that $\Phi \circ g = f$ is equivalent to finding $\Phi$ such that $\Phi \circ G \circ W = F \circ W$ on some subset of $\Omega_0$. Since $W$ is a homeomorphism, it is enough to find $\Phi$ such that $\Phi \circ G = F$ on some relevant domain inside $C^-$. We simply define $\Phi: G(D^c) \to C^+$, $\Phi(z) = F(G^{-1}(z))$, where the inverse is the one taking values in $D$. This completes the proof. 

We recall that, unlike quasiregular functions, “most” analytic functions are determined by their values at the frontier. In particular, a function defined on a domain whose boundary contains an interval from $\mathbb{R}$ and with values in a half-plane as above is determined by its (known to exist a.e.) nontangential limits on any set of nonzero measure (the Fatou and Riesz-Privalov Theorems - see [? , Theorems 2.5 and 8.1]). In the following, we prove Theorem 2.15, namely that $f(0, x) = \lim_{t \to 0} \Phi(g(t, x)) = \lim_{t \to 0} \Phi(z)$, where, as before, $\lim \Phi(z)$ denotes the nontangential limit of $\Phi$ at $x$. This holds of course $\nu$-a.e. almost everywhere on the interval $[a, b]$ in question. In order for our result to be non-vacuous, we need to assume that $\nu$ has a nonzero absolutely continuous part in $[a, b]$, and, in particular, that $a < b$.

**Lemma 7.6.** Let $D \subseteq \mathbb{C}^+$ be a rectangle such that $\partial D \cap \mathbb{R}$ is an interval whose interior contains zero, and consider a non-constant analytic function $\omega: D \to \mathbb{C}^+$. Suppose that the nontangential limit of $\omega$ at zero exists and belongs to $\mathbb{C}^+$. Then there exists $1 > \varepsilon > 0$ and a smooth path $\gamma: (0, \varepsilon) \to \mathbb{C}^+$ such that:

1. $\gamma(t) = t \omega(\gamma(t)), t \in (0, \varepsilon]$;
2. $\lim_{t \to 0} \gamma(t) = 0$;
3. $\lim_{t \to 0} \omega(\gamma(t))$ exists and equals the nontangential limit of $\omega$ at zero.

Moreover, for $\varepsilon$ small enough, the path $\gamma$ satisfying properties 1, 2, and 3 above is unique.
Proof. Let \( t \in \mathbb{C}^+ \) denote the nontangential limit of \( \omega \) at zero. Consider a cone
\[
\Gamma_c = \{ z \in \mathbb{C}^+ : |\text{Re } z| < c |\text{Im } z| \},
\]
choose \( c = 1 + 2 |\text{Re } l|/|\text{Im } l| \) (so that \( l \in \Gamma_c \)), and denote \( \Gamma_c(\eta) = \{ z \in \Gamma_c : \text{Im } z < \eta \} \). Since zero belongs to the interior of the interval which is the intersection of the boundary of \( D \) with the real line, there is an \( \eta > 0 \) sufficiently small such that \( \Gamma_c(\eta) \subset D \).

By definition of nontangential limit, there exists \( \eta > 0 \) such that \( \omega(\Gamma_c(\eta)) \subset \Gamma_c(\eta) \), and from the continuity of \( \omega \) on the closure of \( \Gamma_c(\eta) \) we conclude that the set \( \omega(\Gamma_c(\eta)) \) is bounded. Thus, there exists \( \varepsilon > 0 \) such that \( t \omega(\Gamma_c(\eta)) \subset \Gamma_c(\eta) \) for all \( t \in (0, \varepsilon] \). Fix now such a \( t \). The analytic function \( \varphi : \Gamma_c(\eta) \to \Gamma_c(\eta) \) defined by \( \varphi(z) = t \omega(z) \) has, according to the Denjoy-Wolff Theorem [7, 7], a unique interior fixed point (observe that the point must indeed be interior, since zero is not a fixed point, and \( t \omega(\Gamma_c(\eta)) \) is a proper subset of \( \Gamma_c(\eta) \cup \{0\} \)). Denote this point by \( \gamma(t). \) The implicit function theorem guarantees the smoothness of the correspondence \( t \mapsto \gamma(t) : \) indeed, according to the Schwarz-Pick lemma, \( |\varphi'(\gamma(t))| < 1. \) This proves the first part of the lemma.

The second part follows from the fact that \( \gamma(t) \in t \omega(\Gamma_c(\eta)) \) and the diameter of the set \( t \omega(\Gamma_c(\eta)) \) tends to zero as \( t \to 0 \).

Item 3 follows from item 2, the fact that \( \gamma(t) \in \Gamma_c \) for all \( t \in (0, \varepsilon] \), and Fatou’s Theorem.

Assume towards contradiction that there exists another path \( \delta : (0, \varepsilon] \to D \) satisfying conditions 1-3 in the Lemma. Observe that both \( \gamma \) and \( \delta \) are right inverses for the function \( \Psi : D \to \mathbb{C} \setminus (-\infty, 0] \) defined by \( \Psi(z) = z/\omega(z) \), and thus they are injective. Moreover, \( \gamma((0, \varepsilon]) \cap \delta((0, \varepsilon]) = \emptyset \). Indeed, assume that \( \gamma(t_1) = \delta(t_2) \). Then \( t_1 = \Psi(\gamma(t_1)) = \Psi(\delta(t_2)) = t_2. \) Denote \( s = t_1 = t_2. \) We have
\[
\Psi'((\gamma(s))) = s^2 \omega(\gamma(s)) - \alpha(s) \omega'(\gamma(s))/\gamma(s)^2 = s^2 \gamma(s)(1/s - \omega'(\gamma(s))/\gamma(s)^2),
\]
so the derivative of \( \Psi \) in the point \( \gamma(s) \) is zero if and only if \( 1 - s \omega'(\gamma(s)) = 0 \), or, equivalently, if \( \varphi'((\gamma(s)) = 1. \) But \( \gamma(s) \) is the Denjoy-Wolff point of \( \varphi(z) = s \omega(z) \). Since, as seen above, \( z \mapsto \varphi(z) \) sends \( \Gamma_c(\eta) \) strictly inside itself, we obtain a contradiction with the Schwarz-Pick Lemma. We conclude that the derivative of \( \Psi \) in the point \( \gamma(s) \) cannot be zero, so that \( \Psi \) must be injective on some neighborhood of \( \gamma(s) \), and hence \( \gamma \) and \( \delta \) must coincide on a whole subinterval of \( (0, \varepsilon] \) centered at \( s \), and hence on all \( (0, s]. \) This is a contradiction. So indeed \( \gamma((0, \varepsilon]) \cap \delta((0, \varepsilon]) = \emptyset. \)

Now consider the open, connected and simply connected set \( \mathcal{D}_0 \subset D \) delimited by \( \gamma, \delta \), and a third simple smooth curve \( \beta \) included in \( D \) which has its endpoints at \( \gamma(\varepsilon/2) \) and \( \delta(\varepsilon/2) \), intersects \( \gamma((0, \varepsilon]) \cup \delta((0, \varepsilon]) \) in no other point, and such that zero belongs to the closure of \( \mathcal{D}_0 \). Observe that \( \Psi(\gamma(t)) = \Psi(\delta(t)) = t \), so by a theorem of Lindelöf [7, Theorem 2.3.1] applied to \( \Psi \) on \( D \), and a corollary of the Iverson Theorem [7, Theorem 5.2] applied to \( \Psi \) on \( \mathcal{D}_0 \), we have \( \lim_{t \to 0, z \in \mathcal{D}_0} \Psi(z) = 0. \) Of course, \( \Psi(\mathcal{D}_0) \) is in its own turn an open connected set. Since \( \Psi(\gamma((0, \varepsilon/2])) = \Psi(\delta((0, \varepsilon/2])) = (0, \varepsilon/2] \), and so, as seen above, \( \lim_{t \to 0, z \in \mathcal{D}_0} \Psi(z) = 0 \), we must have that \( 0 \in \Psi(\mathcal{D}_0). \) Since \( \Psi \) is an analytic map, which is open, \( \partial \Psi(\mathcal{D}_0) \subseteq \Psi(\partial \mathcal{D}_0) = \Psi(\hat{\gamma} \cup \gamma(\beta)). \) \( \Psi(\mathcal{D}_0) \) is a bounded open connected set, so its topological boundary is a compact set in \( \mathbb{C}. \) This compact set must thus be included in the continuous curve \( [0, \varepsilon/2] \cup \Psi(\beta). \) Therefore, \( \Psi(\beta) \) must describe a curve in \( \mathbb{C} \) which, together with \( [0, \varepsilon/2] \), surrounds an open connected set (it may enter it too, but must surround it entirely). By construction, \( \Psi \) maps the two ends of \( \beta \) in \( \varepsilon/2 \), so \( \Psi(\beta) \) is a closed curve in \( \mathbb{C}. \) We note that this closed curve cannot be included in the complement of \((-\infty, 0)\), and conclude that \( \Psi(\beta) \cap (-\infty, 0) \neq \emptyset. \) But \( \beta \subset D, \) and \( \Psi(z) = z/\omega(z), \) which is a product of two numbers one has positive imaginary part, and one has negative imaginary part. So \( \Psi(\beta) \) is contained in \( \mathbb{C} \setminus (-\infty, 0) \). This is a contradiction. So indeed \( \gamma \) is unique. \[\square\]

Proof of Theorem 2.15. Pick an arbitrary connected component \( \Omega_0. \) With the notations from Proposition 7.5, consider a point \( x \in \mathbb{R} \) such that \( (0, \varepsilon) \in \overline{\Omega_0 \setminus \{(0, a), (0, b)\}} \) and \( \lim \Phi(\varepsilon) \in \mathbb{C}^+. \) We let \( \omega(z) = -\Phi(x + z). \) By Proposition 7.5, this function is defined on a small rectangle included in the lower
half-plane and having zero at the middle of its upper edge. According to Lemma 7.6, there exists a unique path γ in this rectangle such that γ(t) = Ω(γ(t)) = −tΦ(x + γ(t)). However, at the same time −tf(t, x) = −tΦ(x − tf(t, x)). The uniqueness part of Lemma 7.6 guarantees that γ(t) = −tf(t, x), and items 1–3 of the same lemma allow us to conclude.

Now we finally can state that

\[ f(0, x) = \lim_{t \to 0} \Phi(g(t, x)) = \lim_{s \to x} \Phi(z). \]

This relation holds for \( \nu^{ac} \)-almost all \( x \in \mathbb{R} \). That is, for each connected component of \( \Omega \), we find a set \( O \) on which the above can be written. From this relation, it follows immediately that

\[ z \mapsto \Phi(z) - \int_{\mathbb{R}} \frac{d\nu(s)}{z - s} \]

is an analytic function on the intersection of the domain of \( \Phi \) with \( \mathbb{C}^- \). Its nontangential limits are real a.e. on \( \mathbb{R} \cap \partial K \). Thus, as seen before, since \( \Phi(z) - \int_0^1 (1/(z - s))d\nu(s) \) has real nontangential limits a.e. on the relevant subset of \( \mathbb{R} \), either the Schwarz reflection principle applies at a given point \( x \), or the cluster set of the function at \( x \) is equal to \( \overline{\mathbb{C}} \) or to \( \mathbb{C} \).

We focus next on the issue of proving Corollary 2.16 (the uniqueness of the Brownian bridge when provided with the complete initial data, i.e. with \( f(0, \cdot) \)). The following lemma clarifies the relation between \( \Phi \) and \( f(0, \cdot) \) if \( \nu \) contains a non-singular part.

**Lemma 7.7.** With the notations from Proposition 7.5, assume that \( \text{supp} \nu^{ac} \cap \overline{\mathbb{R}}_0 \neq \emptyset \). Then \( f(0, \cdot) \) determines uniquely \( \Phi \) on \( \Omega_0 \).

**Proof.** We continue to use the notations from Proposition 7.5 and its proof. It has been noted just after the proof of Theorem 2.15 that if \( \text{supp} \nu^{ac} \cap \partial \Omega_0 \neq \emptyset \), then \( \lim_{s \to x} \Phi(z) \) for \( \nu^{ac} \)-almost all \( x \), that is for all \( x \) in a set of nonzero Lebesgue measure. An application of [7, Theorem 8.1] guarantees that \( \Phi \) is determined by these values. \( \square \)

**Proof of Corollary 2.16.** For simplicity of notations we write \( f(t, x) = f^{\nu^{ac}}(t, x), f'(t, x) = f^{\nu^{ac}}(t, x) \), and let

\[ g(t, x) = x - tf(t, x), \quad g'(t, x) = x - tf'(t, x). \]

We restrict ourselves to two connected subsets \( \Omega_0 \subset \Omega \) and \( \Omega'_0 \subset \Omega' \) such that \( \{0\} \times \mathbb{R} \cap (\partial_1 \Omega_0) \cap (\partial_1 \Omega'_0) \neq \emptyset \) (see notations in the proof of Proposition 7.5, which we shall use throughout this proof as well). According to Proposition 2.5 2 and 3, in order to prove our corollary, it is enough to show that \( f = f' \) (and thus, in particular, \( \Omega_0 = \Omega'_0 \)). By Proposition 7.5, there exist sets \( K, K' \subset \mathbb{C}^- \) and maps \( \Phi, \Phi' \) such that \( \partial K \cap \partial K' \supset (\{0\} \times \mathbb{R}) \cap (\partial_1 \Omega_0) \cap (\partial_1 \Omega'_0) \), \( \Phi: K \to \mathbb{C}^+, \Phi': K' \to \mathbb{C}^+ \) are analytic and satisfy \( \Phi \circ g = f, \Phi' \circ g' = f' \). According to our hypothesis, the restriction of \( \nu^{ac} \) to the connected component(s) of \( \text{supp} \nu \) included in the intersection \( \{0\} \times \mathbb{R} \cap (\partial_1 \Omega_0) \cap (\partial_1 \Omega'_0) \) is non-zero, so that, by Theorem 2.15, there exist points \( c \) in this set such that \( \lim_{t \to 0} f(t, c), \lim_{t \to 0} f'(t, c) \) exist, are equal, and \( \lim_{t \to 0} \text{Im} f(t, c) = \lim_{t \to 0} \text{Im} f'(t, c) \in (0, \infty) \). Lemma 7.7 guarantees that \( \Phi = \Phi' \) on their (nonempty) joint domain, which means that they are extensions of each other to the respective domains. Lemma 7.6 forces \( f(t, c) = f'(t, c) \) for all points \( c \) as above, and \( t \in (0, 1) \) sufficiently small. As seen in the proof of Lemma 7.6, the existence and uniqueness of the solution \( f(t, c) \) to the equation \( \Phi \circ g = f \) is provided via Denjoy-Wolff Theorem applied to a properly chosen domain (in order to prove the existence of a solution) and the implicit function theorem (in order to prove analyticity of the correspondence in \( t \)). This last result however provides, via the relation \( \Phi'(x - tf(t, x))(1 - t\partial_x f(t, x)) = \partial_x f(t, x) \) and the analytic implicit function theorem, the analyticity of the correspondence \( x \mapsto f(t, x) \) on a neighborhood.

58
(in \(C\)) of a given \(c \in R\) as above, for fixed \(t > 0\) such that \(|t\Phi'(x-tf(t,x))| < 1\). Thus, \(f, f'\) have analytic extensions as functions of two complex variables to an open set in \(C^2\). Thanks to their equality on an open subset of \(R^2\) showed above, we have \(f = f'\) on this whole open set, and necessarily on all of their common domain of analyticity.

The existence and differentiability (established independently - see Proposition 2.5) of the solution \(f\) guarantees that the set \(\{(t,x) \in O: \Phi(g(t,x)) = \frac{-1}{2}\}\) is empty (recall from Proposition 7.5 that \(g(O) = K\)). Indeed, otherwise the equality \(\Phi'(x-tf(t,x))(1-t\partial_t f(t,x)) = \partial_x f(t,x)\) would imply \(0 = \frac{1}{2}\), an obvious contradiction. In particular, this shows the implicit function theorem argument from the above applies on all of \(O \cap O'\) in order to conclude that \(f' = f\) on this set.

Proposition 7.5 guarantees that the upper part of the boundary of the union of the two sets \(K, K'\) covers entirely the convex hull of the components of \(\text{supp} f, f\). Thus, to conclude that \(\Phi = \Phi\) extends to an open, simply connected set that contains this convex hull in its boundary. Thus, the domain of analyticity of each of \(f, f'\) extends to a simply connected subset \(C\) of \(\Omega_0 \cup \Omega'_0\) which contains in its boundary all of the components of \(\text{supp} \nu\) included in the union \(\partial \Omega_0 \cup \partial \Omega'_0\), and on which, as seen above, \(f = f'\).

Consider an arbitrary \(s \in (0, 1)\). Define \(g_t(t,x) = x - (t-s)f(t,x) = g(t,x) + sf(t,x)\). As seen above, \(g_s\) is automatically a solution of the Beltrami equation (7.2) on \(\Omega_0\). Moreover, \(g_s(s,x) = x, \text{Im} g_s(t,x) < 0\) for \(1 > t > s\), \(\text{Im} g_s(t,x) > 0\) for \(0 < t < s\). Lemma 7.4 applies to \(g_s\) on \(\{s\} \times R\), and thus there exists a set\(^2\) \(K_s \subseteq C^-\), a set \(O_s \subseteq \{(t,x) \in \Omega_0: 1 > t > s\}\), and an analytic function \(\Phi_s\) as in Proposition 7.5, with the only difference that \(\Phi_s(g_s(t,x)) = f(t,x), f(t,x) \in O_s\). In addition, the proof of Proposition 7.5 applies without modification to show that there exist \(L_s \subseteq C^+, P_s \subseteq \{(t,x) \in \Omega_0: s > t > 0\}\), and an analytic function \(\Psi_s\) such that \(\Psi_s(g_s(t,x)) = f(t,x), f(t,x) \in P_s\). The similar objects derived from \(\Omega'_0\) and \(f'\) will be denoted the same way, except that each will receive a ‘\(^*\).’

Returning now to the question of the equalities \(f = f', \Omega_0 = \Omega'_0\), consider an \((s,x) \in \Omega_0 \cap \Omega'_0\). We claim that \(f(s,x) = f'(s,x)\). (Note again that trivially if \(f = f'\) on \(\Omega_0 \cap \Omega'_0\), then \(\Omega_0 = \Omega'_0\).) If \((s,x) \in C\), then there is nothing to prove. If \((s,x) \in \{s\} \times I\) for some interval \(I \subseteq R\) such that \(\{s\} \times I \cap C \neq \emptyset\), then we apply the considerations above to find sets \(L_s, L'_s, P_s, P'_s\), and functions \(\Psi, \Psi'\) such that \(\Psi_s(g_s(t,x)) = f(t,x), \Psi'_s(g'_s(t,x)) = f'(t,x)\) for \((t,x) \in P_s \cap P'_s\). The arguments above yield the existence of a connected set \(C_s\) in \(P_s \cup P'_s\) containing \(\{s\} \times I\) in its boundary such that \(f = f'\) on all of \(C_s\), and hence in \((s,x)\). We immediately observe that this guarantees the equality \(f = f'\) on the whole subset below \(\{s\} \times I\). More precisely, we look at the segments \(\{(r) \times R\} \cap \Omega_0 \cap \Omega'_0\) for each \(r \in (0, 1)\). If \(\tilde{s}\) such that \(\tilde{s} \times R \cap C \neq \emptyset\), then there are disjoint intervals \(I_1, I_2, \ldots, I_k\) such that \(\{s\} \times I \cap C \neq \emptyset\), and a (possibly smaller) subfamily \(\{j_1, \ldots, j_k\} \subseteq \{I_1, I_2, \ldots, I_k\}\) such that \(\{\tilde{s}\} \times I_i \cap C \neq \emptyset, 1 \leq i \leq k\). Then the set of points \((r,x)\) that can be connected to \(\{\tilde{s}\} \times I_i\) for some \(i \in \{1, \ldots, k\}\) by a path in \(\Omega_0\) starting at \((r,x)\) and whose first coordinate does not decrease satisfies the condition that \(f(r,x) = f'(r,x)\) and implicitly that \((r,x) \in \Omega_0 \cap \Omega'_0\). Thus, we have succeeded in proving that the set of points on which \(f = f'\) is bounded in \((0, 1) \times R\) by a family of segments \((s_j) \times I_j\). Moreover, below (in the sense just described) these segments, \(\Omega_0\) and \(\Omega'_0\) coincide. For simplicity, re-denote this set by \(C\).

Finally, assume that \((s,x)\) is above \(C\), meaning that, with the notation \(I\) from the above, \((\{s\} \times I) \cap C = \emptyset\). For any segment \((r) \times J\) bordering \(C\), we may perform the construction described in the first part of the proof in order to increase \(C\) strictly above the “level” \(r\) (i.e. to find \(r' > r\) and an interval \(J' \subset R\) such that points from \((r) \times J\) can be united to points from \((r') \times J'\) by smooth paths whose first coordinate in \((0, 1) \times R\) does not decrease). Continuity of \(f\) guarantees that this process will reach any “level” which is strictly less than 1, and in particular level \(s\).

Recalling that the definition of \(\Omega\) is given as the set in \((0, 1) \times R\) on which \(|\text{Im} f| > 0\) allows us to conclude that \(\Omega_0 = \Omega'_0\), and thus complete the proof.

---

\(^2\)It may be that \(\{(t,x) \in \Omega_0: 1 > t > s\}\) and/or \(\{(t,x) \in \Omega_0: s > t > 0\}\) are not connected anymore, but our arguments apply as well to each connected component of these sets.
We record next two facts about the free Brownian bridge which might be of some independent interest.

**Corollary 7.8.** The function \( f \) reaches its supremum on \( \partial \Omega \). Moreover, \( \text{Im}[f(t,x)] \) reaches its supremum on \( \{0,1\} \times \mathbb{R} \), \( -\text{Im}[g(t,x)] \) on \( \{1\} \times \mathbb{R} \), and \( \text{Im}[h(t,x)] \) on \( \{0\} \times \mathbb{R} \).

**Proof.** Let us recall that \( f \) satisfies Beltrami’s equation (7.2) and thus, \( f \) is an open mapping. In particular, as an open mapping, it cannot have a local maximum inside \( \Omega \). We shall argue that on each simply connected component \( \Omega_0 \) of \( \Omega \), the imaginary part \( \text{Im}[f] \) can reach its maximum only on \( \partial \Omega_0 \).

Since on \( \partial \Omega \cap \{(0, 1) \times \mathbb{R}\} \) we know that \( \text{Im}[f] \) is zero, it remains that this maximum is reached at a point of either \( \{0\} \times \mathbb{R} \) or \( \{1\} \times \mathbb{R} \).

Thus, assume towards contradiction that there exists a component \( \Omega_0 \) of \( \Omega \) (simply connected by Lemma 7.2), a point \((t_0, x_0) \in \Omega_0 \) and a neighbourhood \( V_0 \subseteq \Omega_0 \) of it so that \( \text{Im}[f(t, x)] \leq \text{Im}[f(t_0, x_0)] \) for all \((t, x) \in V_0 \) (denote for simplicity \( c = \text{Im}[f(t_0, x_0)] \in (0, +\infty) \)). By shrinking \( V_0 \) if necessary, we may assume that \( V_0 \subset \Omega_0 \). This means that

\[
\text{Im}[f(V_0)] \subset \{ z \in \mathbb{C}^+ | \text{Im}[z] \leq c \},
\]

and in addition that \( f(V_0) \cap (\mathbb{R} + ic) \neq \emptyset \), as it contains the point \( f(t_0, x_0) \). But any neighbourhood of \( f(t_0, x_0) \) contains elements from \( \{ z \in \mathbb{C}^+ : \text{Im}[z] > c \} \), so the point \( f(t_0, x_0) \) is in the boundary of the set \( f(V_0) \) while \((t_0, x_0) \) belongs to the open set \( V_0 \). This contradicts the openness of \( f \) at \((t_0, x_0) \). The last two statements are obvious consequences of the previous. \( \square \)

Finally we observe that the function \( \Phi \) found in Proposition 7.5 is intimately related to the function \( u_0^* \) identified in Theorem 2.3 in the case when \( \partial_2 \Omega_0 \) and/or \( \partial_4 \Omega_0 \) is not mapped by \( g \) to a single point. In that case (with the notations form Proposition 7.5) there exists a \( \delta > 0 \) such that \( \Phi \) reflects analytically through the complement of \( \text{supp} \nu \) in the interval \([a - \delta, b + \delta]\) (see Proposition 7.5). We have

**Corollary 7.9.** With the notations of Proposition 7.5, and Proposition 2.5, assume that \( \nu \) is absolutely continuous with respect to the Lebesgue measure. Then the function \( u_0^* \) is defined on \([a - \delta, b + \delta] \setminus \text{supp} \nu \) as \( u_0^*(x) = \Phi(x) \).