LARGE DEVIATIONS FOR FUNCTIONS OF TWO 
RANDOM PROJECTION MATRICES

FUMIO HIAI$^{1,2}$ AND DÉNES PETZ$^{1,3}$

Abstract. In this paper two independent and unitarily invariant projection matrices $P(N)$ and $Q(N)$ are considered and the large deviation is proven for the eigenvalue density of all polynomials of them as the matrix size $N$ converges to infinity. The result is formulated on the tracial state space $TS(A)$ of the universal $C^*$-algebra $A$ generated by two selfadjoint projections. The random pair $(P(N), Q(N))$ determines a random tracial state $\tau_N \in TS(A)$ and $\tau_N$ satisfies the large deviation. The rate function is in close connection with Voiculescu’s free entropy defined for pairs of projections.

Mathematics Subject Classification: 15A52, 60F10, 46L54.
Key words: Eigenvalue density, large deviation, random matrices, free entropy, universal $C^*$-algebra, tracial state space.

Introduction

Large deviation results for the empirical eigenvalue density of random matrices started with the paper of Ben Arous and Guionnet [2] in which generalized Wigner theorem concerning Gaussian symmetric (or selfadjoint) matrices was proven. The paper was followed by large deviation results for several other kind of random matrices (as Wishart, etc); see the monograph [9] for a detailed discussion and the survey [8] for more recent developments.

Up to now the typical large deviation results on random matrices have dealt with the empirical eigenvalue density of a certain sequence of matrices; occasionally these matrices were algebraically expressed from two (as in [12]). In this paper two independent projection matrices are considered and the large deviation is proven for all polynomials (even for more general functions) of them. More precisely, the main result is a $C^*$-algebraic formulation of large deviations for the sequence of two random selfadjoint projection matrices $P(N)$ and $Q(N)$ having independent and unitarily invariant distribution provided moreover $\alpha := \lim_N \text{rank}(P(N))/N$ and $\beta := \lim_N \text{rank}(Q(N))/N$ exist. The main theorem is formulated on the tracial state space $TS(A)$ of the universal $C^*$-algebra $A := \mathcal{C}^*(\mathbb{Z} \ast \mathbb{Z})$ generated by two selfadjoint projections $e$ and $f$. The random pair $(P(N), Q(N))$ determines a random tracial state $\tau_N \in TS(A)$ as follows:

$$\tau_N(h) = \frac{1}{N} \text{Tr}(\psi(h)), \quad h \in A,$$

where $\psi : A \to M_N(\mathbb{C})$ is the unique $^*$-homomorphism such that $\psi(e) = P(N)$ and $\psi(f) = Q(N)$. The random $\tau_N$ induces a measure $\nu_N$ on $TS(A)$ and the sequence $\nu_N$ satisfies the large deviation principle in the scale $1/N^2$ with a rate function $I : TS(A) \to [0, \infty]$ in the ordinary sense. It is very remarkable that the rate function $I$ is in close relation with

---

1 Supported in part by Japan-Hungary JSPS-HAS Joint Project.
2 Supported in part by Strategic Information and Communications R&D Promotion Scheme of MPHPT.
3 Supported in part by OTKA T032662.
Voiculescu's free entropy $\chi(p,q)$ defined for a pair of projections in a $W^*$-probability space. Namely, the GNS-construction from $(\mathcal{A}, \tau)$ yields a $W^*$-probability space $(\pi_{\tau}(\mathcal{A})'', \tilde{\tau})$ and for the projections $p = \pi_{\tau}(e)$ and $q = \pi_{\tau}(f)$, we have $I(\tau) = -\chi(p,q)$.

The result includes a bunch of traditional large deviation results for the eigenvalue density of different polynomials of $P(N)$ and $Q(N)$. The corresponding rate function can be obtained from $I$ by the contraction principle and computed explicitly in some examples as $P(N)Q(N) + Q(N)P(N)$ and $aP(N) + bQ(N)$.

The paper is organized as follows. First we establish a large deviation theorem for the empirical eigenvalue density of the random matrix $P(N)Q(N)P(N)$. This result is obtained via the joint eigenvalue density and the cases $\alpha + \beta \leq 1$ and $\geq 1$ are somewhat separated but treated parallel. A few facts about the Jacobi ensemble are used here. Since polynomials of two projections are easily controlled by the powers of $P(N)Q(N)P(N)$, we can move to the $C^*$-algebraic formulation mentioned above. The tracial state space $TS(A)$ has a convenient representation in terms of four numbers and a measure on $(0,1)$. The large deviation theorem or more precisely the rate function is first identified in terms of the representation of tracial states and the description à la Voiculescu comes afterwards. The last section is the application of the contraction principle and contains very concrete computations.

1. Joint distribution of two projections

Let $M_N(\mathbb{C})$ be the algebra of $N \times N$ complex matrices. By an $N \times N$ random projection matrix $P$ we always mean a random orthogonal (or selfadjoint) projection matrix, and the unitary invariance of $P$ means that the distribution of $VPV^*$ is equal to that of $P$ for any unitary $V \in M_N(\mathbb{C})$.

The aim of this section is to analyze the joint distribution of two independent and unitarily invariant random projection matrices $P,Q$ in $M_N(\mathbb{C})$, when their ranks $\text{rank}(P) = k$ and $\text{rank}(Q) = l$ are fixed; we may assume that $0 \leq k \leq l \leq N$. Throughout this section, we keep these assumptions on $P$ and $Q$.

The joint eigenvalue distribution of $PQP$ is related to the Jacobi ensemble. Let $(A,B)$ be an independent pair of $N \times N$ complex Wishart matrices of $p$ degrees of freedom and of $q$ degrees of freedom, respectively, that is, $A = YY^*$ and $B = ZZ^*$ with complex $N \times p$ and $N \times q$ random matrices $Y$ and $Z$ such that $\text{Re}Y_{ij}, \text{Im}Y_{ij}, \text{Re}Z_{ij}$ and $\text{Im}Z_{ij}$ are independent standard Gaussians. Assume here that $p,q \geq N$. Then the random positive semidefinite matrix

$$(A + B)^{-1/2}A(A + B)^{-1/2}$$

is called an $N \times N$ Jacobi ensemble of parameter $(p - N, q - N)$. It has the probability distribution

$$\text{Constant} \times \text{Det}(X)^{p-N}\text{Det}(I - X)^{q-N}1_{\{0 \leq X \leq I\}}(X) \, dX \quad (1.1)$$

on the space of $N \times N$ selfadjoint matrices (see [4, Lemma 2.1]), where $1_{\{0 \leq X \leq I\}}$ denotes the characteristic function of $\{X \in M_N(\mathbb{C}) : 0 \leq X \leq I\}$. The density formula (1.1) implies the joint distribution of the eigenvalues

$$\text{Constant} \times \prod_{i=1}^{N} x_i^{p-N}(1 - x_i)^{q-N} \prod_{1 \leq i < j \leq N} (x_i - x_j)^2 \prod_{i=1}^{N} 1_{[0,1]}(x_i) \, dx_i,$$

see also [5] or [7, Chapter 2].
The next lemma is from [4, Theorem 2.2].

**Lemma 1.1.** Assume that \( k + l \leq N \). Then \( PQP \), when considered as a random matrix in \( M_k(\mathbb{C}) = PM_N(\mathbb{C})P \), has the distribution of a Jacobi ensemble of parameter \( (l - k, N - k - l) \).

Hence, the joint eigenvalue distribution of the nonzero eigenvalues of \( PQP \) is given by

\[
\frac{1}{Z_{N,k,l}} \prod_{i=1}^{k} x_i^{l-k}(1-x_i)^{N-k-l} \prod_{1 \leq i < j \leq k} (x_i - x_j)^2 \prod_{i=1}^{k} 1_{[0,1]}(x_i) \, dx_i
\]

with a normalization constant \( Z_{N,k,l} \).

Let \((A, B)\) and \((A', B')\) be pairs of selfadjoint \( N \times N \) random matrices. We say that they have the same joint distribution if

\[
\text{tr}_N(h(A, B)) = \text{tr}_N(h(A', B')) \quad \text{almost surely}
\]

for any polynomial \( h \) of two non-commuting variables, where \( \text{tr}_N \) denotes the normalized trace on \( M_N(\mathbb{C}) \).

Our strategy is to modify the pair \((P, Q)\) of projections in such a way that they are easy to handle but their joint distribution does not change. As the first step, we may assume that \((P, Q)\) are of the forms

\[
P = I_k \oplus 0_{N-k}, \quad Q = U(I_l \oplus 0_{N-l})U^*,
\]

where \( I_k \oplus 0_{N-k} \) stands for the diagonal matrix whose \( k \) first diagonal entries are 1 and the remaining are 0, and \( U \) is an \( N \times N \) Haar-distributed random unitary matrix. In this way, randomness belongs to only \( Q \), while \( P \) is a constant projection matrix.

**Proposition 1.2.**

(a) If \( k + l \leq N \), then the joint distribution of \((P, Q)\) coincides with that of the pair

\[
P \quad \text{and} \quad \begin{bmatrix} \sqrt{X} & \sqrt{X(I_k - X)} & 0 & 0 \\ \sqrt{X(I_k - X)} & I_k - X & 0 & 0 \\ 0 & 0 & I_{l-k} & 0 \\ 0 & 0 & 0 & 0_{N-k-l} \end{bmatrix},
\]

where \( X := \text{Diag}(x_1, \ldots, x_k) \) and \((x_1, \ldots, x_k) \in [0,1]^k \) is distributed under the distribution (1.2).

(b) If \( k + l > N \), then the joint distribution of \((P, Q)\) coincides with that of the pair

\[
P \quad \text{and} \quad \begin{bmatrix} I_{k+l-N} & 0 & 0 & 0 \\ 0 & \sqrt{X(I_{N-l} - X)} & 0 & 0 \\ 0 & 0 & I_{l-k} & 0 \\ 0 & 0 & 0 & 0_{N-k-1} \end{bmatrix},
\]

where \( X := \text{Diag}(x_1, \ldots, x_{N-l}) \) and \((x_1, \ldots, x_{N-l}) \in [0,1]^{N-l} \) is distributed under

\[
\frac{1}{Z_{N,k,l}} \prod_{i=1}^{N-l} x_i^{l-k}(1-x_i)^{k+l-N} \prod_{1 \leq i < j \leq N-l} (x_i - x_j)^2 \prod_{i=1}^{N-l} 1_{[0,1]}(x_i) \, dx_i.
\]
Proof. (a) Assume \( k+l \leq N \). By the structure theorem of two projections (see [14, pp. 306–308]), after a (random) unitary conjugation, \((P,Q)\) can be represented as

\[
P = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \oplus I \oplus I \oplus 0 \oplus 0,
\]

\[
Q = \begin{bmatrix} X \sqrt{X(I-X)} \\ \sqrt{X(I-X)} \sqrt{X(I-X)} \end{bmatrix} \oplus I \oplus 0 \oplus I \oplus 0,
\]

where \( 0 \leq X \leq I \) with \( \ker X = \{0\} \) and \( \ker (I-X) = \{0\} \) on \( \mathcal{H}_0 \), under a decomposition

\[
\mathbb{C}^N = (\mathcal{H}_0 \otimes \mathbb{C}^2) \oplus \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}_3 \oplus \mathcal{H}_4.
\]

(Note that \( \mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3 \) and \( \mathcal{H}_4 \) are the ranges of \( P \land Q, P \land Q^\perp, P^\perp \land Q \) and \( P \lor Q^\perp \), respectively, and some of them may be zero spaces.) Since \((PQ|_{\mathcal{H}_0} = X \oplus I \oplus 0 \) on \( \mathcal{H}_0 \oplus \mathcal{H}_1 \oplus \mathcal{H}_2 \), it follows from Lemma 1.1 that \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) are zero spaces almost surely. This shows that there exists an \( N \times N \) random unitary matrix \( V \) such that

\[
V PV^* = P = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} \oplus 0_{l-k} \oplus 0_{N-k-l},
\]

\[
V QV^* = \begin{bmatrix} X \sqrt{X(I-X)} \\ \sqrt{X(I-X)} \sqrt{X(I-X)} \end{bmatrix} \oplus I_{l-k} \oplus 0_{N-k-l},
\]

where \( X = \text{Diag}(x_1, \ldots, x_k) \) and \((x_1, \ldots, x_k) \in [0,1]^k\) is distributed under (1.2) by Lemma 1.1. Hence we have the desired conclusion.

(b) Next, assume \( k+l > N \); then since \( N-l < k \) and \( (N-l) + k \leq N \), one can apply the above case (a) to \((I-Q,P)\) instead of \((P,Q)\). Thus, the joint distribution of \((I-Q,P)\) is almost surely equal to that of the pair

\[
\begin{bmatrix} I_{N-l} & 0 \\ 0 & 0 \end{bmatrix} \oplus 0_{k+l-N} \oplus 0_{l-k}
\]

and

\[
\begin{bmatrix} X \sqrt{X(I_N-I-N)} \\ \sqrt{X(I_N-I-N)} \sqrt{X(I_N-I-N)} \end{bmatrix} \oplus I_{N+m-N} \oplus 0_{m-n}
\]

so that \((P,Q)\) has the same joint distribution almost surely as the pair

\[
\begin{bmatrix} X \sqrt{X(I_N-I-N)} \\ \sqrt{X(I_N-I-N)} \sqrt{X(I_N-I-N)} \end{bmatrix} \oplus I_{k+l-N} \oplus 0_{l-k}
\]

and

\[
\begin{bmatrix} 0 & 0 \\ 0 & I_{N-l} \end{bmatrix} \oplus I_{k+l-N} \oplus I_{l-k}.
\]

Here, \( X = \text{Diag}(x_1, \ldots, x_{N-l}) \) and \((x_1, \ldots, x_{N-l}) \in [0,1]^{N-l}\) is distributed under

\[
\frac{1}{Z_{N,N-l,k}} \prod_{i=1}^{N-l} x_i^{k+l-N}(1-x_i)^{l-k} \prod_{1 \leq i<j \leq N-l} (x_i-x_j)^2 \prod_{i=1}^{N-l} 1_{[0,1]}(x_i) \, dx_i. \tag{1.4}
\]

Since \[
\begin{bmatrix} X \sqrt{X(I_N-I-N)} \\ \sqrt{X(I_N-I-N)} \sqrt{X(I_N-I-N)} \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 0 \\ 0 & I_{N-l} \end{bmatrix}
\]

are respectively transformed into

\[
\begin{bmatrix} I_{N-l} & 0 \\ 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} I_{N-l} - X \sqrt{X(I_N-I-N)} \\ \sqrt{X(I_N-I-N)} \end{bmatrix}
\]

by a conjugation by the unitary matrix
From Proposition 1.2 we can readily obtain joint eigenvalue distributions of some polynomials of $P$ and $Q$. For example, we have:

**Corollary 1.3.**

(i-a) When $k + l \leq N$, the eigenvalues of $PQP$ (or $PQ$) are given as

\[
0, \ldots, 0, x_1, \ldots, x_k
\]

where $k$ times

and the joint distribution of $(x_1, \ldots, x_k)$ is (1.2).

(i-b) When $k + l > N$, the eigenvalues of $PQP$ (or $PQ$) are given as

\[
0, \ldots, 0, 1, \ldots, 1, x_1, \ldots, x_{N-l},
\]

where $k$ times

and the joint distribution of $(x_1, \ldots, x_{N-l})$ is (1.3).

(ii-a) When $k + l \leq N$, the eigenvalues of $PQ + PQ$ are given as

\[
0, \ldots, 0, x_1 \pm \sqrt{x_1}, \ldots, x_k \pm \sqrt{x_k}
\]

where $k$ times

and the joint distribution of $(x_1, \ldots, x_k)$ is (1.2).

(ii-b) When $k + l > N$, the eigenvalues of $PQ + PQ$ are given as

\[
0, \ldots, 0, 2, \ldots, 2, x_1 \pm \sqrt{x_1}, \ldots, x_{N-l} \pm \sqrt{x_{N-l}},
\]

where $k$ times

and the joint distribution of $(x_1, \ldots, x_{N-l})$ is (1.3).

(iii-a) When $k + l \leq N$ and $a, b \in \mathbb{R} \setminus \{0\}$, the eigenvalues of $aP + bQ$ are given as

\[
0, \ldots, 0, a + b - x_1, \ldots, a + b - x_k,
\]

where $k$ times

and the joint distribution of $(x_1, \ldots, x_k)$ is

\[
\frac{2^k}{|ab|^{k(N-k)}Z_{N,k,l}} \prod_{i=1}^{k} \frac{1}{x_i - a + b} \sum \frac{1}{(x_i - x)(x_i - b)} \left( x_i - a \right) \left( x_i - b \right) \left( x_i - a + b - x_i \right)^{N-k-l} \times \prod_{1 \leq i < j \leq k} \frac{1}{(x_i - x_j)^2(a + b - x_i - x_j)^2} \prod_{i=1}^{k} 1_{[A,B]}(x_i) dx_i, \quad (1.5)
\]

where $Z_{N,k,l}$ is the normalization constant in (1.2) and $A, B$ are the first two smallest numbers of $0, a, b, a + b$.

(iii-b) When $k + l > N$ and $a, b \in \mathbb{R} \setminus \{0\}$, the eigenvalues of $aP + bQ$ are given as

\[
b, \ldots, b, a + b, \ldots, a + b, x_1, \ldots, x_{N-l}, a + b - x_1, \ldots, a + b - x_{N-l},
\]

where $k$ times

and the joint distribution of $(x_1, \ldots, x_k)$ is
and the joint distribution of \((x_1, \ldots, x_{N-1})\) is
\[
\frac{2^{N-l}}{|ab|^{l(N-l)}Z_{N,k,l}} \prod_{i=1}^{N-l} \left| \frac{a+b}{2} \right| \prod_{1 \leq i<j \leq N-l} (x_i - x_j)^2 (a + b - x_i - x_j)^2 \prod_{i=1}^{N-l} 1_{[A,B]}(x_i) \, dx_i,
\]
where \(Z_{N,k,l}\) is the normalization constant in (1.3) and \(A, B\) are as in (iii-a).

**Proof.** (i-a) is Lemma 1.1 and (i-b) is immediate from Proposition 1.2 (b).

(ii-a) By Proposition 1.2 (a) we may assume that
\[
PQ +QP = \begin{bmatrix} 2X & \sqrt{X(I_k - X)} \\ \sqrt{X(I_k - X)} & 0 \end{bmatrix} \oplus 0_{N-2k},
\]
where \(X\) is as in Proposition 1.2 (a). Then the result immediately follows because the eigenvalues of the \(2 \times 2\) matrix
\[
\begin{bmatrix} 2x & \sqrt{x(1-x)} \\ \sqrt{x(1-x)} & 0 \end{bmatrix}
\]
for \(0 \leq x \leq 1\) are \(x \pm \sqrt{x}\). The proof of (ii-b) is similar by Proposition 1.2 (b).

(iii-a) By Proposition 1.2 (a) we may assume that
\[
ap + bQ = \begin{bmatrix} aI_k + bX & b\sqrt{X(I_k - X)} \\ b\sqrt{X(I_k - X)} & b(I_k - X) \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.
\]
The eigenvalues of the \(2 \times 2\) matrix
\[
\begin{bmatrix} a + bx & \sqrt{x(1-x)} \\ \sqrt{x(1-x)} & b(1-x) \end{bmatrix}
\]
for \(0 \leq x \leq 1\) are
\[
a + b \pm \sqrt{(a-b)^2 + 4abx}.
\]
Set \(t_i := \frac{a + b - \sqrt{(a-b)^2 + 4abx}}{2}\) for \(1 \leq i \leq k\). Then the eigenvalues of \(aP + bQ\) are
\[
0, \ldots, 0, \underbrace{b, \ldots, b}_{N-2k \text{ times}}, t_1, \ldots, t_k, a + b - t_1, \ldots, a + b - t_k,
\]
and \((t_1, \ldots, t_k)\) is supported in \([A, B]^k\). By noting that
\[
x_i = \frac{(t_i - a)(t_i - b)}{ab}, \quad 1 - x_i = \frac{t_i(a + b - t_i)}{ab}, \quad \frac{dx_i}{dt_i} = 2 \left( t_i - \frac{a + b}{2} \right),
\]
the form (1.5) of the joint distribution of \((x_1, \ldots, x_k)\) can be directly computed from (1.2). The proof of (iii-b) is similar. \(\square\)

2. LARGE DEVIATION FOR \(PQP\)

From now on, for each \(N \in \mathbb{N}\) let \((P(N), Q(N))\) be a pair of independent and unitarily invariant random projection matrices in \(M_N(\mathbb{C})\) with non-random ranks \(k(N) := \text{rank}(P(N))\) and \(l(N) := \text{rank}(Q(N))\). Throughout what follows, we assume that \(k(N)/N \to \alpha\) and \(l(N)/N \to \beta\) as \(N \to \infty\) for some \(\alpha, \beta \in [0,1]\). Our goal is to obtain a large deviation
theorem for the empirical eigenvalue density of \( P(N)Q(N)P(N) \). Concerning large deviation theory, our general reference is [6], but [9] contains many matrix examples.

We have already observed that the two cases \( \alpha + \beta \leq 1 \) and \( \alpha + \beta \geq 1 \) are slightly different. To treat them parallel, we set

\[
n_0(N) := N - \min\{k(N), l(N)\}, \quad n_1(N) := \max\{k(N) + l(N) - N, 0\},
\]

\[
n(N) := N - n_0(N) - n_1(N) = \min\{k(N), l(N), N - k(N), N - l(N)\}.
\]

Then one can combine (i-a) and (i-b) of Corollary 1.3 to see that the eigenvalues of the \( N \times N \) selfadjoint random matrix \( P(N)Q(N)P(N) \) are

\[
\frac{1}{Z(N)} \prod_{i=1}^{n(N)} x_i^{k(N) - l(N)} (1 - x_i)^{l(N) - N} \prod_{1 \leq i < j \leq n(N)} (x_i - x_j)^2 \prod_{i=1}^{n(N)} 1_{[0,1]}(x_i) dx_i
\]

(2.1)

with a normalization constant \( Z(N) \).

When \( \mathcal{X} \) is a Polish space, let \( \mathcal{M}(\mathcal{X}) \) denote the set of all probability measures on \( \mathcal{X} \), which becomes a Polish space with respect to weak topology. For \( \mu \in \mathcal{M}(\mathbb{R}) \) let \( \Sigma(\mu) \) be Voiculescu's free entropy (or the minus of the logarithmic energy) of \( \mu \) defined by

\[
\Sigma(\mu) := \iint \log |x - y| d\mu(x) d\mu(y)
\]

(see [16] and [9, §5.3]). In particular, when \( \mu \) is compactly supported, \( \Sigma(\mu) \in [-\infty, +\infty) \) is well defined.

We first prove large deviation for the sequence of distributions (2.1) with slight modifications of notation.

**Proposition 2.1.** For each \( N \in \mathbb{N} \) consider the distribution

\[
\frac{1}{Z(N)} \prod_{i=1}^{n(N)} x_i^{\kappa(N)} (1 - x_i)^{\lambda(N)} \prod_{1 \leq i < j \leq n(N)} (x_i - x_j)^2 \prod_{i=1}^{n(N)} 1_{[0,1]}(x_i) dx_i
\]

(2.2)

on \([0,1]^{n(N)}\) with \( n(N) \in \mathbb{N}, \kappa(N), \lambda(N) \in [0, \infty) \) and a normalization constant \( Z(N) \). Assume that \( n(N)/N \to \rho, \kappa(N)/N \to \kappa \) and \( \lambda(N)/N \to \lambda \) as \( N \to \infty \) for some \( \rho \in (0, \infty) \) and \( \kappa, \lambda \in [0, \infty) \). Then:

1. The limit \( \lim_{N \to \infty} \frac{1}{n(N)} \log Z(N) \) exists and it equals \( \rho^2 B(\kappa/\rho, \lambda/\rho) \), where

\[
B(s, t) := \frac{(1 + s)^2}{2} \log(1 + s) - \frac{s^2}{2} \log s + \frac{(1 + t)^2}{2} \log(1 + t) - \frac{t^2}{2} \log t
\]

for \( s, t \geq 0 \).

2. When \( (x_1, \ldots, x_{n(N)}) \) is distributed under (2.2), the empirical measure

\[
\delta_{x_1} + \cdots + \delta_{x_{n(N)}}
\]

(2.3)
satisfies the large deviation principle in the scale $1/N^2$ with the rate function

$$I(\mu) := -\rho^2 \Sigma(\mu) - \rho \int_{0}^{1} (\kappa \log x + \lambda \log (1-x)) \, d\mu(x) + \rho^2 B \left( \frac{\kappa}{\rho}, \frac{\lambda}{\rho} \right)$$

(2.4)

for $\mu \in \mathcal{M}([0,1])$. Moreover, there exists a unique minimizer $\mu_0 \in \mathcal{M}([0,1])$ of $I(\mu)$ with $I(\mu_0) = 0$.

Proof. (1) The Selberg integral formula (see [11, §17.1]) gives

$$Z(N) = \int_{[0,1]^{n(N)}} \prod_{i=1}^{n(N)} x_i^{\kappa(N)} (1-x_i)^{\lambda(N)} \prod_{1 \leq i < j \leq n(N)} (x_i - x_j)^2 \prod_{i=1}^{n(N)} dx_i$$

$$= \prod_{j=1}^{n(N)} \frac{\Gamma(j+1) \Gamma(j+\kappa(N)) \Gamma(j+\lambda(N))}{\Gamma(2) \Gamma(j+n(N)+\kappa(N)+\lambda(N))}.$$

By using the Stirling formula, under neglecting the small order $o(N)$, we compute

$$\frac{1}{N^2} \log Z(N)$$

$$= \frac{1}{N^2} \left\{ \sum_{j=1}^{n(N)} \left( j \log j + \sum_{j=1}^{n(N)} (j + \kappa(n)) \log (j + \kappa(n)) + \sum_{j=1}^{n(N)} (j + \rho(n)) \log (j + \rho(n)) \right) \right.$$

$$- \sum_{j=1}^{n(N)} (j + n + \kappa(n) + \rho(n)) \log (j + n + \kappa(n) + \rho(n)) \right\}$$

$$= \frac{n(N)}{N^2} \left\{ \sum_{j=1}^{n(N)} \frac{j}{n(N)} \log \left( \frac{j}{n(N)} \right) + \sum_{j=1}^{n(N)} \left( \frac{j}{n(N)} + \frac{\kappa}{\rho} \right) \log \left( \frac{j}{n(N)} + \frac{\kappa}{\rho} \right) \right.$$

$$+ \sum_{j=1}^{n(N)} \left( \frac{j}{n(N)} + \frac{\lambda}{\rho} \right) \log \left( \frac{j}{n(N)} + \frac{\lambda}{\rho} \right)$$

$$- \sum_{j=1}^{n(N)} \left( \frac{j}{n(N)} + 1 + \frac{\kappa}{\rho} + \frac{\lambda}{\rho} \right) \log \left( \frac{j}{n(N)} + 1 + \frac{\kappa}{\rho} + \frac{\lambda}{\rho} \right) \right\}.$$

Therefore,

$$\lim_{N \to \infty} \frac{1}{N^2} \log Z(N)$$

$$= \rho^2 \left\{ \int_{0}^{1} x \log x \, dx + \int_{0}^{1} \left( x + \frac{\kappa}{\rho} \right) \log \left( x + \frac{\kappa}{\rho} \right) \, dx + \int_{0}^{1} \left( x + \frac{\lambda}{\rho} \right) \log \left( x + \frac{\lambda}{\rho} \right) \, dx \right.$$

$$- \int_{0}^{1} \left( x + 1 + \frac{\kappa}{\rho} + \frac{\lambda}{\rho} \right) \log \left( x + 1 + \frac{\kappa}{\rho} + \frac{\lambda}{\rho} \right) \, dx \right\}$$

$$= \rho^2 B \left( \frac{\kappa}{\rho}, \frac{\lambda}{\rho} \right).$$

(2) Denote the distribution (2.2) by $\nu_{n(N)}$ and define the probability measure $P_N$ on $\mathcal{M}([0,1])$ by

$$P_N(\Lambda) := \nu_{n(N)} \left( \{ x \in [0,1]^{n(N)} : \mu_x \in \Lambda \} \right)$$
for Borel subsets \( \Lambda \) of \( \mathcal{M}([0,1]) \), where \( \mu_x \) denotes the empirical measure (2.3) for \( x = (x_1,\ldots,x_{n(N)}) \). Define the kernel functions on \([0,1]^2\) as follows:

\[
F(x,y) := -\log |x-y| - \frac{\kappa}{2\rho} (\log x + \log y) - \frac{\lambda}{2\rho} (\log(1-x) + \log(1-y)),
\]

\[
F_R(x,y) := \min \{ F(x,y), R \} \quad \text{for } R > 0.
\]

Furthermore, for each \( N \in \mathbb{N} \) we define

\[
\tilde{F}_N(x,y) := -\log |x-y| - \delta_{\kappa>0} \frac{\kappa(N)}{2n(N)} (\log x + \log y)
- \delta_{\lambda>0} \frac{\lambda(N)}{2n(N)} (\log(1-x) + \log(1-y)),
\]

\[
\tilde{F}_{N,R}(x,y) := \min \{ \tilde{F}_N(x,y), R \} \quad \text{for } R > 0,
\]

where \( \delta_{\kappa>0} = 1 \) if \( \kappa > 0 \), \( \delta_{\kappa>0} = 0 \) if \( \kappa = 0 \), and \( \delta_{\lambda>0} \) is similar. Then we observe the following:

(i) \( \tilde{F}_{N,R}(x,y) \leq -\log |x-y| - \frac{\kappa(n)}{2n(N)} (\log x + \log y) - \frac{\lambda(n)}{2n(N)} (\log(1-x) + \log(1-y)) \) for all \( x,y \in [0,1] \).

(ii) For any \( R > 0 \), \( \tilde{F}_{N,R}(x,y) \) converges to \( F_R(x,y) \) uniformly for \( x,y \in [0,1] \) as \( N \to \infty \).

In fact, (i) is obvious by the definition of \( \tilde{F}_{N,R}(x,y) \). For (ii) assume that \( \kappa,\lambda > 0 \) (the proof is similar for other cases). For \( \delta > 0 \) set

\[
T_\delta := \{ (x,y) \in [0,1]^2 : \delta \leq x \leq 1-\delta, \delta \leq y \leq 1-\delta, |x-y| \geq \delta \}.
\]

For any \( R > 0 \) there exist \( \delta > 0 \) and \( N_0 \in \mathbb{N} \) such that \( F(x,y) \geq R \) and \( \tilde{F}_N(x,y) \geq R \) for all \( (x,y) \in [0,1]^2 \setminus T_\delta \) and \( N \geq N_0 \). Obviously, \( \tilde{F}_N(x,y) \) converges to \( F(x,y) \) uniformly on \( T_\delta \) as \( N \to \infty \), and the assertion follows.

According to general theory of large deviations ([6]), the stated large deviation is shown when we prove the following two inequalities for every \( \mu \in \mathcal{M}([0,1]) \):

\[
\inf_G \left[ \limsup_{N \to \infty} \frac{1}{N^2} \log P_N(G) \right] \leq -\rho^2 \iint F(x,y) d\mu(x) d\mu(y) - C, \quad (2.5)
\]

\[
\inf_G \left[ \liminf_{N \to \infty} \frac{1}{N^2} \log P_N(G) \right] \geq -\rho^2 \iint F(x,y) d\mu(x) d\mu(y) - C, \quad (2.6)
\]

where \( C := \rho^2 B(\kappa/\rho,\lambda/\rho) \) and \( G \) runs over neighborhoods of \( \mu \).
Proof of (2.5). For every neighborhood $G$ of $\mu \in \mathcal{M}([0,1])$, setting $\tilde{G} := \{ x \in [0,1]^{n(N)} : \mu_x \in G \}$, by the above (i) we have
\[ P_N(G) = \nu_{n(N)}(\tilde{G}) = \frac{1}{Z(N)} \int_{\tilde{G}} \prod_{i=1}^{n(N)} x_i^{\kappa_i(N)} (1-x_i)^{\lambda_i(N)} \prod_{1 \leq i < j \leq n(N)} (x_i-x_j)^2 \prod_{i=1}^{n(N)} dx_i \]
\[ \leq \frac{1}{Z(N)} \int_{\tilde{G}} \prod_{i=1}^{n(N)} x_i^{\kappa_i(N)/n(N)} (1-x_i)^{\lambda_i(N)/n(N)} \times \exp \left( -2 \sum_{1 \leq i < j \leq n(N)} \bar{F}_{N,R}(x_i,x_j) \prod_{i=1}^{n(N)} dx_i \right) \]
\[ \leq \frac{1}{Z(N)} \left( \int_0^1 x^{\kappa_i(N)/n(N)} (1-x)^{\lambda_i(N)/n(N)} dx \right)^{n(N)} \times \exp \left( -n(N)^2 \inf_{\mu' \in G} \int \int \bar{F}_{N,R}(x,y) d\mu'(x) d\mu'(y) + n(N)R \right). \]

Since the above fact (ii) implies that
\[ \lim_{N \to \infty} \left( \inf_{\mu' \in G} \int \int \bar{F}_{N,R}(x,y) d\mu'(x) d\mu'(y) \right) = \inf_{\mu' \in G} \int \int F_{R}(x,y) d\mu'(x) d\mu'(y), \]
we get
\[ \lim_{N \to \infty} \frac{1}{N^2} \log P_N(G) \leq -\rho^2 \inf_{\mu' \in G} \int \int F_{R}(x,y) d\mu'(x) d\mu'(y) - C \]
thanks to (1). Furthermore, appealing to the continuity of $\mu' \mapsto \int \int F_{R}(x,y) d\mu'(x) d\mu'(y)$, we obtain
\[ \inf_{\mu} \left[ \limsup_{N \to \infty} \frac{1}{N^2} \log P_N(G) \right] \leq -\rho^2 \int \int F_{R}(x,y) d\mu(x) d\mu(y) - C \]
so that (2.5) follows by letting $R \to +\infty$.

Proof of (2.6). If $\mu$ has an atom at 0 or 1, then $\int \int F(x,y) d\mu(x) d\mu(y) = +\infty$ so that we have nothing to do. Otherwise, letting $d\mu_\delta(x) := \mu([\delta,1-\delta])^{-1}1_{[\delta,1-\delta]}(x) d\mu(x)$, we get
\[ \int \int F(x,y) d\mu(x) d\mu(y) = \lim_{\delta \searrow 0} \int \int F(x,y) d\mu_\delta(x) d\mu_\delta(y). \]
Also it is immediate to see that $\mu \in \mathcal{M}([0,1]) \mapsto \inf \left\{ \liminf_{N \to \infty} \frac{1}{N^2} \log P_N(G) : G \text{ is a neighborhood of } \mu \right\}$ is upper semicontinuous. Hence we may assume that $\mu$ is supported in $[a,b]$ with $0 < a < b < 1$. For $\varepsilon > 0$ let $\phi_\varepsilon \geq 0$ be a $C^\infty$-function supported in $[-\varepsilon, \varepsilon]$ such that $\int \phi_\varepsilon(x) dx = 1$. Then we get $\Sigma(\phi_\varepsilon * \mu) \geq \Sigma(\mu)$ (see [9, p. 216]) as well as
\[ \lim_{\varepsilon \searrow 0} \int \log x d(\phi_\varepsilon * \mu)(x) = \int \log x d\mu(x), \]
\[ \lim_{\varepsilon \searrow 0} \int \log(1-x) d(\phi_\varepsilon * \mu)(x) = \int \log(1-x) d\mu(x). \]
so that $\mu$ may be assumed to have a continuous density. Furthermore, by the concavity of $\Sigma(\mu)$, it suffices to prove (2.6) for $(1 - \varepsilon)\mu + \varepsilon m$ for each $0 < \varepsilon < 1$, where $m$ is the uniform measure on an interval including the support $\text{supp} \mu$. After all, we can assume that $\mu$ has a continuous density $f > 0$ on $\text{supp} \mu = [a, b]$ with $0 < a < b < 1$ and $\delta \leq f(x) \leq \delta^{-1}$ on $[a, b]$ for some $\delta > 0$.

For each $N \in \mathbb{N}$ let

$$a < a_1^{(N)} < b_1^{(N)} < a_2^{(N)} < \cdots < a_{n(N)}^{(N)} < b_{n(N)}^{(N)}$$

be such that

$$\int_a^{a_i^{(N)}} f(x) \, dx = \frac{i - \frac{1}{2}}{n(N)}, \quad \int_a^{b_i^{(N)}} f(x) \, dx = \frac{i}{n(N)}, \quad 1 \leq i \leq n(N);$$

then

$$b_i^{(N)} - a_i^{(N)} \geq \frac{\delta}{2n(N)}, \quad 1 \leq i \leq n(N).$$

Define

$$\Delta_{n(N)} := \{x = (x_1, \ldots, x_{n(N)}) \in [0, 1]^{n(N)} : a_i^{(N)} \leq x_i \leq b_i^{(N)}, 1 \leq i \leq n(N)\}.$$ 

For any neighborhood $G$ of $\mu$, whenever $N$ is large enough, we have

$$\Delta_{n(N)} \subset \tilde{G} := \{x \in [0, 1]^{n(N)} : \mu_x \in G\}$$

so that

$$P_N(G) = \nu_{n(N)}(\tilde{G})$$

$$\geq \frac{1}{Z(N)} \int_{\Delta_{n(N)}} \prod_{i=1}^{n(N)} x_i^{\kappa(N)} (1 - x_i)^{\lambda(N)} \prod_{1 \leq i < j \leq n(N)} (x_i - x_j)^2 \prod_{i=1}^{n(N)} dx_i$$

$$\geq \frac{1}{Z(N)} \left( \frac{\delta}{2n(N)} \right)^{n(N)} \prod_{i=1}^{n(N)} (a_i^{(N)})^{\kappa(N)} (1 - b_i^{(N)})^{\lambda(N)} \prod_{1 \leq i < j \leq n(N)} (a_j^{(N)} - b_i^{(N)})^2.$$ 

With $g : [0, 1] \to [a, b]$ being the inverse function of $t \in [a, b] \mapsto \int_a^t f(x) \, dx$, since $a_i^{(N)} = g\left((i - \frac{1}{2})/n(N)\right)$ and $b_i^{(N)} = g(i/n(N))$, we have

$$\lim_{N \to \infty} \frac{\kappa(N)}{N^2} \sum_{i=1}^{n(N)} \log a_i^{(N)} = \rho \int_0^1 \log g(t) \, dt = \rho \int \log x \, d\mu(x),$$

$$\lim_{N \to \infty} \frac{\kappa(N)}{N^2} \sum_{i=1}^{n(N)} \log(1 - b_i^{(N)}) = \rho \int_0^1 \log(1 - g(t)) \, dt = \rho \int \log(1 - x) \, d\mu(x),$$

$$\lim_{N \to \infty} \frac{2}{N^2} \sum_{1 \leq i < j \leq n(N)} \log(a_j^{(N)} - b_i^{(N)})$$

$$= 2\rho^2 \int \int_{0 \leq s < t \leq 1} \log(g(t) - g(s)) \, ds \, dt = \rho^2 \Sigma(\mu).$$

These estimates altogether imply (2.6).

The proof of the large deviation is now completed, and the existence of a unique minimizer of the rate function is known as a general result on weighted logarithmic energy functionals (see [13, I.1.3]).
Now, the large deviation theorem for the random matrix $P(N)Q(N)P(N)$ can be easily shown from Proposition 2.1. Set
\begin{equation}
\rho := \min\{\alpha, \beta, 1 - \alpha, 1 - \beta\},
\end{equation}
and
\begin{equation}
C := \rho^2B\left(\frac{|\alpha - \beta|}{\rho}, \frac{|\alpha + \beta - 1|}{\rho}\right)
\end{equation}
(meant zero if $\rho = 0$), and denote by $\mathcal{M}((0,1))$ the set of all probability measures on $[0,1]$ with no atoms at 0 and 1.

**Theorem 2.2.** The empirical eigenvalue density of $P(N)Q(N)P(N)$ satisfies the large deviation principle in the scale $1/N^2$ with the rate function $\tilde{I}(\tilde{\mu})$ for $\tilde{\mu} \in \mathcal{M}((0,1))$ given as follows: If
\begin{equation}
\tilde{\mu} = (1 - \min\{\alpha, \beta\})\delta_0 + \max\{\alpha + \beta - 1, 0\}\delta_1 + \rho\mu
\end{equation}
with $\mu \in \mathcal{M}((0,1))$, then
\begin{equation}
\tilde{I}(\tilde{\mu}) := -\rho^2\Sigma(\mu) - \rho|\alpha - \beta|\int_0^1 \log x \, d\mu(x) - \rho|\alpha + \beta - 1|\int_0^1 \log(1 - x) \, d\mu(x) + C;
\end{equation}
otherwise $\tilde{I}(\tilde{\mu}) = +\infty$. Moreover, a unique minimizer of $\tilde{I}(\tilde{\mu})$ is given by
\begin{equation}
\tilde{\mu}_0 := (1 - \min\{\alpha, \beta\})\delta_0 + \max\{\alpha + \beta - 1, 0\}\delta_1 + \int \sqrt{\frac{(x - \xi)(\eta - x)}{2\pi x(1 - x)}} 1_{[\xi, \eta]}(x) \, dx
\end{equation}
where
\begin{equation}
\xi, \eta := \alpha + \beta - 2\alpha\beta \pm \sqrt{4\alpha\beta(1 - \alpha)(1 - \beta)}.
\end{equation}
In particular, when $\rho = 0$, $\tilde{I}(\tilde{\mu})$ is identically $+\infty$ except at only $\tilde{\mu}_0 = (1 - \min\{\alpha, \beta\})\delta_0 + \max\{\alpha + \beta - 1, 0\}\delta_1$.

**Proof.** From the fact mentioned at the beginning of the section, the empirical eigenvalue density of $P(N)Q(N)P(N)$ is given by
\begin{equation}
\tilde{R}_N := \frac{n_0(N)}{N}\delta_0 + \frac{n_1(N)}{N}\delta_1 + \frac{n(N)}{N}R_N,
\end{equation}
where $R_N := \frac{1}{n(N)}(\delta_{x_1} + \cdots + \delta_{x_{n(N)}})$ and the joint distribution of $(x_1, \ldots, x_{n(N)})$ is (2.1).
First, assume that $\rho > 0$. Proposition 2.1 says that $(R_N)$ satisfies the large deviation in the scale $1/N^2$ with the rate function $I(\mu)$ for $\mu \in \mathcal{M}((0,1))$ given in (2.4) with $\kappa := |\alpha - \beta|$ and $\lambda := |\alpha + \beta - 1|$. We now proceed as in the proof of [9, 5.5.11]. Let $P_N$ and $\tilde{P}_N$ be the distributions on $\mathcal{M}((0,1))$ of $R_N$ and $\tilde{R}_N$, respectively; then
\begin{equation}
\tilde{P}_N(\Lambda) = P_N\left(\left\{\mu \in \mathcal{M}((0,1)) : \frac{n_0(N)}{n}\delta_0 + \frac{n_1(N)}{n}\delta_1 + \frac{n(N)}{n}\mu \in \Lambda\right\}\right)
\end{equation}
for $\Lambda \subset \mathcal{M}((0,1))$. Let $\mathcal{D}$ denote the set $\{\rho_0\delta_0 + \rho_1\delta_1 + \rho\mu : \mu \in \mathcal{M}((0,1))\}$, where $\rho_0 := 1 - \min\{\alpha, \beta\}$ and $\rho_1 := \max\{\alpha + \beta - 1, 0\}$. If $\tilde{\mu} \notin \mathcal{D}$, then $\tilde{\mu}(\{0\}) < \rho_0$ or $\tilde{\mu}(\{1\}) < \rho_1$ so that letting $\tilde{\mu}_0 := \epsilon < \rho_0$ (or $\tilde{\mu}_1 := \epsilon < \rho_1$) we have a neighborhood $\tilde{G} := \{\mu' \in \mathcal{M}((0,1)) : \tilde{\mu}'(\{0\}) < \epsilon$ (or $\tilde{\mu}'(\{1\}) < \epsilon$) of $\mu$. Since $\tilde{P}_N(\tilde{G}) = 0$ for large $N$, we get $\lim_{N \to \infty} \frac{1}{N} \log \tilde{P}_N(\tilde{G}) = -\infty$. Next, assume that $\tilde{\mu} \in \mathcal{D}$ and $\tilde{\mu} = \rho_0\delta_0 + \rho_1\delta_1 + \rho\mu$. For any
neighborhood \( \tilde{G} \) of \( \tilde{\mu} \), there exists a neighborhood \( G \) of \( \mu \) such that 

\[
\frac{n_0(N)}{N} \delta_0 + \frac{n_1(N)}{N} \delta_1 + \frac{n(N)}{N} G \subset \tilde{G}
\]

for large \( N \) and hence

\[
\liminf_{N \to \infty} \frac{1}{N^2} \log \tilde{P}_N(\tilde{G}) \geq \liminf_{N \to \infty} \frac{1}{N^2} \log P_N(G) \geq -I(\mu).
\]

On the other hand, for any neighborhood \( G \) of \( \mu \), there exists a neighborhood \( \tilde{G} \) of \( \tilde{\mu} \) such that

\[
\left( \frac{N}{n(N)} \tilde{G} - \frac{n_0(N)}{n(N)} \delta_0 - \frac{n_1(N)}{n(N)} \delta_1 \right) \cap \mathcal{M}([0,1]) \subset G,
\]

that is,

\[
\left\{ \mu \in \mathcal{M}([0,1]) : \frac{n_0(N)}{n} \delta_0 + \frac{n_1(N)}{n} \delta_1 + \frac{n(N)}{n} \mu \in \tilde{G} \right\} \subset G
\]

for large \( N \). Therefore,

\[
\inf_{\tilde{G}} \left[ \limsup_{N \to \infty} \frac{1}{n^2} \log \tilde{P}_N(\tilde{G}) \right] \leq \inf_{\tilde{G}} \left[ \limsup_{N \to \infty} \frac{1}{n^2} \log P_N(G) \right] \leq -I(\mu).
\]

Noting that \( \Sigma(\mu) = -\infty \) if \( \mu \in \mathcal{M}([0,1]) \) has an atom at 0 or 1, we obtain the desired large deviation for \( (\tilde{R}_N) \) when \( \rho > 0 \). The proof in the case \( \rho = 0 \) is similar to the above argument for \( \tilde{\mu} \notin \mathcal{D} \).

Finally, the existence of a unique minimizer of \( \tilde{I}(\tilde{\mu}) \) is already known by Proposition 2.1. To obtain the explicit form of the minimizer, we may apply a standard method in free probability theory. In fact, by the asymptotic freeness due to Voiculescu [15, Theorem 3.11] (see also [9, 4.3.5]), the joint distribution of \( (P(N), Q(N)) \) converges to that of \((p, q)\) where \( p \) and \( q \) are free projections in a tracial \( W^* \)-probability space \((\mathcal{M}, \tau)\) with \( \tau(p) = \alpha \) and \( \tau(q) = \beta \). The computation by use of \( S \)-transform in [18] says that the measure \((2.10)\) is the distribution measure of \( pqp \); hence it is the minimizer of \( \tilde{I}(\tilde{\mu}) \).

Note that the rate function \( \tilde{I}(\tilde{\mu}) \) is indeed lower semicontinuous and convex on \( \mathcal{M}([0,1]) \), which is of course a good rate function because of the compactness of \( \mathcal{M}([0,1]) \).

3. \( C^* \)-algebra formulation

The two-dimensional commutative \( C^* \)-algebra \( \mathbb{C} \oplus \mathbb{C} = C^*(\mathbb{Z}_2) \) is the universal \( C^* \)-algebra generated by a single orthogonal projection; hence the universal \( C^* \)-algebra generated two orthogonal projections is

\[
(C \oplus C) \star (C \oplus C) = C^*(Z \star Z)
\]

with projection generators \((1,0)\)'s in two components. As pointed out in [3, p. 14], one can see from the structure theorem for two projections ([14, pp. 306–308]) that \( C^*(Z \star Z) \) is isomorphic to an algebra of \( M_2(\mathbb{C}) \)-valued continuous functions on \([0,1] \); namely

\[
\mathcal{A} := \{ a \in C([0,1]; M_2(\mathbb{C})) : a(0) \text{ and } a(1) \text{ are diagonal} \},
\]

where the corresponding two projection generators are represented as

\[
e(t) := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad f(t) := \begin{bmatrix} t & \sqrt{t(1-t)} \\ \sqrt{t(1-t)} & 1-t \end{bmatrix} \quad \text{for } 0 \leq t \leq 1.
\]

We thus consider the above \( C^* \)-algebra \( \mathcal{A} \) with generators \( e, f \) as the universal \( C^* \)-algebra generated by two projections. We denote by \( TS(\mathcal{A}) \) the set of all tracial states on \( \mathcal{A} \), which
becomes a Polish space with respect to $w^*$-topology. The following lemma is a concrete description of $TS(A)$, the details are left to the reader.

**Lemma 3.1.** For each $\tau \in TS(A)$ there exist $\alpha_{11}, \alpha_{10}, \alpha_{01}, \alpha_{00} \geq 0$ with $\sum_{i,j=0}^{1} \alpha_{ij} \leq 1$ and $\mu \in \mathcal{M}((0, 1))$ such that

$$\tau(a) = \alpha_{10}a_1(0) + \alpha_{01}a_2(0) + \alpha_{11}a_1(1) + \alpha_{00}a_2(1) + \left(1 - \sum_{i,j=0}^{1} \alpha_{ij}\right) \int_{0}^{1} \text{tr}(a(t)) \, d\mu(t)$$

for all $a \in A$ with $a(0) = \text{Diag}(a_1(0), a_2(0))$ and $a(1) = \text{Diag}(a_1(1), a_2(1))$.

In this way, the set $TS(A)$ is parameterized by the set of all $(\{\alpha_{ij}\}_{i,j=0}^{1}, \mu)$ of $\alpha_{ij} \geq 0$, $\sum_{i,j=0}^{1} \alpha_{ij} \leq 1$ and $\mu \in \mathcal{M}((0, 1))$, and we write $\tau = (\{\alpha_{ij}\}_{i,j=0}^{1}, \mu)$ under this parameterization. But, note that $\mu$ is irrelevant if $\sum_{i,j=0}^{1} \alpha_{ij} = 1$. For $\tau = (\{\alpha_{ij}\}_{i,j=0}^{1}, \mu)$ we have

$$\tau(e) = \frac{1}{2}(1 + \alpha_{11} + \alpha_{10} - \alpha_{01} - \alpha_{00}),$$

$$\tau(f) = \frac{1}{2}(1 + \alpha_{11} - \alpha_{10} + \alpha_{01} - \alpha_{00}).$$

Furthermore, let $\pi_{\tau}$ be the GNS representation of $A$ associated with $\tau$ and $\tilde{\tau}$ be the normal extension of $\tau$ to $\pi_{\tau}(A)^{\prime\prime}$. Then, for $p := \pi_{\tau}(e)$ and $q := \pi_{\tau}(f)$ in $\pi_{\tau}(A)^{\prime\prime}$ we have

$$\tilde{\tau}(p \wedge q) = \alpha_{11}, \quad \tilde{\tau}(p \wedge q^\perp) = \alpha_{10}, \quad \tilde{\tau}(p^\perp \wedge q) = \alpha_{01}, \quad \tilde{\tau}(p^\perp \wedge q^\perp) = \alpha_{00}.$$

(3.1)

For any two projections $p, q$ in a tracial $W^*$-probability space $(\mathcal{M}, \tau)$, the universality property of $A$ shows that there exists a (unique) $*$-homomorphism $\psi_{p,q} : A \to \mathcal{M}$ such that $\psi_{p,q}(e) = p$ and $\psi_{p,q}(f) = q$. We simply write $h(p, q)$ for $\psi_{p,q}(h)$ for each $h \in A$, which may be regarded as a sort of “noncommutative functional calculus.” Then a tracial state $\tau_{p,q} \in TS(A)$ is defined by $\tau_{p,q}(h) := \tau(h(p, q))$ for $h \in A$. In particular, for $N \times N$ projection matrices $P, Q$, we have $\tau_{P,Q} \in TS(A)$ given by $\tau_{P,Q}(h) = \text{tr}_N(h(P, Q))$ for $h \in A$. When $P, Q$ are random projection matrices, $\tau_{P,Q}$ is a random tracial state on $A$ regarded as the “noncommutative empirical measure” of the pair $(P, Q)$. Its distribution measure on $TS(A)$ is defined by

$$\nu(A) := \text{Prob}(\{\tau_{P,Q} \in A\})$$

for Borel subsets $A \subset TS(A)$, where $\text{Prob}$ denotes probability measure of the underlying probability space where $P, Q$ are defined.

We are now in a position to state our main large deviation result formulated on the tracial state space $TS(A)$.

**Theorem 3.2.** For each $N \in \mathbb{N}$ let $(P(N), Q(N))$ be a pair of independent and unitarily invariant random projection matrices in $\mathcal{M}_N(\mathbb{C})$ such that $\text{rank}(P(N))/N \to \alpha$ and $\text{rank}(Q(N))/N \to \beta$ as $N \to \infty$. Let $\nu_N$ be the distribution measure of the random tracial state $\tau_N := \tau_{P(N), Q(N)}$ on $TS(A)$. Then $(\nu_N)$ satisfies the large deviation principle in the scale $1/N^2$ with rate function

$$I(\tau) := -\rho^2 \Sigma(\mu) - \rho|\alpha - \beta| \int_{0}^{1} \log x \, d\mu(x) - \rho|\alpha + \beta - 1| \int_{0}^{1} \log(1 - x) \, d\mu(x) + C$$
evaluated at \( \tau = (\{\alpha_{ij}\}_{i,j=0}^1, \mu) \in TS(A) \) if

\[
\begin{align*}
\alpha_{11} &= \max\{\alpha + \beta - 1, 0\}, \\
\alpha_{00} &= \max\{1 - \alpha - \beta, 0\}, \\
\alpha_{10} &= \max\{\alpha - \beta, 0\}, \\
\alpha_{01} &= \max\{\beta - \alpha, 0\},
\end{align*}
\]

otherwise \( \mathcal{I}(\tau) = +\infty. \) (See (2.7) and (2.8) for constants \( \rho \) and \( C. \))

Moreover, the unique minimizer of \( \mathcal{I} \) is the tracial state \( \tau_{p,q} \) corresponding to a pair \( (p,q) \) of free projections with trace values \( \alpha \) and \( \beta \).

**Proof.** First we notice that all mixed moments of \( e, f \) with respect to \( \tau \) are listed as \( \tau(e), \tau(f) \) and

\[
\tau((ef)^k) = \tau((fe)^k) = \tau((ef^2)^k) = \tau((ef)^k), \quad k \geq 1.
\]

Since the moments \( \tau((ef)^k), k \geq 1 \), determine the distribution of \( efe \) with respect to \( \tau \), one can define an affine homeomorphism \( \Psi \) of \( TS(A) \) with \( w^* \)-topology into \([0,1] \times [0,1] \times M([0,1])\) with product topology by \( \Psi(\tau) := (\tau(e), \tau(f), \mu) \) where \( \mu \) is the distribution measure of \( efe \) with respect to \( \tau \). For each \( \tau = (\{\alpha_{ij}\}_{i,j=0}^1, \mu) \in TS(A) \) let \( p := \pi_{\tau}(e) \) and \( q := \pi_{\tau}(f) \) in \((\pi_{\tau}(A)^*, \tilde{\tau})\), and let \( e_{pqp}(\cdot) \) be the spectral measure of \( pqp \). From the structure theorem for two projections, we get

\[
\tilde{\mu}(\{0\}) = \tilde{\tau}(e_{pqp}(\{0\})) = \frac{1}{2}\tilde{\tau}(1 - p \wedge q - p \perp q - p \perp q - p \perp q - q \perp q - q \perp q) + \tilde{\tau}(p \perp q + p \perp q + p \perp q - q \perp q)
\]

\[
= \frac{1}{2}(1 - \alpha_{11} + \alpha_{10} + \alpha_{01} + \alpha_{00})
\]

and

\[
\tilde{\mu}(\{1\}) = \tilde{\tau}(e_{pqp}(\{1\})) = \tilde{\tau}(p \wedge q) = \alpha_{11}
\]

thanks to (3.1). Hence it is straightforward to check that \( \tau \) satisfies (3.2) if and only the following hold:

\[
\begin{align*}
\tau(e) &= \alpha, \\
\tau(f) &= \beta, \\
\tilde{\mu}(\{0\}) &= 1 - \min\{\alpha, \beta\}, \\
\tilde{\mu}(\{1\}) &= \max\{\alpha + \beta - 1, 0\}.
\end{align*}
\]

Furthermore, in this case we obviously have

\[
\tilde{\mu} = (1 - \min\{\alpha, \beta\})\delta_0 + \max\{\alpha + \beta - 1, 0\}\delta_1 + \rho\mu,
\]

where

\[
\rho = \min\{\alpha, \beta, 1 - \alpha, 1 - \beta\} = \frac{1}{2}(1 - \sum_{i,j=0}^1 \alpha_{ij}).
\]

Based on Theorem 2.2 together with these facts, to show the theorem, it suffices to prove the following assertions:
(i) If $\tau \in TS(A)$ and $(\tau(e), \tau(f)) \neq (\alpha, \beta)$, then
$$\inf_{G} \left[ \limsup_{N \to \infty} \frac{1}{N^2} \log \nu_N(G) \right] = -\infty.$$  

(ii) If $\tau \in TS(A)$ and $\Psi(\tau) = (\alpha, \beta, \tilde{\mu})$, then
$$\inf_{G} \left[ \limsup_{N \to \infty} \frac{1}{N^2} \log \nu_N(G) \right] \leq -\tilde{I}(\tilde{\mu}),$$
$$\inf_{G} \left[ \liminf_{N \to \infty} \frac{1}{N^2} \log \nu_N(G) \right] \geq -\tilde{I}(\tilde{\mu}),$$

where $\tilde{I}(\tilde{\mu})$ is the rate function in Theorem 2.2 and $G$ runs over neighborhoods of $\tau$.

When $(\tau(e), \tau(f)) \neq (\alpha, \beta)$, choose $\varepsilon > 0$ such that $\varepsilon < |\tau(e) - \alpha|$ (or $\varepsilon < |\tau(f) - \beta|$), and set $G := \{\tau' \in TS(A) : |\tau'(e) - \alpha| < \varepsilon$ (or $|\tau'(f) - \beta| < \varepsilon$). Since $\tau_N(e) = \text{tr}_N(P(N)) = k(N)/N \to \alpha$ and $\tau_N(f) = \text{tr}_N(Q(N)) = l(N)/N \to \beta$ as $N \to \infty$, we get $\nu_N(G) = 0$ for large $N$ so that (i) follows.

To prove (ii), assume that $\Psi(\tau) = (\alpha, \beta, \tilde{\mu})$. For any neighborhood $\tilde{G}$ of $\tilde{\mu}$, note that $\Psi^{-1}(0, 1] \times [0, 1] \times \tilde{G}$ is a neighborhood of $\tau$ and

$$\nu_N(\Psi^{-1}(0, 1] \times [0, 1] \times \tilde{G}) = \text{Prob}(\{\Psi(\tau_N) \in [0, 1] \times [0, 1] \times \tilde{G}\}) = \text{Prob}(\{\tilde{R}_N \in \tilde{G}\}) = \tilde{P}_N(\tilde{G}),$$

where $\tilde{R}_N$ is the empirical eigenvalue distribution of $P(N)Q(N)P(N)$ and $\tilde{P}_N$ is its distribution on $\mathcal{M}(0, 1]$ (see the proof of Theorem 2.2). Hence we have

$$\inf_{G} \left[ \limsup_{N \to \infty} \frac{1}{N^2} \log \nu_N(G) \right] \leq \inf_{G} \left[ \limsup_{N \to \infty} \frac{1}{N^2} \log \tilde{P}_N(\tilde{G}) \right] \leq -\tilde{I}(\tilde{\mu})$$

by Theorem 2.2. On the other hand, for any neighborhood $G$ of $\tau$, one can choose $\varepsilon > 0$ and a neighborhood $\tilde{G}$ of $\tilde{\mu}$ such that $\Psi^{-1}((\alpha - \varepsilon, \alpha + \varepsilon) \times (\beta - \varepsilon, \beta + \varepsilon) \times \tilde{G}) \subset G$, which implies that

$$\liminf_{N \to \infty} \frac{1}{N^2} \log \nu_N(G) \geq \liminf_{N \to \infty} \frac{1}{N^2} \log \nu_N\left(\Psi^{-1}((\alpha - \varepsilon, \alpha + \varepsilon) \times (\beta - \varepsilon, \beta + \varepsilon) \times \tilde{G})\right) = \liminf_{N \to \infty} \frac{1}{N^2} \log \text{Prob}(\{|\text{tr}_N(P(N)) - \alpha| < \varepsilon, |\text{tr}_N(Q(N)) - \beta| < \varepsilon, \tilde{R}_N \in \tilde{G}\}).$$

Since $|\text{tr}_N(P(N)) - \alpha| < \varepsilon$ and $|\text{tr}_N(Q(N)) - \beta| < \varepsilon$ for large $N$ (as in the proof of (i)), we have

$$\liminf_{N \to \infty} \frac{1}{N^2} \log \nu_N(G) \geq \liminf_{N \to \infty} \frac{1}{N^2} \log \tilde{P}_N(\tilde{G}) \geq -\tilde{I}(\tilde{\mu})$$

by Theorem 2.2, and hence (ii) is proven. Finally, Theorem 2.2 proves the assertion on the minimizer as well (or this is a direct consequence of the asymptotic freeness of $(P(N), Q(N))$).

For $N \in \mathbb{N}$ and $k \in \{0, 1, \ldots, N\}$ let $\mathcal{P}(N, k)$ denote the set of all $N \times N$ orthogonal projection matrices of rank $k$, and $\gamma_{N,k}$ be the unitarily invariant measure on $\mathcal{P}(N, k)$. We note that $\mathcal{P}(N, k)$ is identified with the homogeneous space $U(N)/(U(k) \oplus U(N-k))$ (or the Grassmannian manifold $G(N,k)$) and $\gamma_{N,k}$ corresponds to the measure on that space induced from the Haar probability measure on the unitary group $U(N)$. In fact, an $N \times N$
unitarily invariant random projection matrix of rank \( k \) we have treated is standardly realized by \( P \in \mathcal{P}(N, k) \) distributed under \( \gamma_{N,k} \).

Let \( (p, q) \) be a pair of projections in a tracial \( W^* \)-probability space \( (\mathcal{M}, \tau) \) and let \( \alpha := \tau(p) \) and \( \beta := \tau(q) \). The free entropy \( \chi(p, q) \) of \( (p, q) \) proposed in [17, 14.2] by Voiculescu is defined as follows: Choose sequences \( k(N) \) and \( l(N) \) such that \( k(N)/N \to \alpha \) and \( l(N)/N \to \beta \) as \( N \to \infty \). For each \( m \in \mathbb{N} \) and \( \varepsilon > 0 \) set

\[
\Gamma(p, q; k(N), l(N); N, m, \varepsilon) := \left\{ (P, Q) \in \mathcal{P}(N, k(N)) \times \mathcal{P}(N, l(N)) : |\text{tr}_N(P_1 \cdots P_m) - \tau(p_1 \cdots p_m)| < \varepsilon \right\}
\]

for all \( (P_j, p_j) \in \{(P, p), (Q, q)\}, 1 \leq j \leq m \}, \)

and define

\[
\chi(p, q) := \lim_{\varepsilon \to 0} \limsup_{N \to \infty} \frac{1}{N^2} \log (\gamma_{N,k(N)} \otimes \gamma_{N,l(N)}) \left( \Gamma(p, q; k(N), l(N); N, m, \varepsilon) \right).
\]

Let \( \mathcal{A} \) be the \( C^* \)-algebra with two projection generators \( e, f \) introduced in the previous section. The free entropy of a tracial state \( \tau \in T \mathcal{S}(\mathcal{A}) \) is defined as \( \chi(\tau(e), \tau(f)) \) in the tracial \( W^* \)-probability space \( (\pi_\tau(\mathcal{A})', \tau) \) obtained via the GNS construction associated with \( \tau \).

Next we identify the rate function in Theorem 3.2 as the free entropy \( \chi(\tau) \) (up to a sign).

**Proposition 3.3.** The rate function in Theorem 3.2 given for \( \alpha = \tau(e) \) and \( \beta = \tau(f) \) is

\[
\mathcal{I}(\tau) = -\chi(\tau).
\]

Moreover \( \limsup \) can be replaced by \( \lim \) in definition (3.5).

**Proof.** Let \( p := \tau(e) \), \( q := \tau(f) \) and \( \tilde{\mu} \) be the distribution of \( efe \) with respect to \( \tau \). In view of the form (3.3) of joint moments of \( e, f \) and the choices of \( k(N), l(N) \) as above, one can easily see that for each \( m \in \mathbb{N} \) and \( \varepsilon > 0 \)

\[
\Gamma(p, q; k(N), l(N); N, 2m, \varepsilon)
\]

whenever \( N \) is large enough. This implies that

\[
\left( \gamma_{N,k(N)} \otimes \gamma_{N,l(N)} \right) \left( \Gamma(p, q; k(N), l(N); N, 2m, \varepsilon) \right) = \tilde{P}_N(G(m, \varepsilon)),
\]

where \( \tilde{P}_N \) is the distribution on \( \mathcal{M}([0, 1]) \) mentioned in the proof of Theorem 3.2 and \( \tilde{G}(m, \varepsilon) \) is a neighborhood of \( \tilde{\mu} \) given by

\[
\tilde{G}(m, \varepsilon) := \left\{ \tilde{\mu}' \in \mathcal{M}([0, 1]) : \left| \int x^k \, d\tilde{\mu}'(x) - \int x^k \, d\tilde{\mu}(x) \right| < \varepsilon, 1 \leq k \leq m \right\}.
\]

Now, as in the proof of [9, 5.6.2] we have the limit

\[
\lim_{N \to \infty} \frac{1}{N^2} \log \left( \gamma_{N,k(N)} \otimes \gamma_{N,l(N)} \right) \left( \Gamma(p, q; k(N), l(N); N, 2m, \varepsilon) \right)
\]

and the conclusion follows from Theorem 3.2 and its proof.
Theorem 3.2 implies that the free entropy \( \chi(p, q) \) of two projections \( p, q \) admits a maximal value, i.e., \( \chi(p, q) = 0 \) if and only if \( p, q \) are free. Moreover, note by Proposition 3.3 that the definition (3.5) of \( \chi(p, q) \) is independent of the choices of sequences \( k(N) \) and \( l(N) \), but this fact is easy to directly verify.

A further study of the free entropy \( \chi(p_1, \ldots, p_n) \) for general \( n \)-tuples of projections as well as some related topics will be in a forthcoming paper [10].

4. Applications of the contraction principle

Let \((P(N), Q(N))\) be as before, and let \( \mathcal{A} \) be the \( C^* \)-algebra of two projection generators introduced in the previous section. Our large deviation in Theorem 3.2 is formulated on the tracial state space of \( \mathcal{A} \). The aim of this section is to exemplify how Theorem 3.2 implies, via the contraction principle, the large deviation for the empirical eigenvalue density of various random matrices made from \((P(N), Q(N))\).

For each selfadjoint element \( h \in \mathcal{A} \) and \( \tau \in TS(\mathcal{A}) \), let \( \lambda_h(\tau) \) denote the distribution measure of \( h \) with respect to \( \tau \). Fixing \( h \) we then have a map \( \lambda_h : TS(\mathcal{A}) \to \mathcal{M}(\mathbb{R}) \); in fact, \( \lambda_h(\tau) \in \mathcal{M}(\{ -\|h\|, \|h\| \}) \) for every \( \tau \in TS(\mathcal{A}) \). It is straightforward to see that \( \lambda_h \) is continuous with respect to \( w^*-\)topology on \( TS(\mathcal{A}) \) and weak topology on \( \mathcal{M}(\mathbb{R}) \). Let \( \tau_N := \tau_{P(N), Q(N)} \) be the random tracial state on \( \mathcal{A} \) induced by \((P(N), Q(N))\) and \( \nu_N \) the distribution on \( TS(\mathcal{A}) \) of \( \tau_N \) (see Section 3). We then notice that
\[
\nu_N \circ \lambda_h^{-1}(\Lambda) = \text{Prob}(\{\tau_N \in \lambda_h^{-1}(\Lambda)\}) = \text{Prob}(\{\lambda_h(\tau_N) \in \Lambda\})
\]
for Borel sets \( \Lambda \subset \mathcal{M}(\mathbb{R}) \). Since
\[
\int x^m \, d\lambda_h(\tau_N)(x) = \tau_N(h^m) = \text{tr}_N(h(P(N), Q(N))^m), \quad m \in \mathbb{N},
\]
it follows that \( \lambda_h(\tau_N) \) is nothing but the empirical eigenvalue distribution of an \( N \times N \) selfadjoint random matrix \( h(P(N), Q(N)) \) (via “noncommutative functional calculus” mentioned in Section 3). Therefore, by the contraction principle (see [6, 4.2.1]), Theorem 3.2 implies the following:

**Theorem 4.1.** For every selfadjoint element \( h \in \mathcal{A} \), the empirical eigenvalue distribution of \( h(P(N), Q(N)) \) satisfies the large deviation principle in the scale \( 1/N^2 \) with the good rate function
\[
I_h(\mu) := \inf \{ I(\tau) : \tau \in TS(\mathcal{A}), \lambda_h(\tau) = \mu \}
\]
for \( \mu \in \mathcal{M}(\mathbb{R}) \), and \( \mu_0 := \lambda_h(\tau_0) \) is a unique minimizer of \( I_h \), where \( I \) and \( \tau_0 \) are as in Theorem 3.2.

**Remark 4.2.** For any unitary \( u \in \mathcal{A} \) define a map \( \lambda_u : TS(\mathcal{A}) \to \mathcal{M}(\mathbb{T}) \), \( \mathbb{T} \) being the unit circle, by letting \( \lambda_u(\tau) \) the distribution of \( u \) with respect to \( \tau \). Then a similar large deviation is satisfied for the empirical eigenvalue distribution of the unitary random matrix \( u(P(N), Q(N)) \) and the rate function \( I_u \) is given in the same way as in Theorem 4.1.

In this way, for concrete applications, it remains only to find an explicit form of the rate function \( I_h \) (or \( I_u \)) as well as that of the minimizer \( \mu_0 \). We present a few examples in the rest of the section.
Example 4.3. Consider $h = ef + fe \in A$ and let $\tau = (\{\alpha_{ij}\}_{i,j=0}^1, \mu) \in TS(A)$ as in Section 3. Since $e(t)f(t) + f(t)e(t)$ has the eigenvalues $t \pm \sqrt{t}$, we get
\[
\tau(\varphi(ef + fe)) = (\alpha_{10} + \alpha_{01} + \alpha_{00})\varphi(0) + \alpha_{11}\varphi(2)
+ \left(1 - \sum_{i,j=0}^{1} \alpha_{ij}\right) \int_0^1 \frac{\varphi(t + \sqrt{t}) + \varphi(t - \sqrt{t})}{2} \, d\mu(t)
\]
for every continuous function $\varphi$ on $\mathbb{R}$. By this expression and (3.4), whenever $\tau$ satisfies (3.2), we have
\[
\lambda_{ef+fe}(\tau) = \max\{|\alpha - \beta|, 1 - 2\alpha, 1 - 2\beta\} \delta_0 + \max\{\alpha + \beta - 1\} \delta_2 + \rho(\mu \circ S^{-1} + \mu \circ T^{-1}),
\]
(4.1)
where $S : (0,1) \rightarrow (0,2)$ and $T : (0,1) \rightarrow [-1/4,0)$ are given by $St := t + t\sqrt{t}$ and $Tt := t - \sqrt{t}$. Hence the empirical eigenvalue distribution of $P(N)Q(N) + Q(N)P(N)$ satisfies the large deviation in the scale $1/N^2$ and the good rate function $\tilde{I}(\tilde{\mu})$ for $\tilde{\mu} \in \mathcal{M}(\mathbb{R})$ is given by (2.9) if $\tilde{\mu}$ is of the form in the right-hand side of (4.1) with $\mu \in \mathcal{M}((0,1))$; otherwise $\tilde{I}(\tilde{\mu}) = +\infty$. The minimizer of $\tilde{I}(\tilde{\mu})$ is the right-hand side of (4.1) with $\mu = \mu_0$, where $\rho\mu_0$ is the continuous part of the measure (2.10).

Example 4.4. Consider $h = ae + bf$ with $a, b \in \mathbb{R}\setminus\{0\}$. Since $ae(t) + bf(t)$ has the eigenvalues $\frac{1}{2}(a + b \pm \sqrt{(a - b)^2 + 4abt})$, we get
\[
\tau(\varphi(ae + bf)) = \alpha_{00}\varphi(0) + \alpha_{10}\varphi(a) + \alpha_{01}\varphi(b) + \alpha_{11}\varphi(a + b)
+ \left(1 - \sum_{i,j=0}^{1} \alpha_{ij}\right) \int_0^1 \frac{1}{2} \left(\varphi\left(\frac{a + b - \sqrt{(a - b)^2 + 4abt}}{2}\right)
+ \varphi\left(\frac{a + b + \sqrt{(a - b)^2 - 4abt}}{2}\right)\right) \, d\mu(t)
\]
for every continuous function $\varphi$ on $\mathbb{R}$ and $\tau = (\{\alpha_{ij}\}_{i,j=0}^1, \mu) \in TS(A)$. Let $A, B$ be the first two smallest numbers of $0, a, b, a + b$, and define $S : (0,1) \rightarrow (A,B)$ and $T : (0,1) \rightarrow (a + b - B, a + b - A)$ by
\[
St := \frac{a + b - \sqrt{(a - b)^2 + 4abt}}{2}, \quad Tt := \frac{a + b + \sqrt{(a - b)^2 + 4abt}}{2}.
\]
When $\tau$ satisfies (3.2), the above expression shows that
\[
\lambda_{ae+bf}(\tau) = \max\{1 - \alpha - \beta, 0\} \delta_0 + \max\{\alpha - \beta, 0\} \delta_a
+ \max\{\beta - a, 0\} \delta_b + \max\{\alpha + \beta - 1, 0\} \delta_{a+b} + \rho(\mu \circ S^{-1} + \mu \circ T^{-1}).
\]
Hence the empirical eigenvalue distribution of $aP(N) + bQ(N)$ satisfies the large deviation and the good rate function as well as its minimizer is determined similarly to the above example.

Let us express the rate function $\tilde{I}(\tilde{\mu})$ and the minimizer $\tilde{\mu}_0$ more explicitly. When $\mu \in \mathcal{M}((0,1))$, the measure $\nu := \frac{1}{2}(\mu \circ S^{-1} + \mu \circ T^{-1})$ is supported in $(A,B) \cup (a + b - B, a + b - A)$ and symmetric at $(a + b)/2$ so that $\mu = 2\nu \circ S_{(A,B)} = 2\nu \circ T_{(a + b - B, a + b - A)}$. Since $St = x$ (or $Tt = x$) implies $t = (x - a)(x - b)/ab$, we get
\[
\int_0^1 \log t \, d\mu(t) = 2 \int_A^B \frac{\log \frac{(x - a)(x - b)}{ab}}{ab} \, d\nu(x) = 2 \int_{a + b - A}^{a + b - B} \frac{\log \frac{(x - a)(x - b)}{ab}}{ab} \, d\nu(x)
\]
so that
\[
\int_0^1 \log t \, d\mu(t) = \int_{(A,B) \cup (a+b-B,a+b-A)} \log \frac{(x-a)(x-b)}{ab} \, d\nu(x).
\]
Similarly,
\[
\int_0^1 \log(1-t) \, d\mu(t) = \int_{(A,B) \cup (a+b-B,a+b-A)} \log \frac{x(a+b-x)}{ab} \, d\nu(x).
\]
On the other hand, we get
\[
\Sigma(\mu) = 4 \int_A^B \int_A^B \log \left| \frac{(x-a)(x-b)}{ab} - \frac{(y-a)(y-b)}{ab} \right| d\nu(x) \, d\nu(y)
\]
\[
= 4 \int_A^B \int_A^B \log \left| \frac{(x-a)(a+b-x-y)}{ab} \right| d\nu(x) \, d\nu(y)
\]
\[
= 2\Sigma(\nu) - \log |ab|.
\]
Consequently, the rate function \( \tilde{I}(\tilde{\mu}) \) is written as
\[
\tilde{I}(\tilde{\mu}) = -2\rho^2 \Sigma(\nu) - \rho |\alpha - \beta| \int_{(A,B) \cup (a+b-B,a+b-A)} \log |x-a)(x-b)| d\nu(x)
\]
\[
- \rho |\alpha + \beta - 1| \int_{(A,B) \cup (a+b-B,a+b-A)} \log |x(a+b-x)| d\nu(x)
\]
\[
+ C + \rho \max\{\alpha, \beta, 1-\alpha, 1-\beta\} \log |ab|
\]
if \( \tilde{\mu} \in \mathcal{M}(\mathbb{R}) \) is of the form
\[
\tilde{\mu} = \max\{1-\alpha - \beta, 0\} \delta_0 + \max\{\alpha - \beta, 0\} \delta_a
\]
\[
+ \max\{\beta - \alpha, 0\} \delta_b + \max\{\alpha + \beta - 1, 0\} \delta_{a+b} + 2\rho\nu
\]
with \( \nu \in \mathcal{M}((A,B) \cup (a+b-B,a+b-A)) \) symmetric at \((a+b)/2\); otherwise \( \tilde{I}(\tilde{\mu}) = +\infty \).

Moreover, by transforming the continuous part of (2.10), the explicit form of the minimizer \( \tilde{\mu}_0 \) can be easily computed as follows:
\[
\tilde{\mu}_0 = \max\{1-\alpha - \beta, 0\} \delta_0 + \max\{\alpha - \beta, 0\} \delta_a
\]
\[
+ \max\{\beta - \alpha, 0\} \delta_b + \max\{\alpha + \beta - 1, 0\} \delta_{a+b}
\]
\[
+ \frac{1}{2} \frac{|x - \frac{a+b}{2}| \sqrt{-(x-A_0)(x-B_0)(x-a-b+B_0)(x-a-b-A_0)}}{\pi(x-a)(x-b)(x-a-b)}
\]
\[
\times 1_{(A_0,B_0)\cup(a+b-B_0,a+b-A_0)}(x) \, dx,
\]
where
\[
A_0 := \frac{a+b - \sqrt{(a-b)^2 + 4ab\xi}}{2}, \quad B_0 := \frac{a+b - \sqrt{(a-b)^2 + 4ab\eta}}{2}
\]
(or exchange \( A_0, B_0 \) depending on the sign of \( ab \)) with \( \xi, \eta \) in (2.11). As is guaranteed by the asymptotic freeness ([15]) of \( (P(N), Q(N)) \), the minimizer \( \tilde{\mu}_0 \) is equal to the distribution of \( ap + bq \) where \( (p, q) \) is a pair of free projections in a tracial \( W^* \)-probability space \( (\mathcal{M}, \tau) \) with \( \tau(p) = \alpha \) and \( \tau(q) = \beta \). In fact, the distribution was computed in [1] by use of \( R \)-transform.

Although one can prove the large deviation result for the empirical eigenvalue density of \( aP(N) + bQ(N) \) (also \( P(N)Q(N) + Q(N)P(N) \)) based on the joint eigenvalue distributions given in Corollary 1.3, our stress is that this is just a particular case of grand Theorem 4.1 (or Theorem 3.2).
Example 4.5. For unitaries we consider a simple example \( u = e^{i\pi}e^{-i\pi f} \). Since the eigenvalues of \( e^{i\pi(t)}e^{-i\pi f(t)} \) are \( 2t - 1 \pm 2i \sqrt{t(1 - t)} = e^{\pm i\theta(t)} \) where \( \theta(t) := \cos^{-1}(2t - 1) \) for \( t \in (0, 1) \), we get

\[
\tau(\varphi(u)) = (\alpha_{11} + \alpha_{00})\varphi(1) + (\alpha_{10} + \alpha_{01})\varphi(-1) + \left(1 - \sum_{i,j=0}^{1} \alpha_{ij}\right) \int_0^1 \frac{\varphi(e^{i\theta(t)}) + \varphi(e^{-i\theta(t)})}{2} d\mu(t)
\]

for every continuous function \( \varphi \) on \( T \) and \( \tau = (\{\alpha_{ij}\}_{i,j=0}^{1}, \mu) \in TS(A) \). When \( \tau \) satisfies (3.2), this implies that

\[
\lambda_n(\tau) = |\alpha + \beta - 1|\delta_1 + |\alpha - \beta|\delta_{-1} + \rho(\mu \circ \theta^{-1} + \mu \circ \tilde{\theta}^{-1}),
\]

where \( \tilde{\theta}(t) := -\theta(t) \) for \( t \in (0, 1) \). For \( \mu \in M((0, 1)) \) let \( \nu := \frac{1}{T}(\mu \circ \theta^{-1} + \mu \circ \tilde{\theta}^{-1}) \), which is a probability measure on \( T \) symmetric for the real axis. We then have

\[
\int_0^1 \log t \, d\mu(t) = \int_T \log \frac{1 + \cos \theta}{2} \, d\nu(e^{i\theta}),
\]

\[
\int_0^1 \log (1 - t) \, d\mu(t) = \int_T \log \frac{1 - \cos \theta}{2} \, d\nu(e^{i\theta}),
\]

\[
\Sigma(\mu) = \int_{\mathbb{T}^2} \log |\cos \theta - \cos \psi| \, d\nu(e^{i\theta}) \, d\nu(e^{i\psi}) - \log 2.
\]

Hence we see by Remark 4.2 that the empirical eigenvalue distribution of \( e^{i\pi P(N)}e^{-i\pi Q(N)} \) satisfies the large deviation in the scale \( 1/N^2 \) and the rate function is given by

\[
\tilde{I}(\tilde{\mu}) = -\rho^2 \int_{\mathbb{T}^2} \log |\cos \theta - \cos \psi| \, d\nu(e^{i\theta}) \, d\nu(e^{i\psi}) - \rho |\alpha - \beta - 1| \int_T \log (1 + \cos \theta) \, d\nu(e^{i\theta}) - \rho |\alpha + \beta - 1| \int_T \log (1 - \cos \theta) \, d\nu(e^{i\theta}) + C + \rho \max\{\alpha, \beta, 1 - \alpha, 1 - \beta\} \log 2
\]

if \( \tilde{\mu} \in M(T) \) is of the form \( \tilde{\mu} = |\alpha + \beta - 1|\delta_1 + |\alpha - \beta|\delta_{-1} + 2\rho \nu \) with \( \nu \in M(T) \) having no atoms at \( \pm 1 \) and symmetric for the real axis; otherwise \( \tilde{I}(\tilde{\nu}) = +\infty \). The minimizer \( \tilde{\mu}_0 \) is also easy to compute as

\[
\tilde{\mu}_0 = |\alpha + \beta - 1|\delta_1 + |\alpha - \beta|\delta_{-1} + \sqrt{-\frac{(\cos \theta + 1 - 2\xi)(\cos \theta + 1 - 2\eta)}{|\sin \theta|}} 1_{(\theta_1, \theta_2) \cup (-\theta_2, -\theta_1)}(\theta) \frac{d\theta}{2\pi},
\]

(4.3)

where \( \theta_1 := \cos^{-1}(2\eta - 1) \) and \( \theta_2 := \cos^{-1}(2\xi - 1) \). This measure is the distribution of \( e^{i\pi p e^{-i\pi q}} \) for free projections \( p, q \) sometimes mentioned above. It may be natural that this distribution is rather different (except the same atomic parts) from that of \( e^{i\pi(p-q)} \) computed from (4.2). In particular, when \( \alpha = \beta = 1/2 \) so that \( \xi = 0 \) and \( \eta = 1 \), the minimizer (4.3) is the uniform measure on \( T \) but (4.2) induces the arcsine law on the angular variable \((-\pi, \pi)\).
REFERENCES

[1] M. Akiyama and H. Yoshida, The orthogonal polynomials for a linear sum of a free family of projections, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 2 (1999), 627–643.
[2] G. Ben Arous and A. Guionnet, Large deviation for Wigner’s law and Voiculescu’s non-commutative entropy, Probab. Theory Related Fields 108, (1997), 517–542.
[3] J. Anderson, B. Blackadar and U. Haagerup, Minimal projections in the reduced group C*-algebra of Z_n * Z_m, J. Operator Theory 26 (1991), 3–23.
[4] B. Collins, Product of random projections, Jacobi ensembles and universality problems arising from free probability, Preprint, arXiv: math.PR/0406560, 2004.
[5] A. G. Constantine, Some non-central distribution problems in multivariate analysis, Ann. Math. Statist. 34 (1963), 1270–1285.
[6] A. Dembo and O. Zeitouni, Large Deviation Techniques and Applications, Second edition, Springer, New York, 1998.
[7] P. J. Forrester, Log-Gases and Random Matrices, Chapter 2, Book manuscript, http://www.ms.unimelb.edu.au/~matpjf/matpjf.html, 2004.
[8] A. Guionnet, Large deviations and stochastic calculus for large random matrices, preprint, 2004.
[9] F. Hiai and D. Petz, The Semicircle Law, Free Random Variables and Entropy, Mathematical Surveys and Monographs, Vol. 77, Amer. Math. Soc., Providence, 2000.
[10] F. Hiai and Y. Ueda, In preparation.
[11] M. L. Mehta, Random Matrices, Second edition, Academic Press, Boston, 1991.
[12] D. Petz and J. Réffy, Large deviation theorem for empirical eigenvalue density of truncated Haar unitary matrices, Prob. Theory Rel. Fields., to appear.
[13] E. B. Saff and V. Totik, Logarithmic Potentials with External Fields, Springer, Berlin-Heidelberg-New York, 1997.
[14] M. Takesaki, Theory of Operator Algebras I, Springer-Verlag, New York-Heidelberg-Berlin, 1979.
[15] D. Voiculescu, Limit laws for random matrices and free products, Invent. Math. 104 (1991), 201–220.
[16] D. Voiculescu, The analogues of entropy and of Fisher’s information measure in free probability theory, I, Comm. Math. Phys. 155 (1993), 71–92.
[17] D. Voiculescu, The analogue of entropy and of Fisher’s information measure in free probability theory VI: Liberation and mutual free information, Adv. Math. 146 (1999), 101–166.
[18] D. V. Voiculescu, K. J. Dykema and A. Nica, Free Random Variables, CRM Monograph Ser., Vol. 1, Amer. Math. Soc., 1992.