The difference between two random mixed quantum states: exact and asymptotic spectral analysis

José Mejia, Camilo Zapata and Alonso Botero

1 Departamento de Física, Universidad de los Andes, Cra 1E No 18A-12, Bogotá, Colombia
2 Department of Physics, ETH Zürich, Otto-Stern-Weg 1, 8093 Zürich, Switzerland

Abstract
We investigate the spectral statistics of the difference of two density matrices, each of which is independently obtained by partially tracing a random bipartite pure quantum state. We first show how a closed-form expression for the exact joint eigenvalue probability density function for arbitrary dimensions can be obtained from the joint probability density function of the diagonal elements of the difference matrix, which is straightforward to compute. Subsequently, we use standard results from free probability theory to derive a relatively simple analytic expression for the asymptotic eigenvalue density (AED) of the difference matrix ensemble, and using Carlson’s theorem, we obtain an expression for its absolute moments. These results allow us to quantify the typical asymptotic distance between the two random mixed states using various distance measures; in particular, we obtain the almost sure asymptotic behavior of the operator norm distance and the trace distance.

Keywords: random matrix, free probability, eigenvalue distribution, random quantum state, asymptotic eigenvalue distribution, Wishart–Laguerre ensemble, free convolution

(Some figures may appear in colour only in the online journal)

1. Introduction

Among the key insights in quantum information that follow from the statistical properties of quantum states in high-dimensional Hilbert spaces is that the overwhelming majority of pure bipartite states are very close to being maximally entangled. More precisely, consider pure,
normalized, quantum states $|\psi\rangle \in \mathcal{H}_M \otimes \mathcal{H}_N$, where $\mathcal{H}_M$ and $\mathcal{H}_N$ are Hilbert spaces of finite dimensions $M$ and $N$ respectively, and for a given $|\psi\rangle$, let $\rho$ be its reduced density matrix in $\mathcal{H}_N$; that is, $\rho = \text{Tr}_{\mathcal{H}_M} |\psi\rangle \langle \psi|$, where $\text{Tr}_{\mathcal{H}_M}$ denotes the partial trace over the Hilbert space $\mathcal{H}_M$. A standard measure of the entanglement of $|\psi\rangle$ is the so-called \textit{entanglement entropy}, defined as the von Neumann entropy $S = -\text{Tr} \log \rho$, so that $S = 0$ when $|\psi\rangle$ is separable and $S = \log(\min(N, M))$ when $|\psi\rangle$ is maximally entangled. A well known result, motivated by an initial estimate of Lubkin [Lub78], later conjectured by Page [Pag93], and finally proved by Foong and Kanno [FK94] (see also [Sen96, SR95]), is that the average entanglement entropy over all pure states $|\psi\rangle$, when $N \leq M$, is given by

$$S = \frac{1}{N - \frac{1}{M}}$$

which for $N \gg 1$ gives $\langle S \rangle \approx \log N$, the maximal entanglement entropy, up to corrections of order $N/M$ (here, and henceforth, random states $|\psi\rangle$ are understood to be uniformly sampled on the hypersphere $\langle \langle \psi | \psi \rangle \rangle = 1$; that is, according to the pushforward of the $U(MN)$ Haar measure under the action of $U(MN)$ on a fixed normalized vector of $\mathcal{H}_M \otimes \mathcal{H}_N$). Since then, a better understanding of the ubiquity of near-maximal entanglement in the bipartite setting [DLMS14, HLW06] has been a subject of considerable interest, fuelled in part by its relevance to several important applications in quantum information (see e.g. [Maj11]), such as random quantum circuits [ODP07, PDP08], superdense coding [AHSW06, HHL04], random quantum channels [BH10, CN11, FK10, Has09, HW08] equilibrium thermodynamics [GLTZ06, PSW06, Rei07] and thermalization [GHT15, GOM01, HW13, LPSW09, LPSW10], among others.

One aspect of the problem that has earned particular attention is the statistics of the spectrum of the partial density matrix $\rho$, the non-zero part of which corresponds to the so-called \textit{Schmidt spectrum}, the squares of the coefficients appearing in the Schmidt (i.e. singular value) decomposition of the state $|\psi\rangle$. When $|\psi\rangle$ is sampled uniformly, as defined previously, the resulting ensemble of random reduced density matrices $\rho$ is an ensemble that is also known in the random matrix theory (RMT) literature as the Fixed Trace Wishart–Laguerre (FTWL) ensemble with Dyson index $\beta = 2$ [For10, Nec07]. The joint probability density function (PDF) for the unordered Schmidt spectrum induced by the uniformly-sampled pure state ensemble was first obtained by Lloyd and Pagels [LP88] and is given by

$$\varrho(\lambda_1, \lambda_2, ..., \lambda_N) \propto \prod_{i=1}^{N} \lambda_i^{M-N} \prod_{i<j}^{N} (\lambda_i - \lambda_j)^2,$$

where it is understood that $N \leq M$, $\sum_{i=1}^{N} \lambda_i = 0$ and $\lambda_i \geq 0$ (henceforth we shall use the symbol $\varrho$ for spectral probability densities and the standard $\rho$ for density matrices). This PDF contains all the necessary information to derive, at least in principle, the statistical properties of any function of the quantum state that is invariant under local unitary transformations. Indeed, a fair amount of progress has been made in obtaining exact results for several quantities of interest, such as moments and correlations of traces of powers of $\rho$ and its entropy [SZ04], the average eigenvalue PDF [KAT08, ATK09], the smallest eigenvalue PDF [AV11, CLZ10, MBL08, Žn07] and the largest eigenvalue PDFs [Viv11]. For general values of $M$ and $N$ these exact formulas tend to be rather complicated and are therefore of limited practical use. However, considerable simplifications emerge in the asymptotic limit $M, N \to \infty$, with the ratio $N/M$ fixed, using well-known asymptotic techniques in RMT. In particular, it follows that the asymptotic eigenvalue density (AED) for the rescaled eigenvalues $x = N\lambda$, satisfies
a Marčenko–Pastur law [MP67, Pag93](see also [NMV11]), as in the unconstrained Wishart ensemble, with parameters given by the ratio of dimensions \( c = N/M \):

\[
\tilde{g}(x) = \max\left(1 - \frac{1}{c}, 0\right) \delta(x) + \frac{1}{2\pi c x} \sqrt{(x-x_-)(x_+-x)} I_{[x_-,x_+]}(x) dx,
\]

where \( x_\pm = c \left(1 \pm \frac{1}{\sqrt{c}}\right) \), and \( I_{[x_-,x_+]}(x) \) is the indicator function in the interval \([x_-,x_+]\). Nadal et al [NMV11] also provides useful asymptotic results for the distribution of the Renyi entanglement entropies (including the von Neumann entropy), and the distribution of the largest Schmidt eigenvalue.

Motivated by the general question of how random bipartite states concentrate around the maximally entangled state, in this paper we address a closely related problem: suppose two random bipartite states \( |\psi_1\rangle \) and \( |\psi_2\rangle \) are uniformly sampled independently from \( \mathcal{H}_M \otimes \mathcal{H}_N \), and their corresponding partial density matrices \( \rho_1 = \text{Tr}_{\mathcal{H}_N}(|\psi_1\rangle\langle\psi_1|) \) and \( \rho_2 = \text{Tr}_{\mathcal{H}_N}(|\psi_2\rangle\langle\psi_2|) \) are computed in \( \mathcal{H}_N \). For appropriately large dimensions, with \( N \leq M \), we expect \( \rho_1 \approx \rho_2 \approx 1/N \). Our question is then: how close are the states \( \rho_1 \) and \( \rho_2 \) from each other? More specifically, we will examine the eigenvalue statistics for the difference matrix \( Z \equiv \rho_1 - \rho_2 \), from which various distance measures can be calculated. In fact, the ensemble of matrices \( Z \) defines a unitarily invariant random matrix ensemble, a fact that automatically implies that \( \rho_1 \) and \( \rho_2 \) cannot be ‘too close’, given the well-known phenomenon of eigenvalue repulsion that characterizes unitarily invariant ensembles, among others. Moreover, unitary invariance allows us to use powerful tools to derive exact, though cumbersome, expressions for the joint PDF of the eigenvalues of the difference matrix \( \lambda \in \mathcal{D} \), as well as a relatively simple formula for the AED and its moments; second, using these results, we derive the almost sure asymptotic values of the trace distance \( d_Q(\rho_1, \rho_2) = \frac{1}{2} \sum_{\lambda \in \text{spec} Z} |\lambda| \) and the operator norm distance \( \|\rho_1 - \rho_2\|_\text{op} = \sup_{\lambda \in \text{spec}(Z)} |\lambda| \), both of which are especially relevant for applications to quantum information theory.

The structure and summary of results of the paper is as follows: In section 2 we present our main results in the form of three new theorems (theorems 2–4) and two corollaries (corollaries 1 and 2), and the remaining sections are devoted to the proofs of these results. The first result is a closed-form formula for the joint eigenvalue PDF for the difference matrix ensemble, which follows from applying a powerful technique of Christandl et al [CDKW14], the so-called derivative principle, which we reproduce here as theorem 1. The derivative principle provides a connection between the joint eigenvalue PDF and the joint PDF of the diagonal matrix elements of a unitarily invariant random matrix ensemble that is then exploited in our first main result, theorem 2, to obtain a closed form expression for the joint eigenvalue PDF in terms of associated Laguerre polynomials, valid for arbitrary dimensions \( N \) and \( M \) with \( N \leq M \). The resulting formula is quite complicated, but may nevertheless prove useful for small \( N \) and \( M \); in particular, the specialization to the case \( N = 2 \) and general \( M \) yields a relatively tractable formula. Section 3 provides a proof of these results, including an alternative proof of the derivative principle that uses standard random matrix theory techniques. The next set of results are concerned with the asymptotics of the eigenvalue PDF of the difference matrix \( Z \). Theorem 3 gives the AED for \( N, M \to \infty \) for all fixed values of the ratio \( c = N/M \) (here, the constraint \( N \leq M \) is relaxed), a result that is proved in section 4 using standard results from free probability theory [VDN92]. Our result agrees with a similar formula independently obtained by Puchala et al [PPZ16], which is however valid for \( c < 2 \) only. In fact, we find that the AED
shows an interesting transition at the critical value $c = 2$. For values of $c < 2$, the AED has positive support in the region $|x| < x_+$, with $x_+^2 = \frac{1}{16}(\sqrt{4c+1+3})^2(\sqrt{4c+1} - 1)$, whereas for values of $c \geq 2$, the AED has positive support in the two regions defined by $x_- < |x| < x_+$, with $x_-^2 = \frac{1}{16}(\sqrt{4c+1-3})^2(\sqrt{4c+1} + 1)$, and a Dirac point measure at the origin. Our other main result, presented in theorem 4, is an expression for the absolute moments (including moments of complex order) of the AED, again for all values of $c$, a result that is proved in section 5 using Carlson’s theorem. The two corollaries to our main results involve the asymptotic almost sure behavior of two distance measures between the independent partial density matrices $\rho_1$ and $\rho_2$, namely the operator norm distance (corollary 1), which is obtained from the upper support point $x_+$ of the AED, and the trace norm distance (corollary 2), which follows from theorem 4 when specialized to the first absolute moment.

2. Main results

Let $\mathcal{H}_M$ and $\mathcal{H}_N$ be two Hilbert spaces of dimensions $M$ and $N$ respectively. Now let $|\psi_1\rangle$ and $|\psi_2\rangle$ be two normalized random pure states in the tensor product Hilbert space $\mathcal{H}_M \otimes \mathcal{H}_N$, which are uniformly and independently sampled, and define $\rho_1$ and $\rho_2$ as the corresponding reduced density matrices for the system $\rho_1 = \text{Tr}_{\mathcal{H}_d}(|\psi_1\rangle\langle\psi_1|)$ and $\rho_2 = \text{Tr}_{\mathcal{H}_d}(|\psi_2\rangle\langle\psi_2|)$. Finally, define the difference matrix $Z$ as

$$Z = \rho_1 - \rho_2.$$  \hfill (4)

For this difference matrix ensemble, our main results are concerned with the exact joint eigenvalue probability PDF $\rho(\lambda)$, where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ are the (unordered) eigenvalues of $Z$, and the AED, the single-eigenvalue marginal of $\rho(\lambda)$ in the asymptotic limit $(N, M) \to \infty$, with fixed ratio $N/M$.

2.1. The exact joint eigenvalue density

We begin with the exact joint eigenvalue PDF for all $N, M$, but with the provision that $N \leq M$. The result is presented in the form of two theorems, of which the first is a recasting of a previous result and the second one is the original result. Both theorems will be proved in section 3.

Theorem 1 establishes an extremely useful connection between the joint PDF of the eigenvalues of a unitarily-invariant random matrix ensemble and the joint PDF of the matrix diagonal elements of the same ensemble, which in general is considerably simpler to compute. This result is known in more general form from the theory of Duistermaat–Heckman (DH) measures [Gui94, Hec82] which are measures on the dual Lie algebra of a Lie group with a Hamiltonian action on a symplectic manifold, obtained from the push-forward of the Liouville measure along the corresponding moment map. It concerns the connection between the so-called non-abelian DH-measure for the Hamiltonian action of a compact connected Lie group and the abelian DH-measure for its maximal torus [GP90, Hec82]. Recently, Christandl et al [CDKW14] have used this connection, under the name of the derivative principle, for the computation of the eigenvalue PDFs of reduced density matrices of multipartite-entangled states. As this derivative principle is surprisingly not as well known as it should be in the context of random matrices, we recast it here and prove it in the following section using the standard language of random matrix theory.
Theorem 1 (Derivative principle for unitarily invariant random matrix ensembles [CDKW14]). Let $Z$ be a random matrix drawn from a unitarily invariant random matrix ensemble, $q_Z$ the joint eigenvalue PDF for $Z$ and $\Psi_Z$ the joint PDF of the diagonal elements of $Z$. Then

$$q_Z(\lambda) = \left( \prod_{p=1}^{N} p! \right)^{-1} \Delta(\lambda) \Delta(-\partial_\lambda) \Psi_Z(\lambda),$$

where $\Delta(\lambda) = \prod_{i<j} (\lambda_j - \lambda_i)$ is the Vandermonde determinant and $\Delta(-\partial_\lambda)$ the differential operator $\prod_{i<j} \left( \frac{\partial}{\partial \lambda_i} - \frac{\partial}{\partial \lambda_j} \right)$.

This theorem proves to be particularly useful to compute the joint eigenvalue PDF of the sum of independent random matrices drawn from unitarily invariant ensembles, given that the joint PDF of the diagonal elements is the convolution of the respective joint PDFs. This is precisely the case at hand for the ensemble of random matrices $Z = \rho_1 - \rho_2$, where $\rho_1$ and $\rho_2$ are the partial density matrices of two independent random bipartite states.

Theorem 2 gives a contour integral expression (or what equivalently can be cast as a constant term identity) for the joint PDF of the diagonal elements for our difference ensemble (an alternative expression in terms of Lauricella generalized hypergeometric functions is given at the end of section 3.2):

**Theorem 2.** Let $Z = \rho_1 - \rho_2$ be the ensemble of difference matrices as defined at the beginning of the section. Then, the joint PDF of diagonal elements $\vec{z} = (z_1, z_2, \ldots, z_N) \equiv (Z_{11}, Z_{22}, \ldots, Z_{NN})$ of $Z$ is given by the contour integral

$$\Psi_Z(\vec{z}) = \delta_0(\vec{z}) \Gamma(NM)^2 \frac{(-1)^{N(M-1)}}{2\pi i} \oint_{C} \prod_{i=1}^{N} \psi_{\vec{z}}(s|z_i)),$$

where the contour encircles the origin, $L_{\alpha}^{\beta}(x)$ are the associated Laguerre Polynomials, $\delta_0(\vec{z})$ the surface delta function

$$\delta_0(\vec{z}) = \delta(\sum_{i=1}^{N} z_i),$$

and $I_0(\vec{z})$ is the indicator function on the $N-1$ dimensional region $R$ in $\mathbb{R}^N$ defined by the Minkowski difference set $R = S - S$, where $S$ is the standard probability simplex.

Theorems 1 and 2 provide a systematic, though not necessarily practical, way of computing the joint eigenvalue PDF for the difference of two random partial density matrices, via equations (5) and (6). In figure 1 we show the resulting PDF for the case $N = 3$ and $M = 3$. The PDFs are supported within a convex polytope in the hyperplane of $\mathbb{R}^N$, with vertices at $\vec{e}_i - \vec{e}_j$, $i \neq j$, where the $\vec{e}_i$ are the standard unit vectors in $\mathbb{R}^N$. They are symmetric under reflection ($\lambda \rightarrow -\lambda$) and permutation of the eigenvalues. The multi-lobe shape of the PDF (see figure 1) is a signature of the well-known eigenvalue repulsion phenomenon, and is a consequence of the Vandermonde determinant in (5), which forces the PDF to vanish whenever two of the eigenvalues are the same.

Armed with the joint eigenvalue PDF, the average density of eigenvalues can be obtained by integrating out over $N-1$ eigenvalues. Unfortunately, there does not appear to be any particularly simple expressions for these marginal PDFs, except in the case $N = 2$, where the
exact result for the average eigenvalue PDF for any $M$ can be expressed in terms of a Gauss hypergeometric function $\text{F}_2^1$ (see appendix B) as:

$$\rho(\lambda) = K_M \lambda \frac{\partial}{\partial \lambda} \left[ \frac{(\lambda^2 - 1)^{2M-1}}{1 + |\lambda|} \text{F}_2^1 \left( \frac{1}{2}, 1 - M; \frac{4|\lambda|}{(1 + |\lambda|)^2} \right) \right],$$

where the proportionality constant is

$$K_M = \frac{\Gamma(2M)^2 \Gamma(2M - 1)^2}{\Gamma(4M - 2) \Gamma(M)^4}.$$

In figures 2(a) and (b) we show how our results for $N = 2$, $M = 10$ and $N = 3$, $M = 3$ agree with empirical distributions obtained from independent samplings of $Z$. From the empirical distributions shown in figures 2(c) and (d) for higher values of $N$ (with $M \geq N$), it becomes evident that the average eigenvalue density will show exactly $N$ peaks, which as $N$ grows become progressively less pronounced, tending to a smooth single-peaked, finitely-supported, distribution in the asymptotic limit $N \to \infty$.

2.2. The asymptotic eigenvalue density (AED)

Considerable simplifications ensue in the asymptotic limit $(M, N) \to \infty$, with a fixed ratio $c = N/M$, if we concentrate on the single-eigenvalue marginal of the joint eigenvalue PDF, to which the empirical eigenvalue density almost surely converges asymptotically. As we show in section 4, unitary invariance of the independent matrices $\rho_1$ and $\rho_2$ implies that two ensembles satisfy the so-called freeness condition asymptotically [VDN92], from which it follows that the AED of $Z = \rho_1 - \rho_2$ can be obtained from the free convolution of the AEDs of $\rho_1$ and $-\rho_2$, which are given for all values of $c$ in terms of the Marčenko–Pastur law (3). Our main result involves the asymptotic density function of the rescaled eigenvalues $x = N\lambda$, which
we denote by $\tilde{\rho}(x)$. The resulting AED coincides for $c < 2$ with the expression obtained in [PPZ16], except for the fact that our definition of $c$ is the inverse of the one used in that reference; however, our result also covers the regime $c > 2$, for which the support of the AED is qualitatively different.

Theorem 3. Let $Z = \rho_1 - \rho_2$ be the ensemble of difference matrices as defined at the beginning of the section, and define the constants

$$x_\pm = \frac{1}{\pi} \left( \sqrt{4c + 1} \pm 3 \right)^{1/2} \left( \sqrt{4c + 1} \pm 1 \right)^{1/2},$$

(10)

where $x_+$ is real and positive and $x_-$ purely imaginary for $c < 2$ and real and positive for $c > 2$. Then, in the limit $N \to \infty$, $M \to \infty$ with the ratio $c = N/M$ fixed, the asymptotic rescaled eigenvalue density is, almost surely,

$$\tilde{\rho}(x) = \begin{cases} \sqrt{(2 - c)^2 + 3x^2} \sinh \left( \frac{f(x)}{3} \right) & |x| < x_+, \\ \frac{\sqrt{3} \pi c |x|}{3} & |x| \geq x_+ \end{cases},$$

(11)

for $c \leq 2$, and

Figure 2. Average eigenvalue density of $\rho_1 - \rho_2$ for low dimensions (red line). The blue bars represent the normalized histograms obtained numerically from 30,000 random samples for figures 2(a) and (b). For figures 2(c) and (d) the number of samples was 5,000.
\[
\tilde{\mathcal{G}}(x) = \begin{cases} 
\sqrt{(2-c)^2 + 3x^2} \sinh \left( \frac{f(x)}{3} \right) & |x| \in (x_-, x_+) \\
\frac{c - 2}{c} \delta(x) & |x| \notin (x_-, x_+) 
\end{cases},
\]

for \( c > 2 \), where

\[
l(x) = \log \left( \eta(x) + \sqrt{\eta^2(x) - 1} \right),
\]

and

\[
\eta(x) = \frac{9(c + 1)x^2 + (2 - c)^3}{(2 - c)^2 + 3x^2}^{1/2}.
\]

Note that in expressions (11) and (12) we can alternatively write for the regions of positive support

\[
\frac{\sqrt{(2-c)^2 + 3x^2}}{\sqrt{3 \pi c |x|}} \sinh \left( \frac{f(x)}{3} \right) = \frac{1}{2\pi c |x|} \left( w[x, c] - \frac{x^2 + \frac{1}{3}(2-c)^2}{w[x, c]} \right),
\]

where

\[
w(x) = \left( \sqrt{(x^2(x^2 - x^2)(x_+^2 - x^2))} + \sqrt{3}(c + 1) \left( x^2 + \frac{(2-c)^3}{9(c+1)} \right) \right)^{1/3}.
\]

Figure 3 shows how the AED depends on the parameter \( c \). The most striking feature of this behavior is the fact that for \( c > 2 \) a gap in the eigenvalue density arises for \( |x| < x_- \), with a point distribution appearing at \( x = 0 \) with weight \( \frac{c - 2}{c} \). The appearance of the point distribution for \( c > 2 \) can be understood from the fact that if \( 2M < N \), the difference \( \rho_1 - \rho_2 \) is of rank at most \( 2M \), and generically we expect the ranges of \( \rho_1 \) and \( \rho_2 \) to be linearly independent when \( 2M < N \), in which case the fraction of zero eigenvalues of \( \rho_1 - \rho_2 \) should be \( \frac{N - 2M}{N} = 1 - \frac{2}{c} \).

Also, from the asymptotic behavior of random subspaces (see Collins [Col05]), we expect that when \( M/N \to 0 \) (or \( c \to \infty \)), the ranges of \( \rho_1 \) and \( \rho_2 \) should become asymptotically orthogonal and thus the non-zero eigenvalues of \( \rho_1 - \rho_2 \) to be approximately the union of the non-zero eigenvalues of \( \rho_1 \) and \( -\rho_2 \). The AED of \( \rho_1 - \rho_2 \) should then be the mixture of the AEDs of \( \rho_1 \) and \( -\rho_2 \), which follow the Marchenko–Pastur law (or a reflected version of it). Thus, the existence of a gap for \( c > 2 \) reflects the gap that already exists in the Marchenko–Pastur law in a region in the neighborhood of the origin, and the closing of the gap for \( c < 2 \) may be understood as a consequence of strong mixing due to the fact that when \( 2M > N \) the ranges of \( \rho_1 \) and \( \rho_2 \) cannot be linearly independent subspaces.

An obvious variant of our problem is to consider differences of the density matrices with different weights, i.e. \( Z = \rho_1 - \rho_2 \). The extension of the results of theorem 3 for this more general class of ensembles can also be obtained in closed form, as is shown in appendix A.

As pointed out in [PPZ16], the AED for the case \( N = M \) has been obtained in relation to the commutator of free random variables [DN12, NS98]. Similarly, these results can be connected to the AED of the partial transpose \( \mathcal{W}^{PT} \) of a random rescaled Wishart matrix [Aub12, BN13].
In the latter reference, the authors notice that the AED can be expressed as a free difference of free Poisson distributions with different parameters. Thus, we can see the AED from theorem 3, as a special limiting case where the two free Poisson distributions have the same parameter, a particular situation considered in theorem 5.3 of [BN13] where the resulting AED is said to be a modified free Bessel distribution [BBCC11].

2.3. Absolute moments of the AED

Our final theorem concerns the absolute moments of the AED $\tilde{\varphi}(x)$ (equations (11) and (12)). As shown in section 5, the general theorem for the absolute moments follows from Carlson’s theorem, which warrants an analytic extension of the closed-form expression for the even integer moments that is obtained from the Laurent series of the Cauchy transform of $\tilde{\varphi}(x)$. Our main result gives the absolute complex moments (or equivalently the Mellin transform for the density of $|x|$):

**Theorem 4.** Let $\tilde{\varphi}(x)$ be the AED in theorem 3 and $z \in \mathbb{C}$ with $\text{Re}(z) > 0$. Then the complex absolute moments of $\tilde{\varphi}(x)$,

$$m_z = \int_{-\infty}^{\infty} dx |x|^z \tilde{\varphi}(x),$$

are

![Figure 3](image-url)
\[ m_c(c) = \begin{cases} \frac{\Gamma(z + 1)(2c)^{\frac{z}{2}}}{\Gamma\left(\frac{z}{2} + 1\right)^2} & c \leq 2, \\ 2c^{z-1} \frac{\Gamma\left(\frac{z}{2} + 1\right)^2}{\Gamma\left(\frac{z}{2} + 1\right)^2} & c > 2, \end{cases} \quad \text{where } F \text{ is the Gauss hypergeometric function.} \]

\[ 2.4. \text{ Corollaries: asymptotics of the distance measures} \]

The main application of our results is in quantifying the distance between the two random states \( \rho_1 \) and \( \rho_2 \) using distance measures derived from the spectrum of \( Z = \rho_1 - \rho_2 \). In particular, the almost sure behavior of two distance measures, can be readily obtained as corollaries of theorems 3 and 4.

From the almost sure convergence of the empirical eigenvalue distributions to the AED \([\text{Spe13}]\), the maximum absolute value of the eigenvalues of \( Z \) may be expected to converge to the value \( x_+ / N \), where \( x_+ \) is the upper edge of the AED. As shown in section 4.3, this is indeed the case:

**Corollary 1.** Under the conditions of theorem 3, the operator norm of the difference \( \rho_1 - \rho_2 \) almost surely behaves as

\[ \| \rho_1 - \rho_2 \|_{\text{op}} \rightarrow \frac{\text{const}}{4N} \left( \sqrt{1 + 4c} + 3 \right)^{\frac{3}{2}} \left( \sqrt{1 + 4c} - 1 \right)^{\frac{1}{2}}. \]  

(19)

Note that for small \( c \), the limiting value simplifies to \( 2\sqrt{2c} / N \). This may be compared to the leading behavior of \( \| \rho_1 - \|N\| \|_{\text{op}} \), the operator norm of the difference between (say) the random state \( \rho_1 \) and its average, the totally mixed state \( \frac{1}{N} \). Using the upper support point \( x_+ \) of the Marčenko–Pastur law (3), this leading behavior can be shown to be \( \approx 2\sqrt{c} / N = 2\sqrt{NM} \), so that asymptotically \( \| \rho_1 - \rho_2 \|_{\text{op}} \approx \sqrt{2} \| \rho_1 - \|N\| \|_{\text{op}} \) for \( c \ll 1 \). Similarly, for \( c \gg 1 \), we find that both \( \| \rho_1 - \rho_2 \|_{\text{op}} \) and \( \| \rho_1 - \|N\| \|_{\text{op}} \) behave as \( \approx cN \).

Next, we turn to the trace distance, which is defined as

\[ d_t(\rho_1, \rho_2) = \frac{1}{2} \| \rho_1 - \rho_2 \| = \frac{1}{2} \text{Tr} \left( \sqrt{(\rho_1 - \rho_2)(\rho_1 - \rho_2)^\dagger} \right), \]  

(20)

where \( \| \cdot \| \) is the trace norm. A straightforward consequence of the almost sure convergence of the AED (theorem 3) is therefore that in the limit \( N, M \rightarrow \infty \) (with \( c \) constant), the trace distance tends to a limiting value

\[ d_t(\rho_1, \rho_2) \xrightarrow{a.s.} \frac{1}{2} \int_{-\infty}^{\infty} dx \left| \tilde{g}(x) \right|. \]  

(21)

Thus, the limiting trace distance follows as a corollary of theorem 4 for \( z = 1 \), together with hypergeometric identities (see section 5.3):
Corollary 2. Under the conditions of theorem 3, the trace distance between $\rho_1$ and $\rho_2$ almost surely behaves as

$$d_{tr}(\rho_1, \rho_2) \to \begin{cases} \frac{1}{2\pi c} \left( (c + 1)\sqrt{(2 - c)c} + (4c - 2)\arcsin\left(\frac{\sqrt{c}}{2}\right) \right), & c \leq 2, \\ 1 - \frac{1}{2c}, & c > 2. \end{cases}$$

As shown in figure 4, the trace distance between two random states asymptotically tends to the maximum value of 1 as $c \to \infty$, which can be understood as resulting from $2M$ non-zero eigenvalues of $Z$ each of which is of magnitude $\geq 1/M$. Similarly, as $c \to 0$, the trace distance goes to zero like $\sim \frac{4\sqrt{2\pi c}}{\sqrt{3}}$. Comparing this leading behavior with that of the distance to the maximally mixed state, $d_{tr}(\rho_1, \mathbb{I}/N)$, which can be obtained from the first moment of the Marčenko–Pastur law (3), we find that $d_{tr}(\rho_1, \mathbb{I}/N) \approx \frac{4\sqrt{c}}{\sqrt{3}} \approx 0.42/\sqrt{\pi M}$, so that asymptotically $d_{tr}(\rho_1, \rho_2) = \sqrt{2} d_{tr}(\rho_1, \mathbb{I}/N)$ for $c \ll 1$, which is the same relation obeyed by the operator norm distance. Finally, we note that our expression coincides with the value $d_{tr} = \frac{1}{4} + \frac{1}{\pi} \approx 0.5683$ obtained in [PPZ16] for the case $c = 1$.

3. The exact joint eigenvalue density

We then proceed with the proofs of theorems 1 and 2. As mentioned earlier, despite the fact that theorem 1 is a known result, it is not widely known within the RMT community. We have therefore chosen to prove it here employing random matrix theory techniques.

3.1. Proof of theorem 1 (derivative principle)

Consider an ensemble of random $N \times N$ Hermitian matrices $Z$ that is known to have a unitarily-invariant PDF $P(Z)$; that is,
for any unitary matrix $U$, where $dZ$ is the standard volume element for Hermitian matrices

$$dZ = \prod_{i<j} dz_{ij} \prod_{i} dz'_{ii},$$

and $Z_{ij}^R$ and $Z_{ij}^I$ are the real and imaginary parts of $Z_{ij}$ (with independent variables $Z_{ij}^R$ for $i \leq j$ and $Z_{ij}^I$ for $i < j$). Now let $\Phi_Z(K)$ be the matrix Fourier transform of $P(Z)$,

$$\Phi_Z(K) = \int e^{-i \text{Tr}(KZ)} P(Z) dZ$$

where henceforth we shall assume that $K$ is also a Hermitian $N \times N$ matrix. From the unitary invariance of $P(Z)$, it follows that $\Phi_Z(K)$ is a symmetric function of the eigenvalues $\kappa_i$ of $K$; hence, we shall write the characteristic function as $\Phi_Z(\kappa)$. We then define $\Psi_Z(z)$ as the ordinary Fourier transform of $\Phi_Z(\kappa)$:

$$\Psi_Z(z) = \frac{1}{(2\pi)^N} \int e^{iz^T \kappa} \Phi_Z(\kappa) d\kappa.$$  

Let us now show that $\Psi_Z(z)$ is in fact the PDF for the diagonal elements $z_i \equiv Z_{ii}$ of $Z$. For this, note the identity

$$\delta(Z) = \frac{1}{2^{N/2}}} \int e^{i \text{Tr}(KZ)} dK$$

where $\delta(Z) = \prod_{i<j} \delta(Z_{ij}^R) \prod_{i} \delta(Z_{ii}^I)$ and the matrix integral will be understood to be over the subspace of Hermitian matrices with respect to the volume element (24). The PDF $P(Z)$ can then be expressed as the integral

$$P(Z) = \frac{1}{2^{N/2}}} \int e^{i \text{Tr}(KZ)} \Phi_Z(K) dK.$$  

Now break up $Z$ as $Z_{||} + Z_{\perp}$, where $Z_{||} = \text{diag}(z_1, z_2, \ldots z_N)$ and $Z_{\perp}$ the off-diagonal part of $Z$. The marginal PDF for the diagonal elements is then given by

$$P(z) = \int dZ_{||} P(Z) = \frac{1}{2^{N/2}}} \int dZ_{||} e^{i \sum_{i=1}^N \kappa_i K_{ii} + i \text{Tr}(KZ_{||})} \Phi_Z(K) dK,$$

where $dZ_{||} = \prod_{i} dZ_{ii}^R \prod_{i<j} dZ_{ij}^I$. Exchanging the order of integration, we use the fact that

$$\int e^{i \text{Tr}(KZ_{||})} dZ_{||} = \pi^{N(N-1)/2} \delta(K_{||}),$$

where $K_{||}$ is the off-diagonal part of the matrix $K$. Hence,

$$P(z) = \frac{1}{(2\pi)^N} \int e^{i \sum_{i=1}^N \kappa_i \Phi_Z(\kappa) = \text{spec}(K_{||})} \prod_i dK_{ii},$$

where $K_{ii} = \text{diag}(K_{i1}, K_{i2}, \ldots K_{iN})$. However, if $K$ is diagonal, the $K_{ii}$ are also its eigenvalues. Hence $P(z)$ is the Fourier transform of $\Phi_Z(\kappa)$ and is therefore equal to $\Psi_Z(z)$. 

...
Now, since $Z$ is Hermitian, we can parametrize it as $Z = U \Lambda U^\dagger$, where $U$ is unitary and $\Lambda$ is the diagonal matrix of the eigenvalues $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ of $Z$. A standard result from random matrix theory [Meh04] states the volume element (24) can be written as
\[
dZ = \frac{\pi^{N(N-1)/2}}{\prod_{p=1}^N p!} \Delta^2(\lambda) \left( \prod_{i=1}^N d\lambda_i \right) DU,
\]  
where $DU$ is shorthand for the normalized Haar measure on the group of $N \times N$ unitary matrices. From this it follows that if $P(Z)$ is unitarily invariant, the eigenvalue PDF can be written as
\[
g(\lambda) = \frac{\pi^{N(N-1)/2}}{\prod_{p=1}^N p!} \Delta^2(\lambda) P(\Lambda).
\]  
Using equations (31) and (33), and similarly applying (32) to the measure $dK$, we can then express $g(\lambda)$ as
\[
g(\lambda) = \frac{\Delta^2(\lambda)}{(2\pi)^N \prod_{p=1}^N p!} \int \Delta^2(\kappa) (e^{i \text{Tr} [U \Lambda U^\dagger]} \Phi_Z(\kappa)) d^N \kappa,
\]  
where $\Lambda_K$ is the diagonal matrix of the eigenvalues of $K$ and
\[
\langle e^{i \text{Tr} [U \Lambda U^\dagger]} \rangle_U = \int e^{i \text{Tr} [U \Lambda U^\dagger]} DU.
\]  
This is the well-known Harish–Chandra–Itzykson–Zuber integral [IZ80]:
\[
\langle e^{i \text{Tr} [U \Lambda U^\dagger]} \rangle_U = i^{-N(N-1)/2} \left( \prod_{p=1}^{N-1} p! \right) \Delta(\kappa) \frac{\det(\exp[i \kappa \lambda_{ij}])_{i,j=1}^{N \times N}}{\Delta(\lambda) \Delta(\bar{\lambda})}.
\]  
Inside the integral (34), we can use permutation symmetry of the integrand to replace $\det(\exp[i \kappa \lambda_{ij}])$ by $N! \exp[i \kappa \cdot \bar{\lambda}]$, to obtain
\[
g(\lambda) = \frac{1^{-N(N-1)/2}}{(2\pi)^N \prod_{p=1}^N p!} \Delta(\lambda) \int \Delta(\kappa) e^{i \kappa \cdot \bar{\lambda}} \Phi_Z(\kappa) d^N \kappa.
\]  
Finally, replacing the arguments $\kappa_i$ in $\Delta(\kappa)$ by the partial derivative operators $-i \frac{\partial}{\partial \lambda_i}$ acting on $e^{i \kappa \cdot \bar{\lambda}}$, we obtain relation (5) between $g(\lambda)$ and $\Psi_Z(\bar{\lambda})$, since $\Psi$ is the Fourier transform of $\Phi$.

3.2. Proof of theorem 2: computation of $\Psi_Z(\bar{\lambda})$

Next we turn to the derivation of (6) in theorem 2. First, suppose $Z = X + Y$, where $X$ and $Y$ are independently drawn from unitarily invariant ensembles. Then, the PDF of the diagonal elements of $Z$ is simply the convolution of $\Psi_X$ and $\Psi_Y$. Specializing this result to the problem of interest, we take $X = \rho_1$ and $Y = -\rho_2$, where $\rho_1$ and $\rho_2$ are sampled independently from the FTWL ensemble, with eigenvalue PDF [SZ04]
\[ \varrho(\bar{\lambda}) = \frac{\Gamma(MN)}{\prod_{j=0}^{N-1} \Gamma(M-j) \Gamma(N-j+1)} \delta_\lambda(\bar{\lambda}) \prod_{i=1}^{N} \lambda_i^{M-N} \prod_{i<j} \frac{(\lambda_i - \lambda_j)^2}{\Gamma} , \quad (38) \]

where \( \delta_\lambda(\bar{\lambda}) = \delta\left(\sum_{i=1}^{N} \lambda_i - 1\right) \delta_\lambda(\bar{\lambda}) \) and \( \delta_\lambda(\bar{\lambda}) \) is the indicator function on the probability simplex \( S \),

\[ S = \left\{ (\lambda_1, \lambda_2, \ldots, \lambda_N) \bigg| \lambda_i \geq 0, \sum_{i=1}^{N} \lambda_i = 1 \right\} . \quad (39) \]

In the FTWL ensemble, a random reduced density matrix \( X = \rho_1 \) can be expressed as

\[ X = \frac{GG^\dagger}{\text{Tr}(GG^\dagger)} , \quad (40) \]

where \( G \) is an \( N \times M \) matrix with independent Gaussian complex entries. It is then straightforward to show that the PDF of the diagonal elements is the symmetric Dirichlet distribution

\[ \Psi_\lambda(\bar{x}) = \frac{\Gamma(MN)}{\Gamma(M)^N} \delta_\lambda(\bar{x}) \prod_{i=1}^{N} x_i^{M-1} . \quad (41) \]

Next, recalling that \( Y = -\rho_2 \), we have \( \Psi_\lambda(\bar{x}) = \Psi_\lambda(-\bar{x}) \); hence,

\[ \Psi_\mu(\bar{\varepsilon}) = \int_{\mathbb{R}^N} d^N x \Psi_\lambda(x) \Psi_\lambda(\bar{x} - \bar{\varepsilon}) . \quad (42) \]

The region of integration, as well as the region of support of \( \Psi_\mu \), is set by the support conditions of the \( \Psi_\lambda \) in the integrand. The integration region for integral \( (42) \) is the one that lies in the intersection between the probability simplex \( S \) and the shifted simplex

\[ S' = \left\{ (x_1 + z_1, x_2 + z_2, \ldots, x_N + z_N) \bigg| x_i \geq 0, \sum_{i=1}^{N} x_i = 1 \right\} . \quad (43) \]

We will call this region \( T = S \cap S' \). Since the diagonal elements of a density matrix lie in the probability simplex, the diagonal elements of the difference between two random density matrices must lie in the Minkowski difference set \( R = S - S \), or explicitly,

\[ R = \{ \bar{r} - \bar{r}' \big| \bar{r}, \bar{r}' \in S \} , \quad (44) \]

which is the \((N - 1)\)-dimensional convex polytope on the hyperplane \( \{ (x_1, \ldots, x_N) \in \mathbb{R}^N \big| \sum_{i=0}^{N} x_i = 0 \} \) with vertices at the points \( \bar{e}_i - \bar{e}_j \) for \( i \neq j \), where \( \bar{e}_i \) are the standard unit vectors in \( \mathbb{R}^N \). The condition \( \bar{\varepsilon} \in R \) is precisely the condition such that \( S \cap S' \neq \emptyset \) and hence the support condition for \( \Psi_\mu \).

To further characterize the integration region \( T \), let \( P_k \) be the facet of \( S \) with vertices \( \{ \bar{e}_i \big| i \neq k \} \) and likewise \( P'_k \) be the facet of \( S' \) with vertices \( \{ \bar{e}_i + \bar{\varepsilon} \big| i \neq k \} \), with \( \bar{\varepsilon} \in R \). If \( \bar{r} \) is a point in \( T \) then the distances \( d_k(\bar{r}) \) and \( d'_k(\bar{r}) \) of \( r \) to \( P_k \) and \( P'_k \) respectively, are given by

\[ d_k(\bar{r}) = \sqrt{\frac{N}{N - 1}} x_k , \quad d'_k(\bar{r}) = \sqrt{\frac{N}{N - 1}} (x_k - z_k) , \quad (45) \]
where \((x_k)_{k=1}^N\) are the coordinates of \(\vec{r}\) (with \(\sum_{i=1}^N x_i = 1\)). This implies that if \(z_k > 0\), then the facet \(P_k\) is closer to \(\vec{r}\) than the facet \(P_k\), and conversely, if \(z_k < 0\) then the facet \(P_k\) is closer to \(\vec{r}\) than the facet \(P_k\). Hence, \(T\) is the convex polytope bounded by the \(N\) facets \(F_k\) where

\[
F_k = \begin{cases} 
  P_k, & z_k \leq 0 \\
  P_k', & z_k > 0 .
\end{cases}
\]  

(46)

Let us then parameterize points in a facet \(F_k\) by

\[
\vec{r} = \sum_{i=k}^N \hat{\theta}_i \hat{e}_i + s_k \vec{z} = \sum_{i=k}^N \hat{\theta}_i' \hat{e}_i + s_k \vec{z} .
\]  

(47)

where \(s_k = 1\) if \(z_k > 0\) and \(s_k = 0\) if \(z_k \leq 0\). When two facets \(F_k\) and \(F_k'\) meet, the intersection is described by the condition

\[
\sum_{i=k}^N \hat{\theta}_i \hat{e}_i + s_k \vec{z} = \sum_{i=k'}^N \hat{\theta}_i' \hat{e}_i + s_k \vec{z} .
\]  

(48)

Using these constraints and solving for the coefficients \(\hat{\theta}_i\), the vertices \(\{\vec{r}_j\}\) of the integration region can be shown to be given by

\[
\vec{r}_j = \vec{r} + \gamma \vec{e}_j ,
\]  

(49)

where

\[
\vec{r} = \sum_{j:z_j > 0} z_j \vec{e}_j , \quad \gamma = 1 - \sum_{j:z_j > 0} z_j .
\]  

(50)

Thus, the integration region \(T\) is also a regular simplex, obtained from the standard simplex \(S\) by a shift \(\vec{r}\) and uniform rescaling by the factor \(\gamma\). Note that the support condition \(\vec{z} \in R\) implies that \(\sum z_j = 0\), and hence that on \(R\), \(\gamma\) can be written as a symmetric function of the \(z_i\)’s, namely,

\[
\gamma = 1 - \frac{1}{2} \sum_{j=1}^N |z_j| .
\]  

(51)

Introducing the change of variables \(x'_i = \frac{1}{\gamma} (x_i - s_i x_k)\) to undo these transformations, and combining equations (41) and (42), we arrive at an expression for \(\Psi\) in terms of an integral on the standard probability simplex \(S\):

\[
\Psi_{\vec{z}}(\vec{z}) = \delta_\vec{z}(\vec{z}) \frac{\Gamma((MN)^2)}{\Gamma(M)^{2N}} \gamma^{N(2M-1)-1} \times \int_{\mathbb{R}^N} d^{N}\vec{x} \quad \delta_\vec{z}(\vec{x}) \prod_i x_i^{M-1} \left( x_i + \frac{|z_i|}{\gamma} \right)^{M-1} .
\]  

(52)

Where \(\delta_\vec{z}(\vec{z})\) is defined as in equation (7). Expanding the terms \(\left( x_i + \frac{|z_i|}{\gamma} \right)^{M-1}\) in the integrand, we can use the multinomial Beta function

\[
\frac{\prod_{i=1}^N \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^N \alpha_i)} = \int_{\mathbb{R}^N} d^{N}\vec{x} \quad \delta_\vec{z}(\vec{x}) \prod_{i=1}^N x_i^{\alpha_i-1} ,
\]  

(53)
to arrive at the expression
\[
\Psi_2(\zeta) = \varepsilon(\zeta) \frac{\Gamma(MN)^2}{\Gamma(M)} \times \sum_{k_0, \ldots, k_{M-1}} \frac{\gamma^{N(2M-1)-1-\sum k_i}}{(N(2M-1) - 1 - \sum k_i)!} \prod_{i=1}^{N} \frac{(2(M-1) - k_i)!}{k_i! (M-1 - k_i)!} |z_i|^k_i,
\]
(54)
where it is understood that the \(k_i\) run over all values such that the arguments in the factorials are non-negative. We can now use the fact that
\[
\frac{\gamma^f}{f!} = \frac{1}{2\pi i} \oint ds e^s s^{-f-1},
\]
(55)
and noting that the resulting sums under the integral are expressible in terms of the associated Laguerre polynomials as
\[
\sum_{k=0}^{M-1} \frac{(2(M-1) - k_0)!}{k_0! (M-1 - k_0)!} (s|z_i|)^k = (M-1) ! (-1)^{M-1} L_{M-1}^{1-2M}(s|z_i|),
\]
(56)
we finally obtain expression (6).

It is also worth noting that a closed-form expression for \(\Psi_2(\zeta)\) is possible in terms of the so-called Lauricella generalized hypergeometric function of type A [Ext76], defined as
\[
F_a^{(\mathbb{A})} \left( \begin{array}{c} b_1, \ldots, b_n \\ c_1, \ldots, c_n \end{array} \right) = \sum \frac{(a)_k \cdots (a)_0}{(c)_k \cdots (c)_0} (b_1)_0 \cdots (b_n)_0 x_1^k \cdots x_n^k,
\]
(57)
where \((a)_k = \frac{(a + i - 1)!}{(a - 1)!}\) is the rising factorial. Therefore, we can alternatively express \(\Psi_2(\zeta)\) as
\[
\Psi_2(\zeta) = \varepsilon(\zeta) \frac{\Gamma(MN)^2}{\Gamma(M)} \frac{\Gamma(2M-1)}{\Gamma(N(2M-1))} \gamma^{N(2M-1)-1} \times F_a^{(\mathbb{A})} \left( \begin{array}{c} 1 - M, \ldots, 1 - M \\ 2(1 - M), \ldots, 2(1 - M) \end{array} \right) \frac{|z_1|}{\gamma}, \ldots, \frac{|z_N|}{\gamma}.
\]
(58)

4. The asymptotic eigenvalue density

We next proceed with the proof of theorem 3 for the AED. The approach we follow in this section again exploits the fact that the random matrix \(Z = \rho_1 - \rho_2\) is the difference of two independent random matrices drawn from unitarily invariant ensembles. In the asymptotic limit, it is well-known that any two such matrices satisfy the so-called freeness condition underpinning Voiculescu’s theory of free probability for non-commuting variables [VDN92]. In a manner analogous to the case of the sum (or difference) of two independent commuting random variables, the AED of \(Z\) can be shown to be given by the so-called free convolution of the AEDs of \(\rho_1\) and \(-\rho_2\).

4.1. Free convolution and the AED

First, let us briefly review the aspects of Free Probability that will be relevant to our analysis; the interested reader is referred to Nica and Speicher’s book [Nic06] on the topic for further details. As mentioned earlier, the ensemble of \(N \times N\) partial density matrices \(\rho = \text{Tr}_\psi \left( |\psi\rangle \langle \psi| \right)\)
where $|\psi\rangle$ is uniformly sampled from $\mathcal{H}_M \otimes \mathcal{H}_N$, is the unitarily invariant FTWL random matrix ensemble. In the asymptotic limit $N \to \infty$, $M \to \infty$ with $c$ fixed, the normalized traces of integer powers of $\rho$ converge, almost surely, to the moments $m_n$ of a well-defined spectral measure, the asymptotic eigenvalue density (AED) $\varrho(\lambda)$ [Spe13]:

$$m_n = \lim_{N \to \infty} \frac{1}{N} \text{Tr}(\rho_N^n) = \int \varrho(\lambda) d\lambda \quad \lambda^n. \quad (59)$$

The AED for the FTWL ensemble was obtained in the context of random partial density matrices by Page [Pag93] following the Coulomb gas analogy. With the knowledge of $\varrho(\lambda)$, the theory of free probability allows us to calculate the AED for the sum (or difference) of two random matrices $\rho_1$ and $\rho_2$ sampled from the FTWL ensemble.

Free probability is a probability theory for non-commutative random variables satisfying a generalization of the ordinary notion of independence. Specializing to the case of random matrices, we say that two random matrices $A$ and $B$ are said to be free if for all $k, m_i, n_j \in \mathbb{N}$,

$$\langle \text{Tr}(\Delta A^{m_1} \Delta B^{n_1} \Delta A^{m_2} \Delta B^{n_2} \ldots \Delta A^{m_k} \Delta B^{n_k}) \rangle = 0, \quad (60)$$

where $\Delta X \equiv X - \langle \text{Tr}(X) \rangle$. A well-known theorem of Voiculescu [Nic06, Voi91] states that if $(A_N)_{N \in \mathbb{N}}$ and $(B_N)_{N \in \mathbb{N}}$ are sequences of $N \times N$ matrices such that their asymptotic eigenvalue densities $\varrho_A(\lambda)$ and $\varrho_B(\lambda)$ exist, and if $(U_N)_{N \in \mathbb{N}}$ is a sequence of Haar unitary $N \times N$ random matrices, then $U_N A_N U_N^T$ and $B_N$ are asymptotically free as $N \to \infty$. From this theorem it follows that any two random matrices, each independently sampled from a unitarily-invariant ensemble, are asymptotically free.

If $A$ and $B$ are free random variables with AEDs $\varrho_A(\lambda)$ and $\varrho_B(\lambda)$, the AED of the sum $C = A + B$ satisfies a generalized notion of convolution, known as free convolution, and denoted by $\varrho_C(\lambda) = \varrho_A(\lambda) \boxplus \varrho_B(\lambda)$. For a given AED $\varrho(\lambda)$, it is convenient to define its Cauchy transform

$$G_\varrho(z) = \int_{\mathbb{R}} \frac{\varrho(\lambda) d\lambda}{z - \lambda}, \quad (61)$$

which is analytic on $\mathbb{C} \setminus \mathbb{R}$. In free probability theory, this function plays the role of a moment-generating (or characteristic) function. Indeed, the moments $m_n$ can be recovered from the Laurent expansion of $G_\varrho(z)$ for sufficiently large $|z|$:

$$G_\varrho(z) = \sum_{n=0}^{\infty} \frac{m_n}{z^{n+1}}, \quad (62)$$

and using the Stiltjes inversion formula [Nic06], the AED $\varrho(\lambda)$ can be recovered from $G_\varrho(z)$ according to:

$$\varrho(\lambda) = \frac{1}{\pi} \lim_{\varepsilon \to 0} \text{Im} G_\varrho(\lambda + i \varepsilon). \quad (63)$$

Closely related to the moments $m_n$ are the so-called free cumulants $k_n$, which in analogy with ordinary cumulants are those combinations of the moments $m_n$ that are additive under the sum of two free random variables. The free cumulant generating function is the so-called $\mathcal{R}$ transform, which is connected to the Cauchy transform through the functional equations

$$G_\varrho \left( R_\varrho + \frac{1}{z} \right) = z, \quad (64)$$

and
The free cumulants are obtained from the power series of $R_\phi(z)$ about the origin:

$$R_\phi(z) := \sum_{n=0}^{\infty} k_n + i^n z^n. \quad (66)$$

For the sum of two free random variables, additivity of the free cumulants implies additivity of the respective $R$ transforms:

$$R_{\phi_1 \oplus \phi_2}(z) = R_{\phi_1}(z) + R_{\phi_2}(z). \quad (67)$$

Thus, the AED of the sum of two free random variables can be obtained using (63) once the Cauchy transform is retrieved from $R_{\phi_1 \oplus \phi_2}(z)$ by means of (65).

4.2. Proof of theorem 3 (AED for $\rho_1 - \rho_2$)

We now specialize the previous discussion to the case of interest. For a random $N \times N$ partial density matrix $\rho$ playing the role of either $\rho_1$ or $\rho_2$, the AED was first computed by Page [Pag93] (see also the result from Nadal et al [NMV11]). Using the well-known 2D Coulomb gas analogy, he showed that for $c = N/M < 1$ the rescaled AED is

$$\tilde{\rho}(x) = \frac{1}{2\pi c x} \sqrt{x - x_-} \sqrt{x_+ - x}, \quad (68)$$

with $x_{\pm} = c(1 \pm c^{-1/2})^2$. Using the fact that the non-zero eigenvalues of the partial density matrices of a bipartite pure state are equal, it is not difficult to show that this result extends to the result of equation (3) for all values of $c$. As pointed out by Nechita [Nec07], this AED corresponds to the well-known Marčenko–Pastur law that arise from the infinite free Poisson process [Nic06]. In terms of the variable $c$, the free cumulants for the Marčenko–Pastur distribution (3) are well-known and given by $k_n = c^{n-1}$ (see e.g. [Nic06]). Using (66), the expression for the $R$ transform is

$$R_\phi(z) = \sum_{n=0}^{\infty} k_n + i^n z^n = \sum_{n=0}^{\infty} c^n z^n = \frac{1}{1 - cz}. \quad (69)$$

The rescaled AED of $-\rho$ is given by the reflected density $\tilde{\rho}(-x)$ with free cumulants $k_n = (-1)^n c^{n-1}$, and hence $R$ transform $\sum_{n=0}^{\infty} (-1)^n c^n z^n = -\frac{1}{1 + cz}$. We therefore have, from (67), that the $R$ transform of the rescaled AED of $Z = \rho_1 - \rho_2$ is

$$R(z) = \frac{1}{1 - cz} + \left(-\frac{1}{1 + cz}\right) = \frac{2cz}{1 - c^2 z^2}. \quad (70)$$

Making use of relation (65), a cubic equation for $G(z)$ can be obtained:

$$\frac{2cG(z)}{1 - c^2 G(z)^2} + \frac{1}{G(z)} = z. \quad (71)$$

The three roots for $G(z)$, indexed by $k = 0,1,2$, can be given in terms of trigonometric functions as

$$G^{(k)} = \frac{2\sqrt{(2 - c)^2 + 3z^2}}{3cz} \sin \left[\frac{1}{3} \arcsin(\eta(z, c)) + k \frac{2\pi}{3}\right] + \frac{c - 2}{3cz}, \quad (72)$$
where
\[ \eta(z) \equiv \frac{9(c + 1)z^2 + (2 - c)^3}{(2 - c^2 + 3z^2)^{3/2}}. \]  

(73)

and \( \text{Arcsin}(z) = -i\text{Log}(iz + (1 - z^2)^{1/2}) \) and \( \text{Log}(z) \) is the principal branch of the \( \text{log}(z) \) function with branch cut along the negative real axis. The actual function \( G(z) \) is built piecewise from the roots \( G^{(k)} \) in such a way that \( G(z) \) is analytic in the region \( \mathbb{C} \setminus \mathbb{R} \), and decays as \( 1/z \) as \( |z| \to \infty \). As mentioned previously, the solution depends on whether \( c < 2 \) or \( c > 2 \), so each case must be discussed separately.

As our interest is in the AED obtained via the Stieltjes inversion formula (63), we only need to concentrate in obtaining the appropriate expression for \( G(x + i\varepsilon) \). In the limit \( x \to \infty \), the root decaying as \( 1/x \) is \( G(0) \). To examine the behavior of this function in the real axis we can express the \( \text{Arcsin} \) in this root as
\[ \text{Arcsin}(\eta + i\varepsilon) = \left\{ \begin{array}{ll}
\frac{\pi}{2} + \text{sgn}(\eta'(x)) \log(\eta(x) + \sqrt{\eta^2(x) - 1}) & |\eta(x)| > 1 \\
\text{arcsin}(\eta(x)) & |\eta(x)| \leq 1
\end{array} \right. \]  

(74)

where \( \text{sgn}(\cdot) \) is the sign function and \( \text{arcsin}(x) \) is the ordinary inverse sine function with domain \([-\pi/2, \pi/2]\). From (74) we can see that a discontinuity arises when \( \eta'(x) \) changes sign, which is indeed the case in both \( c \leq 2 \) and \( c > 2 \) as shown in figures 5(a) and (b), respectively. To analytically continue \( G \), we must then switch to the root \( G(1) \) in the region \( x < c \). Thus we can express \( G(x + i\varepsilon, c) \) for all values of \( x \) as
\[ G(x + i\varepsilon) = \frac{1}{(x + i\varepsilon)} \left[ f_1(x) - i f_2(x) \right], \]

(75)

where \( f_1(x) \) and \( f_2(x) \) are given by
\[ f_1(x) = \left\{ \begin{array}{ll}
\frac{\sqrt{(2 - c)^2 + 3x^2}}{3c} \cosh \left( \frac{l(x)}{3} \right) + \frac{c - 2}{3c} & |\eta(x)| > 1 \\
2\sqrt{(2 - c)^2 + 3x^2} \sin \left[ \frac{1}{3} \text{arcsin}(\eta(x, c)) + \frac{2\pi}{3} \right] + \frac{c - 2}{3c} & |\eta(x)| \leq 1
\end{array} \right. \]  

(76)

and
\[ f_2(x) = \left\{ \begin{array}{ll}
\text{sgn}(x)\frac{\sqrt{(2 - c)^2 + 3x^2}}{\sqrt{3c}} \sinh \left( \frac{l(x)}{3} \right) & |\eta(x)| > 1 \\
0 & |\eta(x)| \leq 1
\end{array} \right. \]  

(77)

and where \( l(x) = \log(\eta(x) + \sqrt{\eta^2(x) - 1}) \). Finally, to obtain the PDF \( \tilde{y}(x) \) via the Stieltjes inversion formula (63), we use the identity
\[ \lim_{\varepsilon \to 0^+} \frac{1}{x + i\varepsilon} = \mathcal{P} \left( \frac{1}{x} - i\pi\delta(x) \right), \]

(78)

where \( \mathcal{P}(\cdot) \) is the principal value of a function. Explicitly,
From equation (79) we can obtain the results from theorem 3. For \( c < 2 \) we have that 
\[ \eta(0) = 1, \] 
therefore 
\[ f_1(0) = 0, \] 
thus there is no point distribution for \( c < 2 \). The non-zero region in (11) arises where 
\[ f_2(x) \neq 0 \] 
which occurs when \( x \leq x_+ \). For \( c \leq 2 \) we have that 
\[ \eta(0) = -1, \] 
so 
\[ f_1(0) = \frac{c-2}{x}, \] 
thus obtaining the point distribution at \( x = 0 \). The non-zero region in (12) arises where 
\[ f_2(x) \neq 0 \] 
that is for \( x \leq x_- \).

4.3. Corollary 1: asymptotics of the operator norm

The upper edge \( x_+ \) of the AED tells us that the expected largest eigenvalue of \( Z \) is 
\( \langle \lambda_{\text{max}} \rangle \approx x_+ / N \). However, this does not guarantee convergence of the operator norm to \( x_+ / N \). This is because there remains the possibility of a \( o(N) \) fraction of extreme eigenvalues that in the limit \( N \to \infty \) ‘do not stick’ to the edge of AED, a possibility is ruled out if \( \|Z\|_{\text{op}} \to x_+ / N \) almost surely. Almost sure convergence of the extreme eigenvalues to the edges of the AED has been verified for a number of ensembles, and in particular for the Wishart ensemble [Bai99, Gem80, Joh00, Sil85], where the extreme eigenvalues are known to follow a Tracy-Widom law [BFP98, Joh01, Sos02]. Moreover, it has been shown that the operator norm of any fixed polynomial of independent and asymptotically free random matrices, sampled from a number of different ensembles [HT05, Mal12], including the Wishart ensemble [DMC07], converges almost surely to the edge of the corresponding AED, a property known as strong asymptotic freeness (SAF).

While the SAF of the fixed trace Wishart–Laguerre ensemble has not been explicitly established in the literature, it is nevertheless not too difficult to deduce from known results related to SAF that the operator norm of the sum (or difference) of two FTWL matrices almost surely converges to the edge of the corresponding AED. In fact, using the SAF of Haar-distributed random unitary matrices, Collins and Male (corollary 2.2 of [CM14]) have shown that for two sequences of independent, unitarily-invariant, random matrices \( A_N \) and \( B_N \), such that in each case the extremal eigenvalues almost surely tend asymptotically to the outmost edges of respective compactly supported AEDs \( \mu \) and \( \nu \), the spectrum of the sum \( A_N + B_N \) will almost surely be contained within any small neighborhood of the support of the free additive convolution \( \mu \boxplus \nu \) for sufficiently large \( N \). Thus, to prove the almost sure convergence of \( N\|Z_N\|_{\text{op}} \) to
$x_\pm$, it remains to verify that the matrices from the FTWL ensemble satisfy the required conditions of Collins and Male’s corollary.

In the FTWL ensemble, it is not too surprising that the extremal eigenvalues should converge almost surely to the edges of the AED, given that the eigenvalues differ from those of the Wishart ensemble by the trace of the Wishart matrix, which concentrates around its mean $MN$ following a $\chi^2$ distribution and hence has $O(N^{-\frac{1}{2}})$ relative fluctuations. Indeed, several results have already connected the probability distributions of the extremal FTWL eigenvalues to their Wishart counterparts [Maj11]. Still, we have found it useful to explicitly establish the connection between the almost sure convergence of the FTWL extremal eigenvalues and the corresponding almost sure convergence in the Wishart Ensemble. For this, one can exploit a particularly transparent relation between the PDF’s for the $j$th eigenvalue in the Wishart and FTWL ensembles, which we term $p_{jW}(\lambda)$ and $p_{jF}(\lambda)$ respectively. Letting $g(t)$ be the $\chi^2$-squared distribution for the trace of the complex Wishart matrix $g(t) = \frac{t^{MN-1}e^{-t}}{\Gamma(MN)}$, the relation reads

$$p_{jW}(s) = \int_0^\infty dt \frac{p_{jF}(sl/t)}{t}. \tag{81}$$

The relation is a straightforward generalization to the complex case of a similar relation established by Davis for the real Wishart ensemble [Dav72], and with a change of variables is easily shown to also be equivalent to an analogous Laplace transform relation [Maj11, MBL08].

Now let $\lambda_{N,\pm}$ be the smallest and largest eigenvalues for the finite matrix ensemble, respectively. By the Borel–Cantelli lemma, almost sure convergence of $\lambda_{N,\pm}$ to the edges $x_{\pm}$ of AED is equivalent to the statement that for any $\epsilon > 0$,

$$\sum_{N=1}^\infty P(|N\lambda_{N,\pm} - x_{\pm}| > \epsilon) < \infty. \tag{82}$$

Now set $M = N/c$ with fixed $c$ (with $c < 1$ w.l.o.g.) in all formulas and let $\delta \tilde{x}_{N,\pm}$ stand for the difference between the empirical rescaled extremal non-zero eigenvalues and the corresponding edge of the respective AED, so that $\delta \tilde{x}_{N,\pm} = N^{-1}\lambda_{N,\pm} - x_{\pm}$ and $\delta \tilde{x}_{N,\pm} = N\lambda_{N,\pm} - x_{\pm}$ for the Wishart and FTWL ensembles respectively. Similarly, let $\delta \tau N = (ct/N^2) - 1$, where $t$ is distributed according to (80). Integrating (81) over $s$ in the interval corresponding to $|\delta \tilde{x}_{N,\pm}| \leq \epsilon$ and splitting the domain of integration in $t$ into a small neighborhood around $t = N^2/c$, where $g(t)$ peaks, and its complement, we can show that for any $k > c$ and any $\epsilon$, with $x_{\pm}^{(W)} > \epsilon > 0$:

$$P(|\delta \tilde{x}_{N,\pm}^{(W)}| > \epsilon) \geq P(|\delta \tau N| < \epsilon')P(|\delta \tilde{x}_{N,\pm}^{(F)}| > k\epsilon), \tag{83}$$

where $\epsilon' = \frac{\epsilon}{c}$. In the inequality, $P(|\delta \tau N| < \epsilon')$ can be replaced by its infimum over $N$, which is non-zero. Thus,

$$\sum_{N=1}^\infty P(|\delta \tilde{x}_{N,\pm}^{(W)}| > \epsilon) < \infty \Rightarrow \sum_{N=1}^\infty P(|\delta \tilde{x}_{N,\pm}^{(F)}| > k\epsilon) < \infty. \tag{84}$$

It then follows from the almost sure convergence of $\lambda_{N,\pm}^{(W)}$ to $N\lambda_{\pm}$ that the extremal eigenvalues of the FTWL ensemble almost surely converge to $x_{\pm}/N$ and hence the spectrum of the
FTWL ensemble satisfies the almost sure convergence conditions required to prove the SAF of the difference of two independent FTLW matrices, according to Collins and Male’s corollary. Therefore, \( \|Z\|_{\text{op}} \) converges almost surely to \( x_1/\sqrt{N} \).

5. Moments and the distance between two random states

Finally, we turn to the proof of theorem 4. To obtain the general expression for the function \( m_c \) in the theorem, we first compute the ordinary moments of the AED, which can be obtained from the Laurent expansion of the Cauchy transform \( G_c(z) \) (62); from the symmetry of the AED it is clear that only even moments exist. The absolute complex moments can be obtained by means of Carlson’s theorem, which allows us to analytically continue the function for the discrete moments.

5.1. Calculating the moments \( m_n \) for even \( n \)

We will begin by calculating the moments \( m_n = \langle x^n \rangle \), for integer \( n \). The moments can be read off from the Laurent expansion \( G_c(z) \) as \( z \to \infty \):

\[
G_c(z) = \sum_{n=0}^{\infty} \frac{m_n}{z^n} = \frac{1}{\Theta(G)}.
\]

(85)

Now, from equation (71) and with the change of variable \( z = \frac{1}{\epsilon} \) we obtain:

\[
\frac{G}{\Theta(G)} = \epsilon,
\]

(86)

where

\[
\Theta(G) = \frac{c^2G^2 - 2cG^2 - 1}{c^2G^2 - 1}.
\]

(87)

This particular form is useful as it allows us to obtain \( G \) as a power series of \( \epsilon \) by applying the Lagrange–Bürmann’s inversion formula [Bü99, Lag70]: namely,

\[
G(z) = \sum_{n=1}^{\infty} \frac{1}{n} \left( \frac{1}{(n-1)!} \lim_{G \to 0} \left( \frac{d^{n-1}}{dG^{n-1}} \Theta(G)^n \right) \right) \epsilon^n.
\]

(88)

The coefficients of the series in (88) will be the moments \( m_{n-1} \) according to (85). Thus, the moments are

\[
m_{n-1} = \frac{1}{n} [G^{n-1}] \Theta(G)^n,
\]

(89)

where \([G^n]\) is an operator that extracts the coefficient of \( n \)th power of \( G \) on the Taylor series expansion of a function. Expanding \( \Theta(G)^n \) as a double sum we get

\[
\Theta(G)^n = \sum_{j,k=0}^{\infty} \frac{n(n+j-1)!}{(n-k)! k!} (2-j^2 + 2c) j^2 G^{2(k+j)},
\]

(90)

and applying the \([G^{n-1}]\) operator we obtain:

\[
m_{n-1} = \sum_{j=0}^{\infty} \frac{(n+j-1)!}{j! \left( \frac{n+1}{2} + j \right)! \left( \frac{n-1}{2} + j \right)! (n-1)!} e^{n-k-1} (2-c)^{2k-j} j^{2j-2k-j}.
\]

(91)
From the extraction of the \((n - 1)\)th power of \(G\) we have that \(n - 1 = 2(k + j)\), and so \(n - 1\) has to be even. Now, letting \(l = \frac{n - 1}{2}\), we obtain
\[
m_{2l} = c^l(2 - c)^l \sum_{j=0}^{\infty} \frac{(2m + j)!}{j!(m + 1 + j)!(m - j)!} \left( \frac{c}{2 - c} \right)^j,
\]
which can be written as a Gauss hypergeometric function \(\mathcal{F}_a(b; c; z)\) as follows:
\[
m_{2l} = \frac{(2l)!}{l!(l + 1)!} c^l(2 - c)^l \mathcal{F}_a \left( \begin{array}{c} 2l + 1, -l; 1 \end{array}; \frac{c}{c - 2} \right)
\]
To analytically extend to complex values of \(l\), we will need expressions for \(\mathcal{F}_a(b; c; x)\) in which \(x < 1\). For any integer \(l > 0\), we can use standard transformations of the hypergeometric function on (93) to show that \(m_{2l}(c) = \mu(l)\), where (for positive integer \(l\) so far),
\[
\mu(l) = \begin{cases} 
\frac{\Gamma(2l + 1)(2c)^l}{\Gamma(l + 1)!\Gamma(l + 2)} \mathcal{F}_a \left( \begin{array}{c} l - l, -l; 1 \end{array}; \frac{c}{2} \right) & \text{for } c < 2, \\
2^{2l-1} \mathcal{F}_a \left( \begin{array}{c} 1 - l, -2l; 2 \end{array}; \frac{c}{c} \right) & \text{for } c > 2.
\end{cases}
\]
Notice that if \(\mu(l)\) is evaluated at \(l = 0\), it yields the result \(2/c\) for \(c > 2\), which corresponds to the integral of \(\tilde{g}(x)\) excluding the Dirac atom at the origin. In the next section we will show that the equality \(m_{2l} = \mu(l)\) extends to complex values of \(l\) with \(\text{Re}(l) > 0\).

5.2. Proof of theorem 4 (the absolute moments \(m_z\))

To show that \(m_{2l} = \mu(l)\) is indeed the correct expression for all absolute moments with complex \(l\) with positive real part, let us define \(f(z)\) as
\[
f(z) = m_{2z} - \mu(z), \quad \text{Re}(z) \geq 0,
\]
where \(m_{2z}\) is the moment function
\[
m_{2z} = 2 \int_0^\infty |x|^{2z} \tilde{g}(x) dx,
\]
and \(\mu(z)\) is as defined in (94). The lower limit \(0^+\) of the integral (96) signifies that the integral is performed excluding the Dirac atom in \(\tilde{g}(x)\) at the origin for \(c > 2\). This choice guarantees that \(m_{2z}\) is analytic for all \(\text{Re}(z) \geq 0\) and that \(f(0) = 0\) for all \(c\). Moreover, from the analyticity of the hypergeometric function with respect to its parameters, it is also ensured that \(f(z)\) is analytic (and hence continuous) for all \(\text{Re}(z) \geq 0\) (for \(\text{Re}(z) < 0\), we understand \(f(z)\) to be the corresponding analytic extension). From the results of the previous section, we know that \(f(z) = 0\) for all \(z \in \mathbb{N}\). Using Carlson’s theorem we can then show that in fact, \(f(z) = 0\) for all \(z \in \mathbb{C}\). Since the integral in (96) is only valid in the region \(\text{Re}(z) \geq 0\), the vanishing of \(f(z)\) in that region suffices for theorem 4. Outside this region, the vanishing of \(f(z)\) tells us that the analytic extension of the integral in (96) is given by \(m_{2z}\).

Carlson’s theorem states that if \(f(z)\) is analytic in \(\text{Re}(z) > 0\), continuous in \(\text{Re}(z) \geq 0\), and satisfies the conditions:
\[
|f(z)| \leq Ce^{\tau |z|}, \quad \text{Re}(z) \geq 0, \quad \tau, C < \infty,
\]

\[ f(iy) \leq Ce^{-|y|}, \quad c < \pi, \quad (97b) \]

\[ f(n) = 0, \quad n \in \mathbb{N}_0, \quad (97c) \]

then \( f \) is identically zero \([\text{Bai35, Rub55}]\). Analyticity and continuity of \( f(z) \) in the required domains was previously ascertained, as was condition (3). Conditions (1) and (2) are satisfied if they can be verified separately for \( m_2 \) and \( \mu(z) \). Indeed, for \( m_2 \), we have \( |m_2| \leq x_+^{2 \text{Re}(z)} \) and \( |m_2(iy)| \leq 1 \). To ascertain conditions (1) and (2) for \( \mu \), the asymptotic behavior of the \( \text{F}_1 \) function for large parameters must be examined. These asymptotics are covered in two papers by Paris \([\text{Par13a, Par13b}]\). Using the expansions of that reference, as well as the asymptotics of the Gamma function, it is not difficult to show that for \( |z| \to \infty \), expression (94) behaves asymptotically as

\[
\mu(x) \approx \begin{cases} 
(2c)(1 - \frac{c}{2})^{3c+1} f_1(t_1(c), c) e^{\pi(3d(c), c)} & \text{for } c < 2, \\
\frac{(z + 1)^{1/2}}{2\sqrt{\pi} e^c} & \text{for } c > 2,
\end{cases}
\]

(98)

where

\[
t_1(c) = \frac{4}{1 + \sqrt{1 + 4c}}, \quad (99a)
\]

\[
\psi_1(t, c) = 2 \log[t] - \log[t - 1] - 2 \log[1 - c/2t], \quad (99b)
\]

\[
f_1(t, c) = \frac{1}{(1 - \frac{c}{2t})^2}, \quad (99c)
\]

\[
t_2(c) = \frac{4(1 - \frac{c}{2})}{3(1 - \sqrt{1 - 4c/9}(1 - \frac{c}{2}))}, \quad (99d)
\]

\[
\psi_2(t, c) = 2 \log\left[\frac{t}{t - 1}\right] + \log\left[1 - \frac{t}{1 - \frac{t}{2}}\right], \quad (99e)
\]

\[
f_2(t, c) = \frac{t}{(t - 1)(1 - \frac{t}{1 - \frac{t}{2}})^2}, \quad (99f)
\]

The asymptotic expansion acquires a particularly simple form when expressed in terms of \( x_+ \), namely,
\[ \mu(x) \approx \begin{cases} \frac{\sqrt{2}}{(z + 1) \sqrt{\pi}} \left( 1 - \frac{c}{2} \right) \frac{f_1(t_0(c), c)}{\sqrt{2\pi}\psi'(t_0(c), c)} & \text{for } c < 2, \\ \frac{\sqrt{2}}{(2z + 1) \sqrt{\pi}} \left( 1 - \frac{c}{2} \right) \frac{f_2(t_2(c), c)}{\sqrt{2\pi}\psi'(t_2(c), c)} & \text{for } c > 2, \end{cases} \] (100)

which clearly are bounded by exponential functions of the form of (97a) and (97b). Thus, by Carlson’s theorem it follows that \( f(z) = 0 \), and thus for \( \text{Re}(z) \geq 0 \), \( m_{zz} = \mu(z) \). Apart from a trivial change of variable \( (2z \to z) \), theorem 4 follows by noting that when \( z \) is restricted to \( \text{Re}(z) > 0 \), the lower limit of integral (96) can be replaced by 0, so as to include the Dirac atom, with no effect whatsoever.

5.3. Corollary 2: the case \( z = 1 \)

For the asymptotic trace distance in corollary 2 we need to specialize to the case \( z = 1 \) in theorem 4, in which case,

\[ m_1 = \begin{cases} \frac{8 \sqrt{2} \sqrt{\pi}}{3\pi} \left( 1 - \frac{1}{2} \right) \left( 1 - \frac{5}{2} \right) \left( 1 - \frac{c}{2} \right) & \text{for } c \leq 2, \\ \frac{2c - 1}{c} & \text{for } c > 2. \end{cases} \] (101)

To obtain the simplified expression (22) from (101), we can make successive use of the recurrence relations

\[ _2F_1(a, b; c; z) = \frac{((c - 1)(c - 2)(1 - z)) _2F_1(a, b; c - 2; z)}{z(a - c + 1)(b - c + 1)} + \frac{((c - 1)(-z(a + b - 2c + 3) - c + 2)) _2F_1(a, b; c - 1; z)}{z(a - c + 1)(b - c + 1)} \] (102)

and

\[ _2F_1(a, b; c; z) = \frac{(z(a - b - 1) + 2b - c + 2) _2F_1(a, b + 1; c; z)}{b - c + 1} + \frac{((b + 1)(z - 1)) _2F_1(a, b + 2; c; z)}{b - c + 1} \] (103)

together with the identity

\[ \text{arcsin}(z) = _2F_1\left( \frac{1}{2}, \frac{1}{2}; \frac{3}{2}; z^2 \right). \] (104)

With these identities, it follows that

\[ _2F_1\left( \frac{1}{2}, \frac{1}{2}; \frac{5}{2}; z^2 \right) = \frac{3(2c - 1) \text{arcsin}\left( \frac{c}{\sqrt{2}} \right)}{4\sqrt{2}c^{3/2}} + \frac{3 \sqrt{1 - c^2} (c + 1)}{8c}, \] (105)

which yields the result of corollary 2 for \( c < 2 \) when multiplied by \( \frac{\sqrt{2}c^{3/2}}{6c} \).
6. Conclusions

The aim of this paper has been to quantify the distance between two random density matrices, each independently sampled from the so-called Fixed Trace Wishart–Laguerre (FTWL) Ensemble. To this end, we have successfully obtained expressions for the joint PDF of the eigenvalues of the difference matrix given finite Hilbert space sizes, the limiting eigenvalue density (AED) in the asymptotic limit of infinite Hilbert space dimensions, and closed-form expressions for the absolute moments of the asymptotic density. The asymptotic expressions were used to quantify the almost sure asymptotic behavior of two distance measures for the two random states; namely, the operator norm distance and the trace distance, both of which are of the same order as the corresponding distances to the maximally mixed state, as expected from the concentration of measure phenomenon of partial density matrices around the maximally mixed state [HLW06, Pag93, PSW06].

Acknowledgments

The authors are grateful to Matthias Christandl, and Satya Mujumdar for helpful discussions, and the two referees for valuable suggestions, particularly in connection with corollary 1. AB gratefully acknowledges funding by Uniandes, proyecto Ciencias Básicas No. 114-2013.

Appendix A.

The main result of section 4 can be generalized to the case where the random matrices do not have the same weight in the difference. In this appendix we will show how to compute the AED for this asymmetric case, namely

\[ Z = p\rho_1 - q\rho_2. \]  

We can define a normalized matrix \( \hat{Z} = Z/p \) and if \( \eta := q/p \), we can find the AED for

\[ \hat{Z} = \rho_1 - \eta\rho_2, \]  

and then go back to the original variables. We begin by calculating the \( \mathcal{R} \) transforms for the AEDs of \( \rho_1 \) and \( -\eta\rho_2 \) which we will call \( \mathcal{R}_1 \) and \( \mathcal{R}_2 \) respectively. The \( n \)th cumulant is a homogeneous function of degree \( n \), i.e. \( k_n[\eta \hat{G}] = (-\eta)^n k_n[\hat{G}] \) thus the \( \mathcal{R} \) transforms are:

\[ \mathcal{R}_1(z) = \frac{1}{1 - cz}, \]  

\[ \mathcal{R}_2(z) = -\frac{\eta}{1 + c\eta z}, \]  

where \( c = N/M \) as in section 4. The \( \mathcal{R} \) transform for the AED of \( \hat{Z} \) will be the sum of the \( \mathcal{R} \) transforms, explicitly

\[ \mathcal{R}(z) = \frac{1}{1 - cz} + \left( -\frac{\eta}{1 + c\eta z} \right) = \frac{(1 - \eta)(1 + 2cz)}{(1 - cz)(1 + \eta cz)}. \]  

Following the same procedure as in section 4.2 we obtain an equation for the Cauchy transform \( G(z) \)
Equation (A.6) is cubic on $G(z)$, analyzing the possible roots with the same criteria of analyticity as in the symmetrical case we find the desired AED

\[
\frac{(1 - \eta) + \eta(2cG(z))}{(1 - cG(z))(1 + \eta cG(z))} + \frac{1}{G(z)} = z. \tag{A.6}
\]

Equation (A.6) is cubic on $G(z)$, analyzing the possible roots with the same criteria of analyticity as in the symmetrical case we find the desired AED

\[
\varphi(x, c, \eta) = \frac{1}{2\pi|x|c\eta} \left( w(x, c, \eta) - \frac{(1 + \eta + \eta^2)x^2 - (1 + c)(\eta - 1)\eta x + (c - 2\eta^2)}{3w(x, c, \eta)} \right), \tag{A.7}
\]

with

\[
w(x, c, \eta) = \left( \sqrt{\frac{\eta^2}{4}((x - x_{c,\eta})^2 - \alpha^2)((x - x_{c,\eta})^2 - \beta^2)} - \frac{f(x, c, \eta)}{6\sqrt{3}} \right)^{\frac{1}{3}}, \tag{A.8}
\]

where

\[
x_{c,\eta} = \frac{(1 + c)(1 - \eta)}{2} \tag{A.9}
\]

and
\[ f(x, c, \eta) = ((x - a) - \gamma)((x - a) - \tilde{\gamma})(x - a - \epsilon), \quad (A.10) \]

where
\[
a = \frac{(1 + c)\eta(1 + \eta(4 + \eta))}{(\eta - 1)(2 + \eta)(1 + 2\eta)}, \quad (A.11)\]

The parameters \(\alpha, \beta, \gamma\) and \(\epsilon\) are functions of \(c\) and \(\eta\) and have cumbersome expressions which we will not write here.

The structure is very similar to the result obtained in (15) with the difference that the limits \([x_-, x_+]\) are changed with \(\eta\). Although it is a complicated expression for \(\varrho\), the absolute maximum limit can be found in the case \(\eta > 1\). This limit is equal to \(\frac{(1 + \alpha(1 - \gamma)}{\beta} - \beta\). In figures A1 and A2 we compare our result with numerical simulations, note that we have only considered here the case \(c < 2\).

Appendix B.

In this appendix we derive average eigenvalue density (8) for the case \(N = 2\). The first step is to note that for \(N = 2\), the two eigenvalues satisfy the condition \(\lambda_1 = -\lambda_2\), hence there is only one independent eigenvalue. Using this constraint, equations (5) and (52), with \(\gamma = 1 - |\lambda|\), the PDF for the independent eigenvalue can be written as
\[
\varrho(\lambda) = -\lambda \frac{d}{d\lambda} \left( \frac{\Gamma(2M)^2}{\Gamma(M)^2} (1 - |\lambda|)^{M-3} W \left( \frac{|\lambda|}{(1-|\lambda|)^2} \right) \right). \quad (B.1)\]

where
\[
W(\alpha) = \int_0^1 dx \, (\alpha(1-x))^{M-1}(\alpha(1-x) + \alpha)^{M-1}. \quad (B.2)\]

Expanding the second factor in the integral in a binomial expansion, and using the Beta integral, \(W(\alpha)\) can be expressed in terms of a hypergeometric function
\[
W(\alpha) = \sum_{k=0}^{M-1} \left( \begin{array}{c} M-1 \\ k \end{array} \right) \frac{(2(M-1)-k)!^2}{(4(M-1)-2k+1)!} \alpha^k \quad (B.3)\]
\[
= \frac{(2M-1)!^2}{(4M-3)!} \binom{3/2 - 2M}{2(1-M)} - 4\alpha. \quad (B.4)\]

Using the standard hypergeometric identity
\[
\binom{a}{c} \binom{b}{z} = (1-z)^{-b} \binom{c-a}{c} \left( \frac{z}{z-1} \right). \quad (B.5)\]

\(W(\alpha)\) can further be simplified to
\[
W(\alpha) = \frac{(2M-1)!^2}{(4M-3)!} (1 + 4\alpha)^{M-1} \binom{1/2, 1-M}{2(1-M)} \frac{4\alpha}{1+4\alpha}. \quad (B.6)\]

Substituting into equation (B.1), we finally obtain (8) after simplification.
References

[AHSW06] Abeyesinghe A, Hayden P, Smith G and Winter A J 2006 Optimal superdense coding of entangled states IEEE Trans. Inf. Theory 52 3635–41

[ATK09] Adachi S, Toda M and Kubotani H 2009 Random matrix theory of singular values of rectangular complex matrices I: exact formula of one-body distribution function in fixed-trace ensemble Ann. Phys. 324 2278–358

[Aub12] Aubrun G 2012 Partial transposition of random states and non-centered semicircular distributions Random Matrices: Theory Appl. 01 1250001

[AV11] Akemann G and Vivo P 2011 Compact smallest eigenvalue expressions in Wishart Laguerre ensembles with or without a fixed trace J. Stat. Mech. P05020

[Bü99] Bürmann H H 1799 Essai de calcul fonctionnaire aux constantes ad-libitum Mem. Inst. Nat. Sci Arts Sci. Math. Phys. 2 13–7

[Bai35] Bailey W N 1935 Generalized Hypergeometric Series (Cambridge: Cambridge University Press)

[Bai99] Bai Z D 1999 Methodologies in spectral analysis of large dimensional random matrices, a review Stat. Sin. 9 611–62

[BBCC11] Banica T, Belinschi S T, Capitaine M and Collins B 2011 Free Bessel laws Can. J. Math. 63 3–37

[BFP98] Baker T H, Forrester P J and Pearce P A 1998 Random matrix ensembles with an effective extensive external charge J. Phys. A: Math. Gen. 31 6087

[BH10] Brandão F G S L and Horodecki M 2010 On Hastings’ counterexamples to the minimum output entropy additivity conjecture Open Syst. Inf. Dyn. 17 31–52

[BN13] Banica T and Nachita I 2013 Asymptotic eigenvalue distributions of block-transposed Wishart matrices J. Theor. Probab. 26 855–69

[CDKW14] Christandl M, Doran B, Kousidis S and Walter M 2014 Eigenvalue distributions of reduced density matrices Commun. Math. Phys. 332 1–52

[CLZ10] Chen Y, Liu D-Z and Zhou D-S 2010 Smallest eigenvalue distribution of the fixed-trace Laguerre beta-ensemble J. Phys. A: Math. Theor. 43 315303

[CM14] Collins B and Male C 2014 The strong asymptotic freeness of Haar and deterministic matrices Ann. Sci. de l’ENS 47 147–63

[CN11] Collins B and Nechita I 2011 Random quantum channels II: entanglement of random subspaces, Rényi entropy estimates and additivity problems Adv. Math. 226 1181–201

[Col05] Collins B 2005 Product of random projections, Jacobi ensembles and universality problems arising from free probability Probab. Theory Relat. Fields 133 315–44

[Dav72] Davis A W 1972 On the ratios of the individual latent roots to the trace of a wishart matrix J. Multivariate Anal. 2 440–3

[DN12] Deya A and Nourdin I 2012 Convergence of Wigner integrals to the tetilla law Latin Am. J. Probab. Math. Stat. 9 101–27

[DLMS14] Dahlsten O C O, Lupo C, Mancini S and Serafini A 2014 Entanglement typicality J. Phys. A: Math. Theor. 47 363001

[DMC07] Donati-Martin C and Capitaine M 2007 Strong asymptotic freeness for Wigner and Wishart matrices Indiana Univ. Math. J. 56 767–803

[Ext76] Exton H 1976 Multiple Hypergeometric Functions and Applications (Mathematics and its Applications) (New York: Halsted Press)

[FK94] Forrester P J and Kanno S 1994 Proof of page’s conjecture on the average entropy of a subsystem Phys. Rev. Lett. 72 1148–51

[FK10] Fukuda M and King C 2010 Entanglement of random subspaces via the Hastings bound J. Math. Phys. 51 042201

[For10] Forrester P J 2010 Log-Gases and Random Matrices (Princeton, NJ: Princeton University Press)

[Gem80] Geman S 1980 A limit theorem for the norm of random matrices Ann. Probab. 8 252–61

[GHT15] Goldstein S, Haro T and Tasaki H 2015 Extremely quick thermalization in a macroscopic quantum system for a typical nonequilibrium subspace New J. Phys. 17 045002

[GLTZ06] Goldstein S, Lebowitz J L, Tumulka R and Zanghì N 2006 Canonical typicality Phys. Rev. Lett. 96 050403
[GOM01] Gemmer J, Ote A and Mahler G 2001 Quantum approach to a derivation of the second law of thermodynamics Phys. Rev. Lett. 86 1927–30

[GP90] Guillemin V and Prato E 1990 Heckman, Kostant and Steinberg formulas for symplectic manifolds Adv. Math. 82 160–79

[Gu94] Guillemin V 1994 Moment Maps and Combinatorial Invariants of Hamiltonian Tn-Spaces (Boston, MA: Birkhäuser Boston)

[Has09] Hastings M B 2009 Superadditivity of communication capacity using entangled inputs Nat. Phys. 5 255–7

[Hec82] Heckman G J 1982 Projections of orbits and asymptotic behavior of multiplicities for compact connected Lie groups Inventiones Math. 67 333–56

[HHL04] Harrow A, Hayden P and Leung D 2004 Superdense coding of quantum states Phys. Rev. Lett. 92 187901

[HLW06] Hayden P, Leung D W and Winter A 2006 Aspects of generic entanglement Commun. Math. Phys. 265 95–117

[HT05] Haagerup U and Thorbjørnsen S 2005 A new application of random matrices: Ext(c *red (f2)) is not a group Ann. Math. 2 711–75

[HW08] Hayden P and Winter A 2008 Counterexamples to the maximal p-norm multiplicativity conjecture for all p > 1 Commun. Math. Phys. 284 263–80

[HW13] Hutter A and Wehner S 2013 Dependence of a quantum-mechanical system on its own initial state and the initial state of the environment it interacts with Phys. Rev. A 87 012121

[IZ80] Itzykson C and Zuber J-B 1980 The planar approximation. II J. Math. Phys. 21 411–21

[Joh01] Johnstone I M 2001 On the distribution of the largest eigenvalue in principal components analysis Ann. Stat. 29 295–327

[Joh00] Johansson K 2000 Shape fluctuations and random matrices Commun. Math. Phys. 209 437–76

[KAT08] Kabotani H, Adachi S and Toda M 2008 Exact formula of the distribution of Schmidt eigenvalues for dynamical formation of entanglement in quantum chaos Phys. Rev. Lett. 100 240501

[Lag70] Lagrange J-L 1770 Nouvelle méthode pour résoudre les équations littérales par le moyen des séries Mém. Acad. R. Sci. Belles-Lettres Berlin 24 251–326

[LP88] Lloyd S and Pagels H 1988 Complexity as thermodynamic depth Ann. Phys. 188 186–213

[LPSW09] Linden N, Popescu S, Short A J and Winter A 2009 Quantum mechanical evolution towards thermal equilibrium Phys. Rev. E 79 061103

[LPSW10] Linden N, Popescu S, Short A J and Winter A 2010 On the speed of fluctuations around thermodynamic equilibrium New J. Phys. 12 055021

[Lab78] Lubkin E 1978 Entropy of an n-system from its correlation with a k-reservoir J. Math. Phys. 19 1028–31

[Maj11] Majumdar S 2011 Extreme eigenvalues of Wishart matrices and entangled bipartite system The Oxford Handbook of Random Matrix Theory (Oxford: Oxford University Press) ch 37

[Mal12] Male C 2012 The norm of polynomials in large random and deterministic matrices Probab. Theory Relat. Fields 154 477–532

[MBL08] Majumdar S N, Bohigas O and Lakshminarayan A 2008 Exact minimum eigenvalue distribution of an entangled random pure state J. Stat. Phys. 131 33–49

[Meh04] Mehta M L 2004 Random Matrices (San Diego, CA: Academic)

[MP67] Marčenko V A and Pastur L A 1967 Distribution of eigenvalues for some sets of random matrices Math. USSR Sb. 1 457–83

[Nec07] Nechaï I 2007 Asymptotics of random density matrices Ann. Henri Poincaré 8 1521–38

[Nic06] Nica A 2006 Lectures on the Combinatorics of Free Probability (London Mathematical Society Lecture Note Series vol 335) (Cambridge: Cambridge University Press)

[NMV11] Nadal C, Majumdar S N and Vergassola M 2011 Statistical distribution of quantum entanglement for a random bipartite state J. Stat. Phys. 142 403–38

[NS98] Nica A and Speicher R 1998 Commutators of free random variables Duke Math. J. 92 553–92

[ODP07] Oliveira R, Dahlsten O C O and Plenio M B 2007 Generic entanglement can be generated efficiently Phys. Rev. Lett. 98 130502

[Pag93] Page D N 1993 Average entropy of a subsystem Phys. Rev. Lett. 71 1291–4

[Par13a] Paris R B 2013 Asymptotics of the Gauss hypergeometric function with large parameters, I J. Class. Anal. 2 183–203
[Par13b] Paris R B 2013 Asymptotics of the Gauss hypergeometric function with large parameters, II J. Class. Anal. 3 1–15
[PDP08] Douglas A, Plato K, Dahlsten O C and Plenio M B 2008 Random circuits by measurements on weighted graph states Phys. Rev. A 78 042332
[PPŻ16] Puchała Z, Pawela Ł and Życzkowski K 2016 Distinguishability of generic quantum states Phys. Rev. A 93 062112
[PSW06] Popescu S, Short A J and Winter A 2006 Entanglement and the foundations of statistical mechanics Nat. Phys. 2 754–8
[Rei07] Reimann P 2007 Typicality for generalized microcanonical ensembles Phys. Rev. Lett. 99 160404
[Rub55] Rubel L A 1955 Necessary and sufficient conditions for Carlson’s theorem on entire functions Proc. Natl Acad. Sci. USA 41 601–3
[Sen96] Sen S 1996 Average entropy of a quantum subsystem Phys. Rev. Lett. 77 1–3
[Sil85] Silverstein J W 1985 The smallest eigenvalue of a large dimensional wishart matrix Ann. Probab. 13 1364–8
[Sos02] Soshnikov A 2002 A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices J. Stat. Phys. 108 1033–56
[Spe13] Speicher R 2013 Asymptotic eigenvalue distribution of random matrices and free stochastic analysis Random Matrices and Iterated Random Functions (Springer Proceedings in Mathematics and Statistics vol 53) ed G Alsmeyer and M Löwe (Berlin: Springer)
[SR95] Sánchez-Ruiz J 1995 Simple proof of Page’s conjecture on the average entropy of a subsystem Phys. Rev. E 52 5653–5
[SŻ04] Sommers H-J and Życzkowski K 2004 Statistical properties of random density matrices J. Phys. A: Math. Gen. 37 8457
[VDN92] Voiculescu D V, Dykema K J and Nica A 1992 Free Random Variables: a Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras and Harmonic Analysis on Free Groups (Providence, RI: American Mathematical Society)
[Viv11] Vivo P 2011 Largest Schmidt eigenvalue of random pure states and conductance distribution in chaotic cavities J. Stat. Mech. P01022
[Voi91] Voiculescu D 1991 Limit laws for random matrices and free products Inventiones Math. 104 201–20
[Žn07] Žnidarič M 2007 Entanglement of random vectors J. Phys. A: Math. Theor. 40 F105