# Quantum Entanglement in high dimensions

Guillaume AUBRUN

October 26, 2016

These lecture notes study some mathematical aspects of the phenomenon of entanglement from quantum mechanics. While the questions we consider are motivated by quantum information theory, where entanglement plays a fundamental role, our exposition targets mostly mathematicians who are not assumed to be familiar with quantum information theory.

We look at entanglement through the prism of "Asymptotic Geometric Analysis", a branch of functional analysis also known as "local theory of Banach spaces" whose objects of study are the normed spaces of large but finite dimension. Indeed, we especially focus on the case of quantum systems of large dimension, for which numerical approaches are usually doomed by the curse of dimensionality.

These notes are organized as follows: in Section 1 we introduce the dichotomy between entangled vs separated states. In Section 2 we explain various approaches to quantify how much entanglement contains a quantum state, notably the "entanglement of formation". Section 3 explains how to use concentration of measure in the form of Dvoretzky's theorem to prove that the entanglement of formation is not additive, a major result first obtain by Hastings [19]. The last two sections study random states, and in particular the question whether they are entangled or separable. The answer relies on volume estimates connected to the convex body of separable states.

We also introduce notation which will be used throughout the text. We consider a complex Hilbert space $\mathcal{H}$ of finite dimension. We denote by $B(\mathcal{H})$ the set of operators on $\mathcal{H}$ and $B^{\mathrm{sa}}(\mathcal{H})$ the subset of self-adjoint operators. We always identify operators with matrices; we denote by $\mathsf{M}_{k,d}$ the space of $k \times d$ matrices with complex entries.

We use the convention from physics to take the scalar product on $\mathcal{H}$ to be anti-linear in the first variable and linear in the second variable. We also use Dirac notation: given $x, y \in \mathcal{H}$, we denote by $|x\rangle\langle y|$ the rank one operator which maps $z \in \mathcal{H}$ to $\langle y, z\rangle x$.

Most of the material presented here will appear also in greater detail in the forthcoming book [4], to which we refer the reader for more information. Other sources are [34] for local theory of Banach spaces, [2] for random matrices theory and [31, 44] for quantum information theory.

# 1 The fundamental dichotomy: entanglement vs separability

## 1.1 Quantum states

A main object of interest in quantum information theory is the set of quantum states. A *quantum state* on $\mathcal{H}$ is a positive self-adjoint operator with trace 1. The set of quantum states is denoted

$$\mathrm{D}(\mathcal{H}) = \{\rho \in B^{\mathrm{sa}}(\mathcal{H}) \ : \ \rho \geqslant 0, \mathrm{Tr}\,\rho = 1\}.$$

The letter D stands for the alternative name "density matrix". Note that for $\rho \in \mathrm{D}(\mathcal{H})$, the linear form defined on $B(\mathcal{H})$ by $X \mapsto \mathrm{Tr}(X\rho)$ if positive with norm 1 and is therefore a state in the usual functional-analytic sense. In the following we simply say "state" to mean "quantum state".

The set $\mathrm{D}(\mathcal{H})$ is a compact convex set with (real) dimension $d^2 - 1$. Its extreme points are the *pure states* on $\mathcal{H}$, i.e., the rank 1 orthogonal projections of the form $|x\rangle\langle x|$. We often consider abusively a unit vector $x$ as a pure state; what is really meant is the pure state $|x\rangle\langle x|$. Note that $|x\rangle\langle x| = |y\rangle\langle y|$ if and only if $x = e^{i\theta}y$ for some $\theta \in \mathbb{R}$. In particular, the set of pure states naturally identifies with the projective space on $\mathcal{H}$.

Elements of $\mathrm{D}(\mathcal{H})$ are often called mixed states. Indeed, we know from the spectral theorem that any quantum state is a convex combination (a "mixture") of pure states:

$$\mathrm{D}(\mathcal{H}) = \mathrm{conv}\{|x\rangle\langle x| \ : \ x \in \mathcal{H}, |x| = 1\}.$$

The "less pure" quantum state is the maximally mixed state $\rho_* := \mathrm{Id}/d$.

Whenever we apply concepts from Euclidean geometry to quantum states, it is tacitly understood that the reference Euclidean structure is given by the Hilbert–Schmidt scalar product on $B^{\mathrm{sa}}(\mathcal{H})$, defined as $\langle A, B\rangle = \mathrm{Tr}(AB)$. The corresponding norm is denoted $\|\cdot\|_{\mathrm{HS}}$. For example, it is a simple exercise to compute the inradius (=the radius of a largest inscribed Hilbert–Schmidt ball) and the outradius (=the radius of the smallest circumscribed Hilbert–Schmidt) of $\mathrm{D}(\mathcal{H})$.

*Exercise* 1. Show that the outradius of $\mathrm{D}(\mathcal{H})$ equals $\sqrt{(d-1)/d}$ and that the inradius of $\mathrm{D}(\mathcal{H})$ equals $1/\sqrt{d(d-1)}$.

For $d = 2$, the in- and out-radii coincide: $\mathrm{D}(\mathbb{C}^2)$ is a Hilbert–Schmidt ball, called the Bloch ball. This should be compared to the classical identification $S^2 \simeq \mathbb{CP}^1$ at the level of pure states. This is specific to the 2-dimensional case: in higher dimensions, $\mathrm{D}(\mathbb{C}^d)$ should rather be considered as the non-commutative analog of a simplex (incidentally, the radii computed at Exercise 1 equal the radii of a $(d-1)$-dimensional simplex embedded in $\mathbb{R}^d$).

*Exercise* 2. Describe the faces of $\mathrm{D}(\mathcal{H})$ of maximal dimension.

## 1.2 Symmetries of $\mathrm{D}(\mathcal{H})$

It is often fruitful to classify the symmetries of a set. As we will see now, the study of the symmetries of $\mathrm{D}(\mathcal{H})$ will put forward the transposition map.

We denote by $P(\mathcal{H})$ the projective space over $\mathcal{H}$, equipped with the metric $\delta$ obtained as the quotient metric from the geodesic metric on the sphere. Given a unit vector $\psi \in \mathcal{H}$, let $|\psi\rangle$ the corresponding element in $P(\mathcal{H})$, so that $|e^{i\theta}\psi\rangle = \psi$ for $\theta$ real.

A result known as Wigner's theorem [43] states that isometries of the metric space $(P(\mathcal{H}), \delta)$ are of the form $|\psi\rangle \mapsto |U\psi\rangle$ or $|\psi\rangle \mapsto |\overline{U\psi}\rangle$ where $U$ is a unitary transformation (the bar denotes complex conjugation with respect to a fixed basis in $\mathcal{H}$).

This is especially transparent for $d = 2$: as we already said, $P(\mathbb{C}^2) \simeq S^2$ so that Wigner's theorem is simply the decomposition of $\mathsf{O}(3)$ as $\mathsf{SO}(3)$ vs $\mathsf{O}(3)\backslash\mathsf{SO}(3)$.

*Exercise* 3. State and prove the real version of Wigner's theorem. Proving the complex version is more delicate, see [38].

Note that

$$|\overline{\psi}\rangle\langle\overline{\psi}| = \overline{|\psi\rangle\langle\psi|} = |\psi\rangle\langle\psi|^T,$$

where $T$ denotes the transposition with respect to the fixed basis.

*Exercise* 4. Let $A \mapsto A^T$ and $A \mapsto A^{T'}$ denote transpositions with respect to two different bases. Check that $A^{T'} = VA^TV^\dagger$ for some unitary map $V$.

An easy consequence of Wigner's theorem is the description of the isometry group of $D(\mathcal{H})$ ("isometry" is meant with respect to the Hilbert–Schmidt Euclidean structure): since isometries on $D(\mathcal{H})$ induce isometries at the level of extreme points, they must be of the form

$$\rho \mapsto U\rho U^\dagger$$

or

$$\rho \mapsto U\rho^T U^\dagger$$

for some unitary map $U$.

### 1.3 Entanglement vs separability

We now assume that $\mathcal{H}$ is a multipartite Hilbert space, i.e., of the form $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$ (we often consider the simpler bipartite case $\mathcal{H} = \mathbb{C}^d \otimes \mathbb{C}^d$). There are canonical isomorphisms

$$B(\mathcal{H}) \simeq B(\mathcal{H}_1) \otimes \cdots \otimes B(\mathcal{H}_k),$$

$$B_{\mathrm{sa}}(\mathcal{H}) \simeq B_{\mathrm{sa}}(\mathcal{H}_1) \otimes_{\mathbb{R}} \cdots \otimes_{\mathbb{R}} B_{\mathrm{sa}}(\mathcal{H}_k). \tag{1}$$

Note that the analogue of (1) would be false for real Hilbert spaces!

A state $\rho \in \mathrm{D}(\mathcal{H})$ is a *product state* if

$$\rho = \rho_1 \otimes \cdots \otimes \rho_k$$

for some states $\rho_i \in \mathrm{D}(\mathcal{H}_i)$. We now introduce the most important definition of these notes: the notion of a separable state which was introduced by Werner [42]

A state is called separable if it can be written as a convex combination of product states

States which are not separable are called *entangled*. We denote by $\mathrm{Sep}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k)$ or simply $\mathrm{Sep}(\mathcal{H})$ the set of all separable states on $\mathcal{H}$. It is easily checked that

$$\begin{aligned} \mathrm{Sep}(\mathcal{H}) &= \mathrm{conv}\{\rho_1 \otimes \cdots \otimes \rho_k \ : \ \rho_i \in \mathrm{D}(\mathcal{H}_i)\} \\ &= \mathrm{conv}\{|\psi_1 \otimes \cdots \otimes \psi_k\rangle\langle\psi_1 \otimes \cdots \otimes \psi_k| \ : \ \psi_i \in \mathcal{H}_i, |\psi_i| = 1\}. \end{aligned}$$

Given closed convex sets $K \subset \mathbb{R}^n$ and $K' \subset \mathbb{R}^{n'}$, we may define

$$K \widehat{\otimes} K' = \overline{\mathrm{conv}}\{x \otimes x' \ : \ x \in K, x' \in K'\} \tag{2}$$

(the closure operation may be dropped when $K$ and $K'$ are compact. If $K$ and $K'$ are unit balls for some norms, we recover the notion of projective tensor product of normed spaces. We have

$$\mathrm{Sep}(\mathcal{H}_1 \otimes \mathcal{H}_2) = \mathrm{D}(\mathcal{H}_1) \widehat{\otimes} \mathrm{D}(\mathcal{H}_2).$$

*Exercise* 5. Show that $K$ and $K'$ have both non-empty interior if and only if $K \widehat{\otimes} K'$ has nonempty interior.

We denote the symmetrization of a convex compact set $K$ as $K_{\mathrm{sym}} = \mathrm{conv}(-K \cap K)$ (this operation increases dimension by 1 if $0 \notin K$). Then the symmetrization of the set of states is the self-adjoint part of the trace class unit ball

$$\mathrm{D}(\mathcal{H})_{\mathrm{sym}} = \{A \in B^{\mathrm{sa}}(\mathcal{H}_1) \ : \ \|A\|_1 \leqslant 1\}.$$

It also holds (denoting $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$) that

$$\mathrm{Sep}(\mathcal{H})_{\mathrm{sym}} = \mathrm{D}(\mathcal{H}_1)_{\mathrm{sym}} \widehat{\otimes} \mathrm{D}(\mathcal{H}_2)_{\mathrm{sym}},$$

from which one checks that

$$\dim \mathrm{Sep}(\mathcal{H}) = \dim(\mathcal{H})^2 - 1 = \dim \mathrm{D}(\mathcal{H}),$$

and similarly for larger number of factors.

*Exercise* 6. Define $L(\mathcal{H})$ as the smallest integer $N$ such that any separable state $\rho \in \mathrm{Sep}(\mathcal{H})$ can we written as a convex combination of length $N$ of pure product states

$$\rho = \sum_{i=1}^{N} \lambda_i |\psi_1^i \otimes \cdots \otimes \psi_k^i\rangle\langle\psi_1^i \otimes \cdots \otimes \psi_k^i|.$$

Show that, for some constant $c$

$$c d^3 \leqslant L(\mathbb{C}^d \otimes \mathbb{C}^d) \leqslant d^4$$

(the right inequality follows from Carathéodory's theorem and the left inequality from a dimension-counting argument, see [12]). The asymptotic growth of $L(\mathbb{C}^d \otimes \mathbb{C}^d)$ is unknown.

## 1.4 The Barnum–Gurvits theorem

Let $\mathcal{H}$ be a multipartite Hilbert space. Although $\mathrm{Sep}(\mathcal{H})$ is a smaller set than $\mathrm{D}(\mathcal{H})$ (being defined via the convex hull of a smaller set), they both have the same dimension. A remarkable result due to Barnum and Gurvits is that in the bipartite case both sets also have the same inradius.

**Theorem 1** (Barnum–Gurvits, [17]). *Let $\mathcal{H} = \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ and denote $n = d_1 d_2 = \dim \mathcal{H}$. If a state $\rho \in \mathrm{D}(\mathcal{H})$ satisfies $\|\rho - \rho_*\|_{\mathrm{HS}} \leqslant \frac{1}{\sqrt{n(n-1)}}$, then $\rho$ is separable.*

We give a sketch of proof due to Hans-Jürgen Sommers [39]. Denote

$$K = \left\{ \rho \in \mathrm{D}(\mathcal{H}) \ : \ \|\rho - \rho_*\|_{\mathrm{HS}} \leqslant \frac{1}{\sqrt{n(n-1)}} \right\}$$

the Hilbert–Schmidt ball inscribed inside $\mathrm{D}(\mathcal{H})$. The inclusion $K \subset \mathrm{Sep}(\mathcal{H})$ is equivalent to the inclusions of cones $\mathbb{R}^+ K \subset \mathbb{R}^+ \mathrm{Sep}(\mathcal{H})$. By the Hahn–Banach separation theorem, this is further equivalent to the following statement: whenever $M \in B^{\mathrm{sa}}(\mathcal{H})$ satisfies

$$\mathrm{Tr}\, M\rho \geqslant 0 \text{ for any } \rho \in \mathrm{Sep}(\mathcal{H}), \tag{3}$$

then $\mathrm{Tr}\, M\rho \geqslant 0$ for any $\rho \in K$. A matrix $M$ satisfying the condition (3) is called *block-positive*. Block positivity means that $\langle \psi_1 \otimes \psi_2 | M | \psi_1 \otimes \psi_2 \rangle \geqslant 0$ for any $\psi_1 \in \mathbb{C}^{d_1}, \psi_2 \in \mathbb{C}^{d_2}$. A simple computation using the Pythagorean theorem reduces the proof of Theorem 1 to the following statement: for any block-positive matrix $M$, $\mathrm{Tr}(M^2) \leqslant (\mathrm{Tr}\, M)^2$. We use the following lemma.

**Lemma 2.** *If $M = \begin{pmatrix} A & B \\ B^\dagger & C \end{pmatrix}$ is block-positive, then $\|B\|_{\mathrm{HS}}^2 \leqslant \|A\|_1 \|C\|_1$.*

Let $M$ be a block-positive matrix on $\mathcal{H}$. Denote $M = (M_{kl})_{1 \leqslant k,l \leqslant d_1}$, where each block $M_{kl}$ is an element of $B(\mathbb{C}^{d_2})$. Diagonal blocks (of the form $M_{kk}$) are positive operators. By Lemma 2, we have $\|M_{kl}\|_{\mathrm{HS}}^2 \leqslant \|M_{kk}\|_1 \|M_{ll}\|_1 = (\mathrm{Tr}\, M_{kk})(\mathrm{Tr}\, M_{ll})$. Summing over $k, l$ gives the inequality $\|M\|_{\mathrm{HS}}^2 \leqslant (\mathrm{Tr}\, M)^2$.

*Exercise* 7. Prove Lemma 2.

## 1.5 Partial transposition

We now investigate symmetries of $\mathrm{Sep}(\mathcal{H})$. For notational simplicity we consider only the bipartite case.

**Proposition 3** (see [1]). *The group of isometries of $\mathrm{Sep}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is generated by the following transformations*

- *Conjugation by local unitaries, of the form* $\rho \mapsto (U_1 \otimes U_2)\rho(U_1 \otimes U_2)^\dagger$, *where* $U_1$ *and* $U_2$ *are unitary transformations, respectively on* $\mathcal{H}_1$ *and* $\mathcal{H}_2$.
- *The two partial transpositions, defined on product states by* $\rho_1 \otimes \rho_2 \mapsto \rho_1^T \otimes \rho_2$ *and* $\rho_1 \otimes \rho_2 \mapsto \rho_1 \otimes \rho_2^T$, *and extended by linearity.*
- *(when* $\dim \mathcal{H}_1 = \dim \mathcal{H}_2$*) The flip operator, defined on product states by* $\rho_1 \otimes \rho_2 \mapsto \rho_2 \otimes \rho_1$ *and extended by linearity.*

We denote by $\rho^\Gamma = (\mathrm{Id} \otimes T)\rho$ the partial transposition of a state $\rho \in \mathrm{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. An explanation for the notation is that $\Gamma$ is "half" of the letter $T$ used for the usual transposition.

It is clear that for a separable state $\rho$, the operator $\rho^\Gamma$ is positive. However, this is not true for any state: indeed, the transposition map is not completely positive! A state $\rho \in \mathrm{D}(\mathcal{H})$ is said to be PPT (positive partial transpose) when $\rho^\Gamma$ is a positive operator.

Note also that the definition of partial transposition depends on a choice of basis. However, we know from Exercise 4 that the property of being PPT is basis-independent.

*Exercise* 8. Show that a pure state $|\psi\rangle\langle\psi| \in \mathrm{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is separable if and only if it is PPT.

The PPT criterion is a useful tool to prove that some states are entangled [33]. Here is an example. The space $\mathbb{C}^d \otimes \mathbb{C}^d$ can be written as the direct sum $S \oplus A$, where $S = \mathrm{span}\{x \otimes x \ : \ x \in \mathbb{C}^d\}$ is the symmetric subspace and $A = \mathrm{span}\{x \otimes y - y \otimes x \ : \ x, y \in \mathbb{C}^d\}$ is the antisymmetric subspace. Let $P_S$ and $P_A$ be the corresponding orthogonal projections. Note that $P_S = \frac{1}{2}(\mathrm{Id} + F)$ and $P_A = \frac{1}{2}(\mathrm{Id} - F)$ where $F : x \otimes y \mapsto y \otimes x$ is the flip operation. Normalize them to obtain the symmetric and antisymmetric states

$$\rho_S = \binom{d+1}{2}^{-1} P_S, \quad \rho_A = \binom{d}{2}^{-1} P_A.$$

States of the form $\rho_\alpha = \alpha\rho_S + (1-\alpha)\rho_A$ for $\alpha \in [0,1]$ are called Werner states.

*Exercise* 9. Show that $\rho_\alpha$ is non-PPT (hence entangled) for $\alpha > 1/2$. Then (harder) show that $\rho_\alpha$ is separable for $\alpha \leqslant 1/2$.

Generalizations of the PPT criterion give other criteria to prove that a state is entangled; one has to replace the transposition by another non-completely positive map. As an application of the Hahn–Banach theorem, one can prove the following: a state $\rho \in \mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$ is separable if and only if, for any positive map $\Phi : B(\mathbb{C}^d) \to B(\mathbb{C}^d)$, the operator $(\mathrm{Id} \otimes \Phi)(\rho)$ is positive [23].

In the special situation $d = 2$, it has been proved [40] that any positive map $\Phi : B(\mathbb{C}^2) \to B(\mathbb{C}^2)$ is of the form $A + B \circ T$, where $T$ is the transposition and $A, B$ are completely positive. It follows that for states on $\mathbb{C}^2 \otimes \mathbb{C}^2$, separability and PPT are equivalent properties.

There is a simple elegant argument to show that any positive map $\Phi : B(\mathbb{C}^2) \to B(\mathbb{C}^2)$ which is in addition unital (i.e., $\Phi(\mathrm{Id}) = \mathrm{Id}$) and trace-preserving must be of the form

$A + B \circ T$ for $A, B$ completely positive. Indeed, $\Phi$ maps the Bloch ball into the Bloch ball and fixes its center, so it can be thought of as a contraction on $\mathbb{R}^3$. Any contraction can be written as a convex combination of orthogonal transformations; among these rotations yield completely positive maps, while anti-rotations yield maps which becomes completely positive when composed with transposition. The hypothesis that $\Phi$ is unital can be removed by invoking Brouwer's theorem (see [4]).

# 2 Quantifying entanglement

## 2.1 Quantifying entanglement of pure bipartite states

Let $x$ be a unit vector in $\mathcal{H}_1 \otimes \mathcal{H}_2$. How much entanglement is there in the pure state $|x\rangle\langle x|$? Before answering this question, it is convenient to introduce the Schmidt decomposition of $x$ (which is simply a reformulation in the tensor language of the singular value decomposition for matrices)

$$x = \sum_{i=1}^{n} \sqrt{\lambda_i} e_i \otimes f_i, \tag{4}$$

where $(\lambda_i)_{1 \leqslant i \leqslant n}$ are positive numbers summing to 1, and $(e_i)$ (resp., $(f_i)$) an orthonormal family in $\mathcal{H}_1$ (resp., $\mathcal{H}_2$). We have $n \leqslant \min(\dim \mathcal{H}_1, \dim \mathcal{H}_2)$ and moreover $|x\rangle\langle x|$ is separable if and only if $n = 1$.

Another useful notion is the partial trace. Denote by $\mathrm{Tr}_{\mathcal{H}_2} : B(\mathcal{H}_1 \otimes \mathcal{H}_2) \to B(\mathcal{H}_1)$ the partial trace with respect to $\mathcal{H}_2$, i.e., the unique linear operator satisfying $\mathrm{Tr}_{\mathcal{H}_2}(A \otimes B) = (\mathrm{Tr}\, B)A$; in other words $\mathrm{Tr}_{\mathcal{H}_2} = \mathrm{Id} \otimes \mathrm{Tr}$. Similarly introduce $\mathrm{Tr}_{\mathcal{H}_1} = \mathrm{Tr} \otimes \mathrm{Id}$. When $x$ is given as (4), we have

$$\mathrm{Tr}_{\mathcal{H}_1} |x\rangle\langle x| = \sum_{i=1}^{n} \lambda_i |f_i\rangle\langle f_i|,$$

$$\mathrm{Tr}_{\mathcal{H}_2} |x\rangle\langle x| = \sum_{i=1}^{n} \lambda_i |e_i\rangle\langle e_i|.$$

Schmidt coefficients are eigenvalues of the so-called reduced density matrix.

We quantify the amount of entanglement present in $x$ as follows: the entropy of entanglement of a unit vector $x$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$ is defined as

$$E(x) = -\sum_{i=1}^{n} \lambda_i \log \lambda_i = S(\mathrm{Tr}_{\mathcal{H}_1} |x\rangle\langle x|),$$

where the $(\lambda_i)$ are the Schmidt coefficients as in (4), and $S(\rho) = -\mathrm{Tr}(\rho \log \rho)$ is the von Neumann entropy of a state $\rho$.

Note that on $\mathbb{C}^d \otimes \mathbb{C}^d$, the maximal value of the entropy of entanglement equals $\log d$ and is achieved for so called "maximally entangled states", i.e., of the form

$$x = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} e_i \otimes f_i$$

for orthonormal bases $(e_i)$ and $(f_i)$. In the special case $d = 2$, maximally entangled states are called Bell states.

## 2.2 Quantum channels and the LOCC paradigm

There are operational justifications for the definition of entropy of entanglement. We first introduce quantum channels: given two Hilbert spaces $\mathcal{H}^{in}$ and $\mathcal{H}^{out}$, a quantum channel $\Phi : B(\mathcal{H}^{in}) \to B(\mathcal{H}^{out})$ is a completely positive map which is also trace-preserving. In particular it maps states to states: $\Phi(\mathrm{D}(\mathcal{H}^{in})) \subset \mathrm{D}(\mathcal{H}^{out})$.

Quantum channels can be characterized via the Kraus representation: a linear map $\Phi : B(\mathcal{H}^{in}) \to B(\mathcal{H}^{out})$ is a quantum channel if and only if it acts as

$$\Phi(X) = \sum_i A_i X A_i^\dagger$$

for $X \in B(\mathcal{H}^{in})$, where $A_i : \mathcal{H}^{in} \to \mathcal{H}^{out}$ are finitely many operators satisfying $\sum A_i^\dagger A_i = \mathrm{Id}$.

Suppose now that $\mathcal{H}^{in}$ and $\mathcal{H}^{out}$ are bipartite Hilbert spaces, i.e., $\mathcal{H}^{in} = \mathcal{H}_1^{in} \otimes \mathcal{H}_2^{in}$ and $\mathcal{H}^{out} = \mathcal{H}_1^{out} \otimes \mathcal{H}_2^{out}$. Given quantum channels $\Phi_1 : B(\mathcal{H}_1^{in}) \to B(\mathcal{H}_1^{out})$ and $\Phi_2 : B(\mathcal{H}_2^{in}) \to B(\mathcal{H}_2^{out})$, we may consider the product quantum channel $\Phi_1 \otimes \Phi_2 : B(\mathcal{H}^{in}) \to B(\mathcal{H}^{out})$.

Mimicking the definition for states, a quantum channel $\Phi : \mathcal{H}^{in} \to \mathcal{H}^{out}$ is said to be separable is it can be written as

$$\Phi(X) = \sum_i (A_i^1 \otimes A_i^2) X (A_i^1 \otimes A_i^2)^\dagger)$$

for some $A_i^1 : \mathcal{H}_1^{in} \to \mathcal{H}_1^{out}$ and $A_i^2 : \mathcal{H}_2^{in} \to \mathcal{H}_2^{out}$. Product channels are examples of product channels. However, the relevant class for Quantum Information Theory is the related class of LOCC (Local Operations and Classical Communications) channels (see [25, 13] for extensive discussions). To avoid technicalities, we do not define it precisely but simply say that

$$\mathrm{conv}\{\text{product channels}\} \subset \{\text{LOCC channels}\} \subset \{\text{separable channels}\}.$$

We explain now a basic result from quantum Shannon theory. Denote by $\psi$ a Bell state (a maximally entangled state on $\mathbb{C}^2 \otimes \mathbb{C}^2$). Given a unit vector $x \in \mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, we may define its *distillable entanglement* $E_D(x)$ to be the supremum of all $R > 0$ such that, for

any $n > 0$, there exists a LOCC quantum channel $\Phi_n : B(\mathcal{H}^{\otimes n}) \to B((\mathbb{C}^2 \otimes \mathbb{C}^2)^{\otimes \lfloor Rn \rfloor})$ with the property that

$$\lim_{n\to\infty} \|\Phi_n(|x\rangle\langle x|^{\otimes n}) - |\psi\rangle\langle\psi|^{\otimes \lfloor Rn \rfloor}\|_1 = 0.$$

This definition may require some effort to grasp: what is meant is that $E_D(x)$ the largest rate at with the state $|x\rangle\langle x|$ can be transformed into the state $|\psi\rangle\langle\psi|$ via LOCC transformations, with an error vanishing in the limit of many copies. It is known that the number $E_D(x)$ is unchanged if "LOCC quantum channel" is replaced by "separable channel" in its definition (see, e.g., [20]).

The reverse operation would be to transform $|\psi\rangle\langle\psi|$ into $|x\rangle\langle x|$. This leads to the definition of the *entanglement cost* of $x$, $E_C(x)$, defined to be the infimum of all $R$ such that, for any $n > 0$, there exists a LOCC quantum channel $\Phi_n : B((\mathbb{C}^2 \otimes \mathbb{C}^2)^{\otimes \lfloor Rn \rfloor}) \to B(\mathcal{H}^{\otimes n})$ with the property that

$$\lim_{n\to\infty} \|\Phi_n(|\psi\rangle\langle\psi|^{\otimes \lfloor Rn \rfloor}) - |x\rangle\langle x|^{\otimes n}\|_1 = 0.$$

It turns out that the distillable entanglement equals the entanglement cost: manipulation of pure state entanglement is asymptotically reversible.

**Theorem 4** (see [9]). *For any bipartite pure state $x$, we have $E_D(x) = E_C(x) = E(x)$.*

The appearance of the entropy of entanglement is related to the following classical fact about "typical sequences": if $x$ has Schmidt coefficients $\lambda_1, \ldots, \lambda_d$, then the Schmidt coefficients of $x^{\otimes n}$ are products $\lambda_{i_1} \cdots \lambda_{i_n}$, for all $(i_1, \ldots, i_n) \in \{1, \ldots, d\}^n$. It follows from the law of large numbers that for large $n$, most of the mass is concentrated on Schmidt coefficients with value of order $\lambda$, where $\log \lambda = n \sum_{i=1}^d \lambda_i \log \lambda_i$.

### 2.3 The case of multipartite pure states

For a vector $x$ in a multipartite Hilbert space $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$, there is no analogue of the Schmidt decomposition when $k > 2$. However, we may define the analogue of the largest Schmidt coefficient by taking the maximal scalar product with a unit product vector. This coincides with the injective norm on the tensor product of Hilbert spaces, which is dual to the projective norm introduced in (2)

$$\begin{aligned} \lambda(x) &= \max\{|\langle x, x_1 \otimes \cdots \otimes x_k\rangle| \ : \ |x_1| = \cdots = |x_k| = 1\} && (5) \\ &= \max\{|\langle x, y\rangle| \ : \ y \in B_{\mathcal{H}_1} \widehat{\otimes} \cdots \widehat{\otimes} B_{\mathcal{H}_k}\} && (6) \\ &= \|x\|_{\mathcal{H}_1 \check{\otimes} \cdots \check{\otimes} \mathcal{H}_k} && (7) \end{aligned}$$

In order to recover a quantity that scales like the entropy of entanglement, one considers $E_\infty(x) := -2 \log \lambda(x)$. Natural questions are: how small can be $\lambda(x)$? what are the most entangled vectors? The minimal value of $\lambda(x)$ over unit vectors $x$ is the inradius of $B_{\mathcal{H}_1} \widehat{\otimes} \cdots \widehat{\otimes} B_{\mathcal{H}_k}$.

In the bipartite case, when $x \in \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$, one has $\lambda(x) \geqslant 1/\sqrt{\min(d_1, d_2)}$. By induction, estimates follow also in the multipartite case. For simplicity, we consider the case of $k$ qubits: $\mathcal{H}_1 = \mathcal{H}_2 = \cdots = \mathcal{H}_k = \mathbb{C}^2$. For any unit vector $x \in (\mathbb{C}^2)^{\otimes k}$, we have $\lambda(x) \geqslant 2^{-(k-1)/2}$, or $E_\infty(x) \leqslant k - 1$. How sharp is this estimate is unknown.

**Problem 5.** *Can we find a constant $C$, and for any $k$ a unit vector $x \in (\mathbb{C}^2)^{\otimes k}$ such that $\mathbf{E}_\infty(x) \geqslant k - C$ ?*

Curiously, in the real case, there are unit vectors in $(\mathbb{R}^2)^{\otimes k}$ satisfying $\lambda(x) = 2^{-(k-1)/2}$. Equivalently, there is a $k$-linear map $\Phi : (\mathbb{R}^2)^k \to \mathbb{R}$ such that $\Phi(x_1, \cdots, x_k) \leqslant |x_1| \cdots |x_k|$ and the "Hilbert–Schmidt" norm of $\Phi$ equals $2^{(k-1)/2}$

$$\|\Phi\|_{\mathrm{HS}} := \left( \sum_{(i_1, \cdots, i_k) \in \{1,2\}^k} \Phi(e_{i_1}, \cdots, e_{i_k})^2 \right)^{1/2} = 2^{(k-1)/2},$$

where $(e_1, e_2)$ is the canonical basis of $\mathbb{R}^2$. Indeed, if $\theta$ is the canonical identification between $\mathbb{R}^2$ and $\mathbb{C}$, we may define $\Phi$ as

$$\Phi(x_1, \ldots, x_k) = \mathrm{Re} \left( \prod_{i=1}^{k} \theta(x_j) \right).$$

## 2.4 Random multipartite states are very entangled

We are going to prove that most vectors in the unit sphere of $(\mathbb{C}^2)^{\otimes k}$ are very entangled, although they are not entangled enough to provide a positive answer to Problem 5. We will use some standard machinery which we now review. We use concentration of measure in the following form, which is called Lévy's lemma in quantum information literature. It asserts that the fluctuations of 1-Lipschitz functions on a $n$-dimensional sphere are of order $O(1/\sqrt{n})$.

**Lemma 6** (see [28, 27, 4])**.** *Let $f : S^{n-1} \to \mathbb{R}$ be a 1-Lipschitz function and choose $x \in S^{n-1}$ randomly according to the uniform measure $\sigma$. Then, for any $t > 0$,*

$$\mathbf{P}(|f(x) - \mathbf{E}\, f(x)| > t) \leqslant 2 \exp(-(n-1)t^2/2).$$

There are two natural distances on the sphere $S^{n-1}$: the geodesic distance and the restriction to $S^{n-1}$ of the Euclidean distance on $\mathbb{R}^n$. Lemma 6 is true for both distances.

Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, and denote by $|\cdot|$ the Euclidean norm. Denote also $\gamma_n$ the standard Gaussian measure on $\mathbb{R}^n$. We may write $x \in \mathbb{R}^n$ as $x = |x| \frac{x}{|x|}$ and use polar integration to obtain

$$\int_{\mathbb{R}^n} \|x\| \, \mathrm{d}\gamma_n = \kappa_n \int_{S^{n-1}} \|u\| \, \mathrm{d}\sigma(u), \tag{8}$$

where

$$\kappa_n = \left( \int_{\mathbb{R}^n} |x| \, \mathrm{d}\gamma_n \right).$$

It is easily checked that $\sqrt{n-1} < \kappa_n < \sqrt{n}$. If we consider instead a norm on $\mathbb{C}^n$, and denote by $\gamma_n^{\mathbb{C}}$ the standard Gaussian measure on $\mathbb{C}^n$ (i.e., such that $\mathrm{Re}\langle \cdot, \theta \rangle$ has distribution $N(0, 1/2)$ for any $\theta \in S_{\mathbb{C}^n}$), the formula becomes

$$\int_{\mathbb{C}^n} \|x\| \, \mathrm{d}\gamma_n^{\mathbb{C}} = \frac{\kappa_{2n}}{\sqrt{2}} \int_{S_{\mathbb{C}^n}} \|u\| \, \mathrm{d}\sigma(u) \tag{9}$$

We also need a version of the union bound for maximum of Gaussian variables.

**Lemma 7.** *Let $X_1, \ldots, X_N$ be random variables and assume that $X_i$ has distribution $N(0, \sigma_i^2)$ with $\sigma_i^2 \leqslant 1$. Then*

$$\mathbf{E} \max(X_1, \ldots, X_N) \leqslant \sqrt{2 \log N}.$$

*Proof.* For any $\beta > 0$, we compute $\mathbf{E} \exp(\beta X_i) = \exp(\sigma_i^2 \beta^2 / 2) \leqslant \exp(\beta^2/2)$ and may write

$$\begin{aligned} \mathbf{E} \max(X_1, \ldots, X_N) &\leqslant \frac{1}{\beta} \mathbf{E} \log \sum_{i=1}^N \exp(\beta X_i) \\ &\leqslant \frac{1}{\beta} \log \sum_{i=1}^N \mathbf{E} \exp(\beta X_i) \\ &\leqslant \frac{1}{\beta} \left( \log N + \frac{\beta^2}{2} \right). \end{aligned}$$

We then choose the optimal value $\beta = \sqrt{2 \log N}$. This proof is due to Talagrand. □

We are going to prove the following proposition, as a standard application of concentration of measure and $\varepsilon$-nets argument (the quantity $\lambda(x)$ was defined in (5)).

**Proposition 8** (see [16, 11])**.** *Let $x$ be a unit vector in $(\mathbb{C}^2)^{\otimes k}$ chosen at random with respect to the uniform measure on the sphere. Then, with large probability*

$$c \frac{\sqrt{k \log k}}{2^{k/2}} \leqslant \lambda(x) \leqslant C \frac{\sqrt{k \log k}}{2^{k/2}},$$

*where $c > 0$ and $C$ denote numerical constants.*

Equivalently, for typical vectors $x$, one has $E_\infty(x) = k - \log k - \log \log k + O(1)$.

*Proof.* Since the function $\lambda$ is 1-Lipschitz on $S_{(\mathbb{C}^2)^{\otimes k}}$, Proposition 8 follows easily from Lemma 6 once we prove that

$$c\frac{\sqrt{k\log k}}{2^{k/2}} \leqslant \mathbf{E}\,\lambda(x) \leqslant C\frac{\sqrt{k\log k}}{2^{k/2}}.$$

We only prove the upper bound (for the lower bound, see Exercise 10). We take a $\varepsilon$-net $\mathcal{N}$ in $S_{\mathbb{C}^2} \simeq S^3$. Since $S^3$ has dimension 3, we can choose such a net with $\operatorname{card}\mathcal{N} \leqslant C/\varepsilon^3$ for some constant $C$. A simple geometric argument shows that

$$\operatorname{conv}\mathcal{N} \supset \left(1-\frac{\varepsilon^2}{2}\right) B_{\mathbb{C}^2}$$

and therefore

$$\operatorname{conv}\mathcal{N}^{\otimes k} \supset \left(1-\frac{\varepsilon^2}{2}\right)^k B_{\mathbb{C}^2}\hat{\otimes}\cdots\hat{\otimes}B_{\mathbb{C}^2}.$$

If we choose $\varepsilon = 1/\sqrt{k}$, then $(1-\varepsilon^2/2)^k \geqslant c$ for some constant $c > 0$. It follows that

$$\lambda(x) \leqslant c^{-1}\max\{|\langle x, y\rangle| \ : \ y \in \mathcal{N}^{\otimes k}\}.$$

Using (9), we have

$$\int_{S_{(\mathbb{C}^2)^{\otimes k}}} \lambda(x)\,\mathrm{d}\sigma(x) = \frac{1}{\kappa_{2^k}^{\mathbb{C}}}\int_{(\mathbb{C}^2)^{\otimes k}} \lambda(x)\mathrm{d}\gamma_{2^k}^{\mathbb{C}} \leqslant \frac{C}{2^{k/2}}\sqrt{2\log\operatorname{card}(\mathcal{N}^{\otimes k})} \leqslant \frac{C\sqrt{k\log k}}{2^{k/2}}. \quad \square$$

The proof of the lower bound is based on the *Sudakov minoration principle*: if $X_1, \ldots, X_N$ are jointly Gaussian (real or complex) vectors satisfying $\mathbf{E}\,|X_i - X_j|^2 \geqslant \varepsilon^2$ for some $\varepsilon > 0$, then $\mathbf{E}\max(X_1, \ldots, X_N) \geqslant c\varepsilon\sqrt{\log N}$.

*Exercise* 10. Produce enough $\varepsilon$-separated points in $(B_{\mathbb{C}^2})^{\otimes k}$ to show the lower bound $\mathbf{E}\,\lambda(x) \geqslant c\frac{\sqrt{k\log k}}{2^{k/2}}$ using Sudakov minoration principle.

## 2.5 Quantifying entanglement of mixed states

A comprehensive survey of the many ways to quantifiy the entanglement of mixed states is [35]. We only consider a few of them.

The definition of the distillable entanglement and of the entanglement cost can be extended verbatim to the case of a mixed state $\rho \in \mathrm{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. We repeat them informally (recall that $\psi$ denotes a Bell state)

- The entanglement cost $E_C(\rho)$ is the best (i.e., smallest) rate $R$ such that we can, via LOCC channels, transform $|\psi\rangle\langle\psi|^{\otimes Rn}$ into $\rho^{\otimes n}$ with vanishing error as $n \to \infty$.
- The distillable entanglement $E_D(\rho)$ is the best (i.e., largest) rate $R$ such that we can, via LOCC channels, transform $\rho^{\otimes n}$ into $|\psi\rangle\langle\psi|^{\otimes Rn}$ with vanishing error as $n \to \infty$.

It is known that creating entanglement has always a non-zero cost.

**Theorem 9** ([45]). *A state $\rho \in \mathrm{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is entangled if and only if $E_C(\rho) > 0$.*

On the other hand, it is easy to check that if $\Phi$ is a separable channel (which includes the case of LOCC channels), and $\rho$ a PPT state, then $\Phi(\rho)$ is a PPT state. Since Bell states are non-PPT, it follows that $E_D(\rho) = 0$ for any PPT entangled state (such states are called bound entangled).

The distillability problem is the following: does there exist a state $\rho \in \mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$ such that $E_D(\rho) = 0$ while $\rho$ is not PPT ? This is already not known for $d = 3$. We also point out that $E_D(\rho) > 0$ if and only if there exists an integer $n$ and operators $A, B : (\mathbb{C}^d)^{\otimes n} \to \mathbb{C}^2$ such that the $(A \otimes B)\rho(A \otimes B)^\dagger$ is non-PPT. It is also known that if there is a non-PPT state $\rho$ with $E_D(\rho) = 0$, then there is a non-PPT Werner state $\rho_\alpha$ with $E_D(\rho_\alpha) = 0$. For more information on the distillability problem see [24, 14, 4].

We now mention the connexion between the entanglement cost and the entanglement of formation. Let $\rho \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Any decomposition of $\rho$ as a mixture of pure states

$$\rho = \sum \lambda_i |\psi\rangle\langle\psi| \tag{10}$$

yields a protocol to generate $\rho$ from Bell states at a rate $R = \sum \lambda_i E(\psi_i)$. The entanglement of formation $E_F(\rho)$ is the infimum of these rates over decompositions (10)

$$E_F(\rho) = \inf \left\{ \sum \lambda_i E(\psi_i) \ : \ \rho = \sum \lambda_i |\psi_i\rangle\langle\psi_i| \right\}.$$

In other words, $E_F$ is the largest convex function on $\mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $E_F(|\psi\rangle\langle\psi|) = E(\psi)$.

*Exercise* 11. Prove that $E_F(\rho) = 0$ if and only if $\rho$ is separable.

The previous definition yields the inequality $E_C(\rho) \leqslant E_F(\rho)$ and actually even $E_C(\rho) \leqslant \frac{1}{n} E_F(\rho)$ (indeed, the way the entanglement cost is defined gives automatically the additivity property $E_C(\rho^{\otimes n}) = n E_C(\rho)$). This inequality was proved to be sharp in the limit $n \to \infty$

**Theorem 10** (Horodecki–Hayden–Terhal, [22]). *For any bipartite state $\rho$,*

$$E_C(\rho) = \lim_{n \to \infty} \frac{1}{n} E_F(\rho^{\otimes n}).$$

The entropy of entanglement is additive on product vectors: $E(x \otimes y) = E(x) + E(y)$. It follows that the entanglement of formation is subadditive:

$$E_F(\rho \otimes \sigma) \leqslant E_F(\rho) + E_F(\sigma). \tag{11}$$

For a long time the entanglement of formation was conjectured to be additive, i.e., that there is equality in (11). An immediate corollary of Theorem 10 would have been the equality between entanglement of formation and entanglement cost. However, this conjecture turned out to be false, as proved by Hastings [19].

# 3 Non-additivity phenomenon and Dvoretzky's theorem

## 3.1 Minimum output entropy

The conjecture that entanglement of formation is additive mentioned at the end of the previous section was known to be equivalent (after work by Shor [37]) to the additivity of the minimal output entropy: for any quantum channels $\Phi_1, \Phi_2$,

$$S_{\min}(\Phi_1 \otimes \Phi_2) \stackrel{?}{=} S_{\min}(\Phi_1) + S_{\min}(\Phi_2)$$

where, for a quantum channel $\Phi : B(\mathcal{H}^{in}) \to B(\mathcal{H}^{out})$,

$$S_{\min}(\Phi) = \min\left\{S(\Phi(\rho)) \ : \ \rho \in \mathrm{D}(\mathcal{H}^{in})\right\}. \tag{12}$$

By concavity of the von Neumann entropy, the minimum in (12) can be restricted to pure states.

It is convenient to use the Stinespring representation of quantum channels. Any quantum channel $\Phi : B(\mathcal{H}^{in}) \to B(\mathcal{H}^{out})$ can be represented as

$$\Phi(\rho) = \mathrm{Tr}_{\mathcal{H}^e} U\rho U^\dagger \tag{13}$$

where $\mathcal{H}^e$ is an auxiliary Hilbert space (the letter $e$ stands for "environment"), and $U : \mathcal{H}^{in} \to \mathcal{H}^{out} \otimes \mathcal{H}^e$ is an isometric embedding. Denote $V \subset \mathcal{H}^{out} \otimes \mathcal{H}^e$ the range of $U$. When $\Phi$ is given as in (13), its minimal output entropy is

$$S_{\min}(\Phi) = \min_{x \in V} E(x).$$

The existence of channels with large minimum output entropy is therefore connected to the existence of (very) entangled subspaces, i.e., subspaces $V \subset \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ such that any unit vector $x \in V$ is (very) entangled. We search for such subspaces of dimension as large as possible.

## 3.2 Entangled subspace: qualitative problem

This calls for a qualitative version of the problem: how large can be $\dim V$, where $V \subset \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ is a subspace which does not contain any nonzero product vector ? This can be solved by elementary algebraic geometry. Denote by

$$\mathrm{Seg} := \{x \otimes y \ : \ x \in \mathbb{C}^{d_1}, y \in \mathbb{C}^{d_2}\} \subset \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$$

the set of product vectors (which is also called the Segré variety). As projective varieties, Seg has dimension $d_1 + d_2 - 2$ while $V$ has dimension $\dim V - 1$ ($\dim V$ denoting the dimension of $V$ as a linear space). By the projective intersection theorem (see [18]), the intersection $V \cap \mathrm{Seg}$ is non-empty whenever

$$(\dim V - 1) + (d_1 + d_2 - 2) \geqslant d_1 d_2 - 1 \iff \dim V \geqslant d_1 d_2 - (d_1 + d_2) + 2.$$

Conversely, it is not hard to show that a randomly chosen $V$ with $\dim V < d_1 d_2 - (d_1 + d_2) + 2$ intersects Seg with probability 1.

*Exercise* 12 (see [32, 41]). Prove the last statement using your favorite notion of dimension (a possibility is to use the Minkowski dimension, i.e., count how many balls of radius $\varepsilon$ are required to cover a given set, as $\varepsilon \to 0$).

Here and in what follows, whenever we say "random subspace of dimension $k$", it is tacitly understood that the subspace is chosen with respect to the Haar measure on the corresponding Grassmann manifold. Equivalently, it can be defined as the subspace spanned by $k$ independent random vectors uniformly distributed on the sphere.

We now turn to the quantitative version of problem: we are interested in finding subspaces $V \subset \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ on which the function $E$ is not only nonzero, but large. This problem enters within the realm of Dvoretzky-like theorems, which state that a Lipschitz function on a high-dimensional sphere is almost constant on large-dimensional subspaces. We mention that an alternative route is possible from that point to obtain counterexamples where the minimum output entropy is non-additive: following [7, 8], one can work directly on the limit object using free probability and prove the following result. Fix $t \in (0, 1)$ and consider a random subspace $E \subset \mathbb{C}^k \otimes \mathbb{C}^n$ of dimension $\lfloor tkn \rfloor$. As $k$ is fixed and $n$ tends to infinity, the set of all possible Schmidt coefficients of unit vectors from $E$ (which is a subset of $\mathbb{R}^k$) has a deterministic limit.

### 3.3 Dvoretzky's theorem

We already mentioned Lévy's lemma (Lemma 6): Lipschitz functions on the sphere are concentrated around their mean. It is useful to introduce the more flexible notion of central value: a central value for a random variable $X$ is either its mean, or any number $t$ such that $\mathbf{P}(X \leqslant t) \geqslant \frac{1}{4}$ and $\mathbf{P}(X \leqslant t) \leqslant \frac{1}{4}$. Lévy's lemma has a variant for central values: if $f : S^{n-1} \to \mathbb{R}$ is 1-Lipschitz, then for any $t > 0$,

$$\sigma\left(\{|f - \mu| > t\}\right) \leqslant C \exp(-cnt^2)$$

for some absolute constants $C, c$.

*Exercise* 13. Deduce the "central value" version of Lévy's lemma from the "median" version (the latter is an immediate consequence of isoperimetry on of the sphere).

We can now state Dvoretzky's theorem for Lipschitz functions.

**Theorem 11.** *Let $f : S^{n-1} \to \mathbb{R}$ be a 1-Lipschitz function, and $\mu$ a central value for $f$. Let $E \subset \mathbb{R}^n$ a random subspace of dimension $k$. Then, provided $k \leqslant c(\varepsilon)n$, with large probability,*

$$\sup_{x \in S^{n-1} \cap E} |f(x) - \mu| \leqslant \varepsilon.$$

*Proof.* The following argument is essentially due to Milman [29] and ultimately based on a "union bound" argument. Fix any subspace $E_0 \subset \mathbb{R}^n$ with dimension $k$. The random subspace $E$ can be realized as $E = O(E_0)$, where $O$ is a random Haar-distributed element of $\mathsf{O}(n)$. Consider also a $\varepsilon/2$-net $\mathcal{N}$ in $S^{n-1} \cap E_0$. Such a net can be chosen with $\operatorname{card} \mathcal{N} \leqslant (C/\varepsilon)^k$. Since $f$ is 1-Lipschitz, it is enough to prove that $|f - \mu| \leqslant \varepsilon/2$ on $O(\mathcal{N})$ with large probability. For any $x \in \mathcal{N}$, the vector $O(x)$ is uniformly distributed on the sphere, and therefore we have

$$\begin{aligned} \mathbf{P}(\exists x \in \mathcal{N} \ : \ |f(O(x)) - \mu| > \varepsilon/2) &\leqslant \operatorname{card}(\mathcal{N}) \sigma(\{|f - \mu| > \varepsilon/2\}) \quad (14) \\ &\leqslant (C/\varepsilon)^k C \exp(cn\varepsilon^2). \quad (15) \end{aligned}$$

The right-hand side of (15) is (much) smaller than 1 provided $k \log(1/\varepsilon) \leqslant cn\varepsilon^2$, or $k \leqslant c(\varepsilon)n$ where $c(\varepsilon) = c\varepsilon^2/\log(1/\varepsilon)$. □

This argument can be improved to obtain the dependence $c(\varepsilon) = c\varepsilon^2$ by using a chaining argument à la Dudley. We also need the complex analogue of Theorem 11.

**Theorem 12** (see [36, 5])**.** *Let $f : S_{\mathbb{C}^n} \to \mathbb{R}$ be a 1-Lipschitz and circled (i.e., such that $f(\alpha x) = f(x)$ for $x \in S_{\mathbb{C}^n}$ and $\alpha \in \mathbb{C}$ with $|\alpha| = 1$) function, and $\mu$ a central value for $f$. Let $E \subset \mathbb{C}^n$ a random subspace of dimension $k$. Then, provided $k \leqslant c\varepsilon^2 n$, with large probability,*

$$\sup_{x \in S_{\mathbb{C}^n} \cap E} |f(x) - \mu| \leqslant \varepsilon.$$

### 3.4 Counterexample to additivity

We now describe how to obtain from Dvoretzky's theorem a pair of channels for which the minimal output entropy is not additive. This result was initially obtained by Hastings [19] and considered as a major breakthrough in quantum information theory. The use of Dvoretzky's theorem allows for a more conceptual approach; we follow the argument from [5].

We consider for $i \in \{1, 2\}$ an random isometry $U_i : \mathcal{H}_i^{in} \to \mathcal{H}_i^{out} \otimes \mathcal{H}_i^e$ and $\Phi_i : B(\mathcal{H}_i^{in}) \to B(\mathcal{H}_i^{out})$ the corresponding channel defined as (13). What only matters is the range of $U_i$, which is a random subspace $V_i \subset \mathcal{H}_i^{out} \otimes \mathcal{H}_i^e$. We are going to adjust later the dimensions $d_{in} = \dim \mathcal{H}_i^{in}, d_{out} = \dim \mathcal{H}_i^{out}$ and $d_e = \dim \mathcal{H}_i^e$ in order to obtain $S_{\min}(\Phi_1 \otimes \Phi_2) < S_{\min}(\Phi_1) + S_{\min}(\Phi_2)$, or equivalently (the minima being restricted to unit vectors)

$$\min_{x \in V_1 \otimes V_2} E(x) < \min_{x_1 \in V_1} E(x_1) + \min_{x_2 \in V_2} E(x_2). \quad (16)$$

We use a trick to ensure that the left-hand side in (16) is small: take $\mathcal{H}_1^{in} = \mathcal{H}_2^{in} = \mathcal{H}^{in}$, $\mathcal{H}_1^{out} = \mathcal{H}_2^{out} = \mathcal{H}^{out}$, $\mathcal{H}_1^e = \mathcal{H}_2^e = \mathcal{H}^e$, $U_2 = \overline{U_1}$ (the entry-wise complex conjugation of $U_1$, with respect to a fixed basis) and $x$ to be the maximally entangled state (with respect to the same basis) in $V_1 \otimes \overline{V_1}$.

**Lemma 13.** *Let $V \subset \mathbb{C}^{d_{out}} \otimes \mathbb{C}^{d_e}$ a subspace with dimension $d_{in}$. Then $V \otimes \overline{V}$ contains a unit vector whose largest Schmidt coefficient is greater that $d_{in}/d_{out}d_e$. Schmidt coefficients are computed with respect to the bipartition $(\mathbb{C}^{d_{out}})^{\otimes 2}$ vs $(\mathbb{C}^{d_e})^{\otimes 2}$.*

Equivalently and perhaps more transparently, the lemma can be stated using the language of matrices: if $V \subset \mathsf{M}_{d_{out},d_e}$ is a subspace of dimension $d_{in}$, then $V \otimes \overline{V}$ contains a matrix $A$ with $\|A\|_{\mathrm{HS}} = 1$ and $\|A\|_\infty \geqslant \sqrt{d_{in}/d_{out}d_e}$.

*Proof.* We prove the matrix version. Let $(e_j)$ be the canonical basis of $\mathbb{C}^{d_{out}}$, $(e_k)$ the canonical basis of $\mathbb{C}^{d_e}$. Let $(A_1, \ldots, A_{d_{in}})$ an orthonormal basis for $V$, with respect to the Hilbert–Schmidt scalar product. Consider

$$A = \frac{1}{\sqrt{d_{in}}} \sum_{i=1}^{d_{in}} A_i \otimes \overline{A_i},$$

$$\phi = \frac{1}{\sqrt{d_{out}}} \sum_{j=1}^{d_{out}} e_j \otimes e_j,$$

$$\psi = \frac{1}{\sqrt{d_e}} \sum_{k=1}^{d_e} f_k \otimes f_k.$$

We have $\|A\|_{HS} = 1$ and $|\phi| = |\psi| = 1$. We compute

$$\begin{aligned}
\langle \phi | A | \psi \rangle &= \frac{1}{\sqrt{d_{in}d_{out}d_e}} \sum_{i,j,k} \langle e_j \otimes e_j | A_i \otimes A_i | f_k \otimes f_k \rangle \\
&= \frac{1}{\sqrt{d_{in}d_{out}d_e}} \sum_{i,j,l} |\langle e_j | A_i | e_k \rangle|^2 \\
&= \frac{1}{\sqrt{d_{in}d_{out}d_e}} \sum_{i} \|A_i\|_{HS}^2 \\
&= \sqrt{\frac{d_{in}}{d_{out}d_e}}
\end{aligned}$$

as needed. □

In order to obtain a counterexample, we consider the following range: $d_{out} = k$, $d_e = k^2$ and $d_{in} = ck^2$ for some fixed constant $c$, and take $k \to \infty$. We know from basic random matrices considerations that typically, Schmidt coefficients of a single random unit vector $x \in \mathbb{C}^k \otimes \mathbb{C}^{k^2}$ are or order $\frac{1}{k}\left(1 + O(\sqrt{k/k^2})\right)$, so that $E(x) = \log(k) - C/k$ (there is an explicit formula for the mean of $E$, see, e.g., [15]). It turns out that this estimate holds uniformly over subspaces of large dimension

**Claim 14.** *There exist constants $c, C' > 0$ such that a random subspace $V \subset \mathbb{C}^k \otimes \mathbb{C}^{k^2}$ of dimension $ck^2$ satisfies*

$$\inf_{x \in V, |x|=1} E(x) \geqslant \log k - \frac{C'}{k}$$

*with large probability.*

This is enough to obtain a counterexample: indeed, for $V$ as in Claim 14, we have

$$S_{\min}(\Phi) = S_{\min}(\overline{\Phi}) \geqslant \log k - \frac{C'}{k}$$

whereas Lemma 13 gives a state $\psi$ such that $(\Phi \otimes \overline{\Phi})(|\psi\rangle\langle\psi|)$ has one eigenvalue larger than $1/k$. A simple computation using the concavity of the von Neumann entropy allows to deduce that

$$S_{\min}(\Phi \otimes \overline{\Phi}) \leqslant \log(k^2) - \frac{c \log k}{k},$$

and therefore $S_{\min}(\Phi \otimes \overline{\Phi}) < S_{\min}(\Phi) + S_{\min}(\overline{\Phi})$ for $k$ large enough. This is really a high-dimension phenomenon and the proof gives a poor estimate for the smallest dimensions in which a counterexample exists. In a slightly different model from [8] (which is based on the limit object) it is proved that counterexamples exist provided $k \geqslant 183$.

### 3.5 Very entangled subspaces

It remains to deduce the Claim 14 from Dvoretzky's theorem. A direct application to the function $E$ fails. The function $E$ can be shown to have Lipschitz constant $C \log k$ on $S_{\mathbb{C}^k \otimes \mathbb{C}^{k^2}}$, but this is not good enough.

*Exercise* 14. When $k \leqslant l$, show that the Lipschitz constant of $E$ on $S_{\mathbb{C}^k \otimes \mathbb{C}^l}$ is smaller than $C \log k$ and larger than $c \log k$, for some absolute constants $C, c$.

A better idea is to use a approximation of $E$ for states close to being maximally entangled

$$E(x) \geqslant \log k - k \left\| \mathrm{Tr}_{\mathbb{C}^{k^2}} |x\rangle\langle x| - \frac{\mathrm{Id}}{k} \right\|_{\mathrm{HS}}^2. \tag{17}$$

*Exercise* 15. Prove (17).

As we already mentioned, the eigenvalues of $\mathrm{Tr}_{\mathbb{C}^{k^2}} |x\rangle\langle x|$ (=the Schmidt coefficients of $x$) are of order $1/k + O(1/k^{3/2})$. If we define

$$g(x) = \left\| \mathrm{Tr}_{\mathbb{C}^{k^2}} |x\rangle\langle x| - \frac{\mathrm{Id}}{k} \right\|_{\mathrm{HS}},$$

this shows that $g$ is typically of order $1/k$ (as required), and we need to show that this holds uniformly over a large subspace. At this point it is more convenient to switch to

the matrix formalism. The function $g$ becomes a function defined on the Hilbert–Schmidt sphere as $g(X) = \|XX^\dagger - \frac{\mathrm{Id}}{k}\|_{\mathrm{HS}}$. We now use another idea: although the function $g$ is globally 2-Lipschitz, this can be improved via the inequality

$$|g(X) - g(Y)| \leqslant \|XX^\dagger - YY^\dagger\|_{\mathrm{HS}} \leqslant \|X\|_\infty \|X^\dagger - Y^\dagger\|_{\mathrm{HS}} + \|Y\|_\infty \|X - Y\|_{\mathrm{HS}}$$

which shows that $\|\cdot\|_\infty$ has Lipschitz constant $6/\sqrt{k}$ when restricted to the subset $\Omega = \{\|\cdot\|_\infty \leqslant 3/\sqrt{k}\}$ of the Hilbert–Schmidt sphere.

We use the following trick: let $\tilde{g}$ be a $6/\sqrt{k}$-Lipschitz extension of $g_{|\Omega}$ to the whole Hilbert–Schmidt sphere. We use Dvoretzky's theorem (Theorem 12) twice to conclude that, for a typical subspace $E \subset \mathbb{C}^k \otimes \mathbb{C}^{k^2}$ of dimension $ck^2$,

- $E \subset \Omega$,
- $\tilde{g} = O(1/k)$ on $E$.

This completes the proof of Claim 14. Note that the median of $g$ is a central value for $\tilde{g}$, and vice versa.

# 4 Random states in high dimension

## 4.1 Random induced states

Let $\mathcal{H}$ be a (finite-dimensional, complex) Hilbert space. How to choose a state on $\mathcal{H}$ at random in a natural way? What is clear is how to pick a pure state at random: simply take $|\psi\rangle\langle\psi|$ with $\psi$ uniformly distributed on the sphere. But what about mixed states?

There is an elegant and physically relevant approach which is based on the fact that any mixed state can be seen as the partial trace of a pure state over some extra Hilbert space. Indeed, if $\rho \in \mathrm{D}(\mathbb{C}^n)$ has the form

$$\rho = \sum_{i=1}^n \lambda_i |e_i\rangle\langle e_i|$$

for some orthonormal basis $(e_i)$ of $\mathbb{C}^n$, then $\rho = \mathrm{Tr}_{\mathcal{H}^e} |\psi\rangle\langle\psi|$ for

$$\psi = \sum_{i=1}^n \sqrt{\lambda_i} e_i \otimes e_i$$

and $\mathcal{H}^e = \mathbb{C}^n$. Therefore one can generate random mixed states as partial traces of random pure states. This leads to a 2-parameters family of probability measures. Given two integers $n, s$, denote by $\mu_{n,s}$ the distribution of $\mathrm{Tr}_{\mathbb{C}^s} |\psi\rangle\langle\psi|$, where $\psi$ is a unit vector with uniform distribution on the sphere $S_{\mathbb{C}^n \otimes \mathbb{C}^s}$; $\mu_{n,s}$ is a probability measure on $\mathrm{D}(\mathbb{C}^n)$. States with distribution $\mu_{n,s}$ are called random induced states and were introduced in [46].

An alternative description is as follows: if $M$ is uniformly distributed in the Hilbert–Schmidt sphere of $n \times s$ matrices, then $\mu_{n,s}$ is the distribution of $MM^\dagger$. Still another description is to consider a $n \times s$ matrix $G$ with independent entries having a standard complex distribution. Then $G/\|G\|_{\mathrm{HS}}$ is uniformly distributed on the Hilbert–Schmidt unit sphere, and therefore the matrix

$$\frac{GG^\dagger}{\operatorname{Tr} GG^\dagger} \tag{18}$$

has distribution $\mu_{n,s}$. This approach is appealing since the matrix $GG^\dagger$ is well-known in random matrix theory under the name of a Wishart matrix (although the real version of Wishart matrices are more frequently encountered) and it allows to transfer directly result from random matrices to random induced states.

In this model, the parameter $s$ has a physical interpretation: it is the dimension of the environment, or the number of non-accessible parameters. When $s \geqslant n$, the measure $\mu_{n,s}$ has a density with respect to the Lebesgue measure given by the formula

$$\frac{1}{Z_{n,s}} (\det \rho)^{s-n} \mathbf{1}_{\rho \geqslant 0}, \tag{19}$$

where $Z_{n,s}$ is a normalization constant [46]. In the special $s = n$, the measure $\mu_{n,n}$ is the uniform measure on $\mathrm{D}(\mathbb{C}^n)$! This can be seen as the non-commutative analogue of the following classical fact: if $x = (x_1, \ldots, x_n)$ is chosen uniformly at random on the unit sphere on $\mathbb{C}^n$, then $(|x_1|^2, \ldots, |x_n|^2)$ is uniformly distributed on the simplex of length $n$ probability vectors.

### 4.2 Limit results for random induced states

Consider the representation of random induced states as normalized Wishart matrices as in (18). First note that $\operatorname{Tr} GG^\dagger$ is strongly concentrated around the value $ns$. An application of the law of large numbers to each entry of $GG^\dagger$ shows that, with $n$ fixed and $s \to \infty$, the measures $\mu_{n,s}$ concentrate towards the maximally mixed state $\mathrm{Id}/n$ (this is also clear from the formula (19) since the maximally mixed state is the unique state with maximal determinant). We will make this statement more quantitative by studying the rate of convergence, and also consider regimes when both $n$ and $s$ tend to infinity.

Given a self-adjoint $n \times n$ matrix $A$ with eigenvalues $(\lambda_1, \ldots, \lambda_n)$, it is convenient to introduce its empirical spectral distribution

$$\mu_A = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A)}.$$

Let $\rho_{n,s}$ be a random induced state with distribution $\mu_{n,s}$. Assume that $n$ and $s$ both tend to infinity in such a way that $\alpha = \lim s/n$ exists. We also assume $\alpha \geqslant 1$ and consider the rescaled states $n\rho_{n,s}$ whose eigenvalues are of order 1. Then, the sequence of

corresponding empirical spectral distributions converges towards a nonrandom measure $\mu_{\mathrm{MP},\alpha}$ which is called the Marčenko–Pastur distribution with parameter $\alpha$ and has support $[(1-1/\sqrt{\alpha})^2,(1+1/\sqrt{\alpha})^2]$. We write

$$\mu_{n\rho_{n,s}} \xrightarrow{*} \mu_{\mathrm{MP},\alpha}. \tag{20}$$

The meaning of $\xrightarrow{*}$ is the following: convergence is in probability (almost sure convergence also holds but is usually irrelevant to our setting) and encompasses both weak convergence of probability measures and convergence of supports.

Similarly, if $n$ and $s$ both to infinity in such a way that $\lim s/n = \infty$, the properly rescaled empirical spectral distributions approach the semicircle law $\mu_{\mathrm{SC}}$

$$\mu_{\sqrt{ns}(\rho-\mathrm{Id}/n)} \xrightarrow{*} \mu_{\mathrm{SC}}. \tag{21}$$

For fixed $n$ and $s \to \infty$, it follows from the multivariate central limit that we have an approximation

$$\rho = \frac{\mathrm{Id}}{n} + \frac{1}{\sqrt{n(n-1)s}}\Gamma_n \tag{22}$$

where $\Gamma_n$ is a standard Gaussian vector is the space of self-adjoint trace zero operators on $\mathbb{C}^n$. This is also a familiar object from random matrix theory: without the trace zero restriction one would get exactly GUE random matrices. Here $\Gamma_n$ can described as a GUE random matrix conditioned to have trace 0, or equivalently as $\Gamma_n = A_n - \mathrm{Tr}(A_n)\mathrm{Id}/n$ where $A$ is a $n \times n$ GUE random matrix.

One checks that formulas (21) and (22) are consistent: by Wigner's theorem we have

$$\mu_{\Gamma_n/\sqrt{n}} \xrightarrow{*} \mu_{\mathrm{SC}}. \tag{23}$$

However, one cannot formally deduce (21) from (22) and (23) because it would require to exchange the order in which limits are taken.

We already explained that for a fixed dimension $n$, the measures $\mu_{n,s}$ concentrate towards $\mathrm{Id}/n$ as $s$ tends to infinity. It makes sense to ask ourselves, given a property of a quantum state, for which values of $s$ does this property typically hold?

For properties which depend only on the spectrum, the answer is provided by the limiting results (20) and (21). However, most properties connected to the entanglement vs probability dichotomy cannot be inferred from spectrum. For such questions to make sense, we assume that the space $\mathbb{C}^n$ is identified with $\mathbb{C}^d \otimes \mathbb{C}^d$. Natural questions are: given a random state $\rho \in \mathrm{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$ with distribution $\mu_{d^2,s}$, is it typically entangled? PPT? What is the typical order or magnitude of the entanglement of formation? of the entanglement cost? of the distillable entanglement?

### 4.3 Threshold for the entanglement of formation

As a warm-up we discuss a threshold phenomenon for the entanglement of formation. Recall its definition

$$E_F(\rho) = \inf\left\{\sum \lambda_i E(\psi_i) \ : \ \rho = \sum \lambda_i |\psi_i\rangle\langle\psi_i|\right\}.$$

We show that when $\rho$ is a random induced state with distribution $\mu_{d^2,s}$, the typical value of $E_F(\rho)$ switches from almost maximal to almost minimal when $s$ is roughly of order $d^2$.

**Proposition 15** (see [21])**.** *Let $\rho$ be a random induced state on $\mathbb{C}^d \otimes \mathbb{C}^d$ with distribution $\mu_{d^2,s}$. Then*

1. *If $s \leqslant cd^2/\log^2 d$, then with high probability $E_F(\rho) \geqslant \log(d) - 1$.*
2. *For any $\varepsilon > 0$, if $s \geqslant C(\varepsilon) d^2 \log^2 d$, then with high probability $E_F(\rho) \leqslant \varepsilon$.*

*Here $C(\varepsilon)$ is a constant depending only on $\varepsilon$.*

Any improvement on Proposition 15, especially in the range when $s$ is of order $d^2$, would be welcome.

*Proof.* For the first part, we lower bound the average by the minimum

$$E_F(\rho) \geqslant \min_{\psi \in \mathrm{Range}(\rho)} E(\psi).$$

The range of $\rho$ is a random $s$-dimensional subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$. Applying Dvoretzky's theorem and using the fact that $E$ is $C \log d$-Lipschitz (see Exercise 14) gives that the minimum of $E$ over range$(\rho)$ is larger $\log d - 1$ with high probability provided $s \leqslant Cd^2/\log^2 d$, as claimed.

For the second part, we are going to use the convexity of $E_F$: let $\lambda_{\min}$ be the smallest eigenvalue of $\rho$ and write

$$\rho = \underbrace{(\rho - \lambda_{\min}\mathrm{Id})}_{(1-d^2\lambda_{\min})\sigma} + d^2\lambda_{\min}\frac{\mathrm{Id}}{d^2}.$$

This gives (note that $\sigma$ is a state)

$$E_F(\rho) \leqslant (1 - d^2\lambda_{\min}) E_F(\sigma) \leqslant (1 - d^2\lambda_{\min}) \log d.$$

We are reduced to estimating $\lambda_{\min}$. This depends only on the spectrum, so we know from (quantitative versions of) the limit-result (20) that

$$\lambda_{\min} = \frac{1}{d^2} - O\left(\frac{1}{d\sqrt{s}}\right).$$

This gives the announced result. □

### 4.4 Threshold for PPT

In some cases the study of thresholds in the spirit of Proposition 15 leads to interesting matrix models. This is the case for the PPT property. By (18), this is equivalent to the following question: when $G$ is a $d^2 \times s$ random matrix with independent standard complex Gaussian entries, when is the matrix

$$(GG^\dagger)^\Gamma$$

positive? Such problems are amenable to the techniques of random matrix theory, especially to the moment method which leads to combinatorial questions about non-crossing partitions.

Let $\rho_{d^2,s}$ be a state on $\mathbb{C}^d \otimes \mathbb{C}^d$ with distribution $\mu_{d^2,s}$. In the regime when both $d$ and $s$ tend to infinity with $\lim s/d^2 = \alpha \in (0, \infty)$, we have [3]

$$\mu_{d^2\rho^\Gamma} \xrightarrow{*} \mu_{\mathrm{SC}(1,1/\alpha)}$$

where $\mu_{\mathrm{SC}(m,\sigma^2)}$ denotes the semicircular distribution with mean $m$ and variance $\sigma^2$. By comparing with (20), we see that partial transposition has a non-trivial effect of the spectrum since it transforms the Marčenko–Pastur distribution $\mathrm{MP}(\alpha)$ into the semicircular distribution $\mathrm{SC}(1, 1/\alpha)$. (Note that both have the same first and second moments, since partial transposition preserves both the trace and the Hilbert–Schmidt norm.)

The support of the distribution $\mathrm{SC}(1, 1/\alpha)$ equals $[1 - 2/\sqrt{\alpha}, 1 + 2/\sqrt{\alpha}]$; it is contained in the positive half-line whenever $\alpha \geqslant 4$. The following dichotomy follows: for any $\varepsilon > 0$,

1. For $s < (4-\varepsilon)d^2$, a random state with distribution $\mu_{d^2,s}$ is non-PPT with probability tending to 1 as $s, d \to \infty$.

2. For $s > (4+\varepsilon)d^2$, a random state with distribution $\mu_{d^2,s}$ is PPT with probability tending to 1 as $s, d \to \infty$.

We may say that the value $s = 4d^2$ is a threshold for the PPT property of random induced states.

### 4.5 Central limit approximation for induced states

For general properties, the problem can be attacked via a geometric approach. Consider a closed convex set $K \subset \mathrm{D}(\mathbb{C}^n)$ and assume that the maximally mixed state $\mathrm{Id}/n$ belongs to the interior of $K$. We think of $\mathrm{Id}/n$ as the origin, making the affine space of trace 1 self-adjoint operators into a vector space.

The gauge associated to $K$ is

$$\|\rho\|_K = \inf \left\{ t > 0 \ : \ \frac{\mathrm{Id}}{n} + \frac{1}{t}\left(\rho - \frac{\mathrm{Id}}{n}\right) \in K \right\}$$

and has the property that $K = \{\rho \ : \ \|\rho\|_K \leqslant 1\}$. Suppose that $K$ corresponds to the of quantum states having a given property (P). The question whether random induced states typically have property (P) is the following: under the probability distribution $\mu_{n,s}$, is the typical value of $\|\cdot\|_K$ larger or smaller than 1?

In most settings, there is enough concentration of measure present to reduce the problem to the estimation of the expectation: if $\mathbf{E}\,\|\cdot\|_K < 1$, then $\rho \in K$ with high probability, while if $\mathbf{E}\,\|\cdot\|_K > 1$, then $\rho \notin K$ with high probability (see [6] for a general statement in this direction).

The following proposition is a quantitative version of the central limit approximation from (22) and compares average of gauges for induced states and for GUE matrices.

**Proposition 16.** *Let $\rho$ be a random induced state with distribution $\mu_{n,s}$, and $\Gamma$ a $n \times n$ GUE random matrix conditioned to have trace 0. For any convex body $K \subset \mathrm{D}(\mathbb{C}^d)$ containing $\mathrm{Id}/n$ in the interior,*

$$C_{n,s}^{-1}\,\mathbf{E}\left\|\frac{\mathrm{Id}}{n}+\frac{\Gamma}{n\sqrt{s}}\right\|_K \leqslant \mathbf{E}\,\|\rho\|_K \leqslant C_{n,s}\,\mathbf{E}\left\|\frac{\mathrm{Id}}{n}+\frac{\Gamma}{n\sqrt{s}}\right\|_K,$$

*where the constants $C_{n,s}$ have the property that $\lim C_{n,s} = 1$ whenever both $n$ and $s/n$ tend to infinity.*

We sketch a proof of Proposition 16. The proof is based on the following coupling argument: let $U$ be a random unitary matrix which is independent both from $\rho$ and from $\Gamma$. Since both models are unitary invariant, we have

- The random matrix $U\,\mathrm{diag}(\mathrm{spec}(\rho))U^\dagger$ has the same distribution as $\rho$,
- The random matrix $U\,\mathrm{diag}(\mathrm{spec}(G))U^\dagger$ has the same distribution as $G$.

By $\mathrm{diag}(\mathrm{spec}(A))$ we mean the diagonal matrix whose elements are the eigenvalues of $A$ (the way they are ordered is irrelevant). Denote by $\mathbb{R}^{n,0} \subset \mathbb{R}^n$ the hyperplane consisting of vectors whose sum of coordinates is 0 and introduce the function $\Phi : \mathbb{R}^{n,0} \to \mathbb{R}$ defined as

$$\Phi(x) = \mathbf{E}\,\|U\,\mathrm{diag}(x)U^\dagger\|_K.$$

The function $\Phi$ is convex and permutation-invariant. Such functions appear naturally in connection with majorization. Majorization is a partial order defined as follows: given $x, y \in \mathbb{R}^{n,0}$, we write $x \prec y$ if one of the following equivalent conditions is satisfied

1. For any $k \in \{1, \ldots, n-1\}$, $S_k(x) \leqslant S_k(y)$, where $S_k(z)$ denotes the sum of the $k$ largest coordinates of a vector $z \in \mathbb{R}^{n,0}$.
2. There is a bistochastic matrix $B$ such that $x = By$.
3. For any function $\Phi : \mathbb{R}^{n,0}$ which is convex and permutation-invariant, we have $\Phi(x) \leqslant \Phi(y)$.

We know that the vectors $\mathrm{spec}(\rho - \mathrm{Id}/n)$ and $\mathrm{spec}(\Gamma/n\sqrt{s})$ are comparable and become more and more comparable as $n$ and $s/n$ tend to infinity, by (21) and (23). If order to translate this into majorization, we use the following lemma, and Proposition 16 follows with little effort (see [6, 4]).

**Lemma 17.** *Let $x, y \in \mathbb{R}^{n,0}$. Assume that $\|x - y\|_\infty \leqslant \varepsilon$ and that $\|y\|_1 \geqslant \alpha n$. Then*

$$x \prec \left(1 + \frac{2\varepsilon}{\alpha}\right) y.$$

*Exercise* 16. Prove Lemma 17.

# 5 Separability of random states and convex geometry

## 5.1 Threshold for separability vs entanglement

In this last section, which in based on [6], we are going to study the following question: for which values of the parameters $d, s$ is it true that a random state $\rho$ on $\mathbb{C}^d \otimes \mathbb{C}^d$ with distribution $\mu_{d^2,s}$ is typically separable?

Assume that both $d$ and $s^2/d$ tend to infinity and apply Proposition 16 to the convex body $K = \mathrm{Sep} = \mathrm{Sep}(\mathbb{C}^d \otimes \mathbb{C}^d)$. We obtain that

$$\mathbf{E}\,\|\rho\|_{\mathrm{Sep}} \simeq \mathbf{E}\left\|\frac{\mathrm{Id}}{d^2} + \frac{\Gamma}{d^2\sqrt{s}}\right\|_{\mathrm{Sep}} \simeq \frac{w(\mathrm{Sep}^\circ)}{\sqrt{s}}$$

where $w(\mathrm{Sep}^\circ)$ denotes the mean width of the polar of Sep—these notions will be explained in Section 5.2. It follows from the discussion preceding Proposition 16 that a threshold for separability occurs at the value

$$s_0(d) = w(\mathrm{Sep}^\circ)^2 \tag{24}$$

in the following sense

1. For $s < (1-\varepsilon)s_0(d)$, a random state with distribution $\mu_{d^2,s}$ is entangled with probability tending to 1 as $s, d \to \infty$.

2. For $s > (1+\varepsilon)s_0(d)$, a random state with distribution $\mu_{d^2,s}$ is separable with probability tending to 1 as $s, d \to \infty$.

To get a complete picture we need to compute the value of $s_0(d)$: we will see in Section 5.3 that

$$cd^3 \leqslant s_0(d) \leqslant Cd^3(\log d)^2. \tag{25}$$

## 5.2 Notions from convex geometry

Let $K \subset \mathbb{R}^n$ be a convex body containing 0 in the interior. Its gauge is defined for $x \in \mathbb{R}^n$ as

$$\|x\|_K = \inf\{t > 0 \ : \ x \in tK\}.$$

Beware that possibly $\|x\|_K \neq \|-x\|_K$ if $K$ is not centrally symmetric! Similarly, the width of $K$ in direction $u$ (it would have been more geometrically correct to call this quantity the half-width) is defined as

$$w(K,u) = \sup_{x \in K} \langle x, u \rangle.$$

Width and gauge are dual quantities: indeed, if we introduce the polar convex body as

$$K^\circ = \{x \in \mathbb{R}^n \ : \ \forall y \in K, \langle x, y \rangle \leqslant 1\}$$

one checks that $\|\cdot\|_{K^\circ} = w(K, \cdot)$ and $\|\cdot\|_K = w(K^\circ, \cdot)$.

*Exercise* 17. Show that $\mathrm{D}(\mathbb{C}^n)^\circ = -n\mathrm{D}(\mathbb{C}^n)$ provided we take the maximally mixed state as the origin.

The average over $S^{n-1}$ of $w(K, \cdot)$ is called the mean width of $K$

$$w(K) = \int_{S^{n-1}} w(K,u) \, \mathrm{d}\sigma(u).$$

It is also convenient to introduce the Gaussian version of the mean width (see (8))

$$w_G(K) = \int_{\mathbb{R}^n} w(K,u) \, \mathrm{d}\gamma_n(u) = \kappa_n w(K) = \mathbf{E} \, \|G\|_K$$

where $G$ is a standard Gaussian vector in $\mathbb{R}^n$.

There is a strong connection between mean width and volume. Denote by $\mathrm{vol}(K)$ the volume (=Lebesgue measure) of $K$. The volume is $n$-homogeneous in the sense that $\mathrm{vol}(\lambda K) = \lambda^n \mathrm{vol}(K)$ for $\lambda > 0$. It is therefore often more convenient to work with a 1-homogeneous variant called the volume radius of $K$ and defined as

$$\mathrm{vrad}(K) = \left( \frac{\mathrm{vol}(K)}{\mathrm{vol}(B_2^n)} \right)^{1/n}$$

where $B_2^n$ is the unit Euclidean ball. It has the properties that $\mathrm{vrad}(B_2^n) = 1$ and $\mathrm{vrad}(\lambda K) = \lambda \, \mathrm{vrad}(K)$ for $\lambda > 0$. Note that the last two properties are also shared by the mean width. A result by Urysohn asserts that the volume radius is always smaller than the mean width.

**Proposition 18** (Urysohn inequality)**.** *For any convex body $K \subset \mathbb{R}^n$, we have* $\mathrm{vrad}(K) \leqslant w(K)$.

The Urysohn inequality belongs to the family of isoperimetric-like inequalities: at fixed volume, the mean width (and surface) of a convex body is minimal for Euclidean balls.

Much deeper results connect to the volume of a convex body to the volume of its polar body (the hardest part in the following proposition is the lower bound, for which several very different proofs are known [10, 26, 30]).

**Proposition 19** (Santaló and reverse Santaló inequalities)**.** *For any convex body $K \subset \mathbb{R}^n$ with center of mass at the origin,*

$$c \leqslant \mathrm{vrad}(K)\,\mathrm{vrad}(K^\circ) \leqslant 1 \tag{26}$$

*for some absolute constant $c$.*

Of course the crucial point is that the constant $c$ is dimension-free: the fact that the lower bound from (26) holds in a fixed dimension follows from a simple compactness argument.

There is no immediate analogue of Proposition 19 for the mean width: even in a fixed dimension the product $w(K)w(K^\circ)$ may be unbounded. In dimension 2, consider for example ellipses with eccentricity close to 1 (however, it is simple to show that the lower bound $w(K)w(K^\circ) \geqslant 1$ always holds).

Let us now consider some examples of convex bodies: the cube and its polar

$$B_\infty^n = [-1,1]^n,$$

$$B_1^n = (B_\infty^n)^\circ = \{x \in \mathbb{R}^n \ : \ |x_1| + \cdots + |x_n| \leqslant 1\},$$

for which one computes via formula (8) that $w(B_\infty^n)$ is of order $\sqrt{n}$ while $w(B_1^n)$ is of order $\sqrt{\log(n)/n}$. Another example is the regular $n$-dimensional simplex $\Delta_n$ which can be rescaled so that $\Delta_n^\circ = -\Delta_n$, and for which the mean width is of order $\sqrt{\log n}$. In these examples the product $w(K)w(K^\circ)$ is at most of order $\log n$. This turns out to be true in general (at least for symmetric convex bodies) provided one allows to apply a suitable linear transformation before computing mean widths.

**Theorem 20** (The $MM^*$-estimate)**.** *For every symmetric convex body $K \subset \mathbb{R}^n$, there is a linear transformation $T \in \mathsf{GL}(n)$ such that, denoting $\tilde{K} = T(K)$, we have*

$$w(\tilde{K})w(\tilde{K}^\circ) \leqslant C \log n \tag{27}$$

*for some absolute constant $C$.*

It is unknown whether the upper bound from (27) could be replaced by $C\sqrt{\log n}$. Another important open question is whether Theorem 20 holds for all convex bodies (without symmetry assumption), where $T$ is allowed to be any invertible affine transformation.

The proof of Theorem 20 is based on estimates on the $K$-convexity constant associated to a normed space. The way $T$ is defined is through an optimization problem; in particular $\tilde{K}$ inherits all the symmetries of $K$. This means that when the subgroup $G < \mathsf{O}(n)$ of isometries preserving $K$ acts irreducibly, Theorem 20 holds with $\tilde{K} = K$.

### 5.3 Estimation of the threshold

Consider $\mathrm{Sep} = \mathrm{Sep}(\mathbb{C}^d \otimes \mathbb{C}^d)$, the set of separable states on $\mathbb{C}^d \otimes \mathbb{C}^d$. We now use the material from Section 5.2 to estimate the threshold function $s_0(d)$. We have to prove—see (25) and (24)—that

$$cd^{3/2} \leqslant w(\mathrm{Sep}^\circ) \leqslant Cd^{3/2}\log(d).$$

Since the convex Sep has a simple description as a convex hull, it is easier to compute its width (which is a maximum over extreme points) than its gauge. We would like to apply Theorem 20 to conclude that

$$w(\mathrm{Sep})w(\mathrm{Sep}^\circ) \leqslant C\log(d). \tag{28}$$

Two issues have to be solved. First, the convex body Sep is not centrally symmetric (an hypothesis crucial in Theorem 20). Second, the isometry group of Sep does not act irreducibly and therefore we cannot conclude that $T = \mathrm{Id}$ is a valid choice in Theorem 20.

The first problem can be remedied via standard symmetrization techniques. For the second, we can use the fact that the isometry group of Sep acts irreducibly on the subspace $E = \mathrm{span}\{A \otimes B \ : \ \mathrm{Tr}\, A = \mathrm{Tr}\, B = 0\}$; since this subspace has a small codimension one can transfer inequalities like (27) from $E$ to the space of trace 1 self-adjoint operators on $\mathbb{C}^d \otimes \mathbb{C}^d$. We refer the reader to [6, 4] for more detail.

Once we know that (28) holds, it remains to estimate $w(\mathrm{Sep})$ via a routine net argument

$$w(\mathrm{Sep}) = \frac{1}{\kappa_{d^4-1}} w_G(\mathrm{Sep}) \sim \frac{1}{d^2}\,\mathbf{E} \sup_{\phi,\psi \in S_{\mathbb{C}^d}} |\langle \phi \otimes \psi | \Gamma | \phi \otimes \psi \rangle|$$

where $\Gamma$ is a $d^2 \times d^2$ GUE matrix. The set of all product vectors $\mathrm{Seg} = \{\phi \otimes \psi\}$ can be approximated by a $\frac{1}{4}$-net with less than $\exp(Cd)$ vertices (for example take all tensors of elements from a net in the sphere), yielding together with Lemma 7 the upper bound $w(\mathrm{Sep}) \leqslant Cd^{-3/2}$. Similarly, since Seg contains $\exp(cd)$ $\frac{1}{4}$ separated points, an application of the Sudakov minoration principle shows that this upper bound is sharp, i.e., that $w(\mathrm{Sep}) \geqslant cd^{3/2}$. This completes the proof of the estimates (25) on the threshold $s_0(d)$.

## References


[1] Erik Alfsen and Fred Shultz. Unique decompositions, faces, and automorphisms of separable states. *J. Math. Phys.*, 51(5):052201, 13, 2010.

[2] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[3] Guillaume Aubrun. Partial transposition of random states and non-centered semicircular distributions. *Random Matrices Theory Appl.*, 1(2):1250001, 29, 2012.

[4] Guillaume Aubrun and Stanisław Szarek. *Alice and Bob meet Banach*. 2017.

[5] Guillaume Aubrun, Stanisław Szarek, and Elisabeth Werner. Hastings's additivity counterexample via Dvoretzky's theorem. *Comm. Math. Phys.*, 305(1):85–97, 2011.

[6] Guillaume Aubrun, Stanisław J. Szarek, and Deping Ye. Entanglement thresholds for random induced states. *Comm. Pure Appl. Math.*, 67(1):129–171, 2014.

[7] Serban Belinschi, Benoît Collins, and Ion Nechita. Eigenvectors and eigenvalues in a random subspace of a tensor product. *Inventiones mathematicae*, 190(3):647–697, 2012.

[8] Serban T Belinschi, Benoit Collins, and Ion Nechita. Almost one bit violation for the additivity of the minimum output entropy. *arXiv preprint arXiv:1305.1567*, 2013.

[9] Charles H. Bennett, Herbert J. Bernstein, Sandu Popescu, and Benjamin Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53:2046–2052, Apr 1996.

[10] J. Bourgain and V. D. Milman. New volume ratio properties for convex symmetric bodies in $\mathbf{R}^n$. *Invent. Math.*, 88(2):319–340, 1987.

[11] Michael J. Bremner, Caterina Mora, and Andreas Winter. Are random pure states useful for quantum computation? *Phys. Rev. Lett.*, 102(19):190502, 4, 2009.

[12] Lin Chen and Dragomir Ž. Đoković. Dimensions, lengths, and separability in finite-dimensional quantum systems. *J. Math. Phys.*, 54(2):022201, 13, 2013.

[13] Eric Chitambar, Debbie Leung, Laura Mančinska, Maris Ozols, and Andreas Winter. Everything you always wanted to know about LOCC (but were afraid to ask). *Comm. Math. Phys.*, 328(1):303–326, 2014.

[14] Lieven Clarisse. The distillability problem revisited. *Quantum Inf. Comput.*, 6(6):539–560, 2006.

[15] S. K. Foong and S. Kanno. Proof of D. N. Page's conjecture on: "Average entropy of a subsystem" [Phys. Rev. Lett. **71** (1993), no. 9, 1291–1294; MR1232812 (94f:81007)]. *Phys. Rev. Lett.*, 72(8):1148–1151, 1994.

[16] D Gross, ST Flammia, and J Eisert. Most quantum states are too entangled to be useful as computational resources. *Physical review letters*, 102(19):190501, 2009.

[17] Leonid Gurvits and Howard Barnum. Largest separable balls around the maximally mixed bipartite quantum state. *Physical Review A*, 66(6):062311, 2002.

[18] Robin Hartshorne. *Algebraic geometry*. Springer-Verlag, New York-Heidelberg, 1977. Graduate Texts in Mathematics, No. 52.

[19] Matthew B Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5(4):255–257, 2009.

[20] Masahito Hayashi. *Quantum information*. Springer-Verlag, Berlin, 2006. An introduction, Translated from the 2003 Japanese original.

[21] Patrick Hayden, Debbie W. Leung, and Andreas Winter. Aspects of generic entanglement. *Comm. Math. Phys.*, 265(1):95–117, 2006.

[22] Patrick M. Hayden, Michał Horodecki, and Barbara M. Terhal. The asymptotic entanglement cost of preparing a quantum state. *J. Phys. A*, 34(35):6891–6898, 2001. Quantum information and computation.

[23] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1–2):1–8, 1996.

[24] Paweł Horodecki and Ryszard Horodecki. Distillation and bound entanglement. *Quantum Inf. Comput.*, 1(1):45–75, 2001.

[25] Ryszard Horodecki, Paweł Horodecki, and Karol Horodecki, Michałand Horodecki. Quantum entanglement. *Rev. Modern Phys.*, 81(2):865–942, 2009.

[26] Greg Kuperberg. From the Mahler conjecture to Gauss linking integrals. *Geom. Funct. Anal.*, 18(3):870–892, 2008.

[27] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[28] Paul Lévy. *Problèmes concrets d'analyse fonctionnelle. Avec un complément sur les fonctionnelles analytiques par F. Pellegrino*. Gauthier-Villars, Paris, 1951. 2d ed.

[29] V. D. Milman. A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. *Funkcional. Anal. i Priložen.*, 5(4):28–37, 1971.

[30] Fedor Nazarov. The Hörmander proof of the Bourgain-Milman theorem. In *Geometric aspects of functional analysis*, volume 2050 of *Lecture Notes in Math.*, pages 335–343. Springer, Heidelberg, 2012.

[31] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, 2000.

[32] K. R. Parthasarathy. On the maximal dimension of a completely entangled subspace for finite level quantum systems. *Proc. Indian Acad. Sci. Math. Sci.*, 114(4):365–374, 2004.

[33] Asher Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, Aug 1996.

[34] Gilles Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[35] Martin B. Plenio and Shashank Virmani. An introduction to entanglement measures. *Quantum Inf. Comput.*, 7(1-2):1–51, 2007.

[36] Gideon Schechtman. A remark concerning the dependence on $\epsilon$ in Dvoretzky's theorem. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 274–277. Springer, Berlin, 1989.

[37] Peter W. Shor. Equivalence of additivity questions in quantum information theory. *Comm. Math. Phys.*, 246(3):453–472, 2004.

[38] Barry Simon. Quantum dynamics: from automorphism to hamiltonian. *Studies in Mathematical Physics. Essays in Honor of Valentine Bargmann*, pages 327–349, 1976.

[39] Hans-Jürgen Sommers. Mini-Workshop: Geometry of Quantum Entanglement. *Oberwolfach Rep.*, 6(4):2993–3031, 2009. Abstracts from the mini-workshop held December 6–12, 2009, Organized by Andreas Buchleitner, Stanisław Szarek, Elisabeth Werner and Karol Życzkowski, Oberwolfach Reports. Vol. 6, no. 4.

[40] Erling Størmer. Positive linear maps of operator algebras. *Acta Math.*, 110:233–278, 1963.

[41] Jonathan Walgate and Andrew James Scott. Generic local distinguishability and completely entangled subspaces. *Journal of Physics A: Mathematical and Theoretical*, 41(37):375305, 2008.

[42] Reinhard F. Werner. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, Oct 1989.

[43] Eugene P. Wigner. *Group theory: And its application to the quantum mechanics of atomic spectra*. Expanded and improved ed. Translated from the German by J. J. Griffin. Pure and Applied Physics. Vol. 5. Academic Press, New York-London, 1959.

[44] Mark M. Wilde. *Quantum information theory*. Cambridge University Press, Cambridge, 2013.

[45] Dong Yang, Michał Horodecki, Ryszard Horodecki, and Barbara Synak-Radtke. Irreversibility for all bound entangled states. *Phys. Rev. Lett.*, 95:190501, Oct 2005.

[46] Karol Życzkowski and Hans-Jürgen Sommers. Induced measures in the space of mixed quantum states. *J. Phys. A*, 34(35):7111–7125, 2001. Quantum information and computation.