CHAPTER 8

Entanglement of Pure States in High Dimensions

Throughout this chapter, we consider a multipartite Hilbert space

$$\mathcal{H} = \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_k}$$

and study the entanglement of pure states on $\mathcal{H}$. We will *always* assume that $k \geqslant 2$ and that $d_1, \ldots, d_k \geqslant 2$.

We identify pure states on $\mathcal{H}$ with elements of $\mathsf{P}(\mathcal{H})$, the projective space on $\mathcal{H}$. The set of product vectors forms the Segré variety $\mathrm{Seg} \subset \mathsf{P}(\mathcal{H})$ (see (B.6) in Appendix B.2). A simple remark, on which we will elaborate, is that most pure states are entangled. Indeed, since the variety $\mathrm{Seg} \subset \mathsf{P}(\mathcal{H})$ has lower dimension and measure zero, it follows that a randomly chosen—in any reasonable sense—pure state in $\mathcal{H}$ is almost surely entangled.

A problem which turns out to be fundamental to several constructions in QIT is to show the existence of large-dimensional subspaces of $\mathcal{H}$, in which every unit vector corresponds to an entangled pure state. There are several variations on this question. We may consider the qualitative version of the problem, where we require the subspace simply to contain no nonzero product vector (see Theorem 8.1). Alternatively, we may insist that the subspace contains only very entangled vectors, once it is specified how to quantify entanglement; for pure states this may be done via the von Neumann or Rényi entropy of the partial trace.

The versions of Dvoretzky's theorem that were discussed in Section 7.2 are obviously relevant to such questions, since they show the existence of large subspaces on which a given function is almost constant. This approach allows us to give a complete presentation of Hastings's counterexample to the additivity problem (Section 8.4.4).

Much of our exposition will be focused on detailed study of the bipartite case $\mathcal{H} = \mathbb{C}^k \otimes \mathbb{C}^d$ (we will always assume that $k \leqslant d$). One reason for such emphasis is the fact that subspaces of a bipartite Hilbert space can provide a convenient description of quantum channels through the Stinespring representation, as we explain in Section 8.2.2. Fine aspects of pure state entanglement in multipartite systems are dealt with in the last part of the chapter (Section 8.5).

8.1. Entangled subspaces: qualitative approach

Let $\mathcal{H} = \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$. A fundamental *qualitative* question we may ask about entangled subspaces is: "What is the maximal dimension of a subspace of $\mathcal{H}$ in which every unit vector corresponds to an entangled pure state?" The answer to this question is $(d_1 - 1)(d_2 - 1)$, as shown by the following theorem, which also settles the multipartite case.

THEOREM 8.1. *Let $\mathcal{H} = \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_k}$, and let $n_0 = d_1 \cdots d_k - (d_1 + \cdots + d_k) + k - 1$. Then*

(1) *If $m > n_0$, then any m-dimensional subspace of $\mathcal{H}$ contains a (nonzero) product vector.*
(2) *If $m \leqslant n_0$, a generic m-dimensional subspace of $\mathcal{H}$ contains no (nonzero) product vector.*

PROOF. We only give an argument for the second part of the Theorem (the first assertion can be proved via the projective dimension theorem from algebraic geometry). The proof is based on dimension counting, and we find it instructive to give a "probabilistic" version of dimension counting, which naturally fits in the general framework of this book. For simplicity, we only consider the case $\mathcal{H} = \mathbb{C}^d \otimes \mathbb{C}^d$ (so that $n_0 = (d-1)^2$), the general case being similar.

We work in the projective space $\mathsf{P}(\mathcal{H})$, which we equip with the distance given by (B.5). The ball of center ψ and radius r is denoted by $B(\psi, r)$. We use bounds on the size of ε-nets in $\mathsf{P}(\mathcal{H})$ and the measure of ε-balls from Theorem 5.11 (and Exercise 5.25; the more elementary results from Section 5.1.2 would actually suffice, cf. Exercise 5.10 and (5.2)). In this proof, as opposed to most material in this book, the dependence of constants on the dimension *is* allowed, and we will denote by C, C' etc. positive constants which may depend on d and m, but are independent of the parameter ε.

Let F be a random m-dimensional subspace of $\mathcal{H}$, chosen with respect to the Haar measure on the Grassmann manifold. More concretely, we may realize F as $F = U(F_0)$, where F_0 is any fixed m-dimensional subspace, and U is a Haar-distributed unitary matrix. Denote also $\mathrm{Seg} \subset \mathsf{P}(\mathcal{H})$ the set of product vectors (the Segré variety).

We are going to show that the event $\mathrm{Seg} \cap F = \varnothing$ has probability 1. Given $\varepsilon > 0$, let $\mathcal{M}_\varepsilon$ be an ε-net inside the projective space $\mathsf{P}(F_0)$ with $\mathrm{card}(\mathcal{M}_\varepsilon) \leqslant (C'/\varepsilon)^{2m-2}$. Next, let $\mathcal{N}_\varepsilon$ be an ε-net inside $\mathsf{P}(\mathbb{C}^d)$ with $\mathrm{card}(\mathcal{N}_\varepsilon) \leqslant (C'/\varepsilon)^{2d-2}$. One checks that $\mathcal{N}_\varepsilon^{\otimes 2} := \{x \otimes y : \ x, y \in \mathcal{N}_\varepsilon\}$ is a 2ε-net inside Seg. We use the union bound in the following way

$$\begin{aligned}
\mathbf{P}(\mathrm{Seg} \cap F \neq \varnothing) &\leqslant \mathbf{P}\left(\bigcup_{\varphi \in \mathcal{N}_\varepsilon^{\otimes 2}} B(\varphi, 2\varepsilon) \cap U\left(\bigcup_{\psi \in \mathcal{M}_\varepsilon} B(\psi, \varepsilon) \right) \neq \varnothing \right) \\
&\leqslant \sum_{\varphi \in \mathcal{N}_\varepsilon^{\otimes 2}, \psi \in \mathcal{M}_\varepsilon} \mathbf{P}\left(B(\varphi, 2\varepsilon) \cap U(B(\psi, \varepsilon)) \neq \varnothing \right) \\
&\leqslant \sum_{\varphi \in \mathcal{N}_\varepsilon^{\otimes 2}, \psi \in \mathcal{M}_\varepsilon} \mathbf{P}\left(d(\varphi, U\psi) < 3\varepsilon \right).
\end{aligned}$$

The quantity $\mathbf{P}(d(\varphi, U\psi) < 3\varepsilon)$ does not depend on the particular points $\varphi, \psi \in \mathsf{P}(\mathcal{H})$, and is equal to the normalized measure of a ball of radius 3ε in $\mathsf{P}(\mathcal{H})$, which is bounded from above by $(C''\varepsilon)^{2d^2-2}$ (or see Exercise 5.11 for the exact value). Consequently,

$$\begin{aligned}
\mathbf{P}(\mathrm{Seg} \cap U(F_0) \neq \varnothing) &\leqslant \mathrm{card}(\mathcal{N}_\varepsilon^{\otimes 2}) \, \mathrm{card}(\mathcal{M}_\varepsilon) (C''\varepsilon)^{2d^2-2} \\
&\leqslant C\varepsilon^{2d^2-2-(2m-2)-2(2d-2)}.
\end{aligned}$$

Provided $m \leqslant (d-1)^2$, the last quantity tends to 0 as ε tends to 0. This shows that the event $\{F$ intersects Seg$\}$ has probability 0, so that F contains no nonzero product vector. □

EXERCISE 8.1 (Universal entanglers). Show that whenever $d \geqslant 4$, a generic unitary matrix $U \in \mathsf{U}(d^2)$ has the property that for every product unit vector $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$, $U|\psi\rangle\langle\psi|U^\dagger$ is entangled.

8.2. Entropies of entanglement and additivity questions

8.2.1. Quantifying entanglement for pure states. The most common way to quantify the entanglement of a bipartite pure state is to use the entropy of entanglement (for operational meanings of the entropy of entanglement, we refer to Notes and Remarks).

Let $\psi \in \mathbb{C}^k \otimes \mathbb{C}^d$ be a unit vector. The *entropy of entanglement* of ψ, denoted by $E(\psi)$, is defined as the von Neumann entropy of the reduced matrix $\rho = \operatorname{Tr}_{\mathbb{C}^d} |\psi\rangle\langle\psi|$.

$$E(\psi) = S(\rho) = -\operatorname{Tr} \rho \log \rho. \tag{8.1}$$

Both parties play a symmetric role since the two reduced matrices $\operatorname{Tr}_{\mathbb{C}^d} |\psi\rangle\langle\psi|$ and $\operatorname{Tr}_{\mathbb{C}^k} |\psi\rangle\langle\psi|$ have the same von Neumann entropy (in the matrix formalism, a consequence of the fact that $MM^\dagger$ and $M^\dagger M$ have the same nonzero eigenvalues for $M \in \mathsf{M}_{k,d}$). If $\psi = \sum \lambda_i \varphi_i \otimes \chi_i$ is a Schmidt decomposition of ψ, then

$$E(\psi) = -\sum \lambda_i^2 \log \lambda_i^2 = -2 \sum \lambda_i^2 \log \lambda_i. \tag{8.2}$$

For any $p \in [0, \infty]$, we introduce the p-entropy of entanglement, defined as

$$E_p(\psi) = S_p(\rho), \tag{8.3}$$

where $\rho = \operatorname{Tr}_{\mathbb{C}^d} |\psi\rangle\langle\psi|$ and S_p is the p-Rényi entropy introduced in Section 1.3.3. Recall that the case $p = 1$ corresponds to the von Neumann entropy, i.e., $E_1(\psi) = E(\psi)$ (as given by (8.1)). The limit cases $p = 0$ and $p = \infty$ should be interpreted as $E_0(\psi) = \log \operatorname{rank}(\psi)$ and $E_\infty(\psi) = -2 \log \max \lambda_1$, where $\operatorname{rank} \psi$ is the Schmidt rank of ψ and λ_1 its largest Schmidt coefficient.

Rényi entropies for $p > 1$ are easier to manipulate since they are closely related to Schatten norms. If we identify a vector $\psi \in \mathbb{C}^k \otimes \mathbb{C}^d$ with a matrix $M \in \mathsf{M}_{k,d}$ as explained in Section 0.8, we obtain (see (2.12))

$$\|\rho\|_p = \|M\|_{2p}^2 \tag{8.4}$$

and therefore

$$E_p(\psi) = \frac{p}{1-p} \log \|\rho\|_p = \frac{2p}{1-p} \log \|M\|_{2p}. \tag{8.5}$$

In all this chapter we assume that $k \leqslant d$, and therefore (for any $p \in [0, \infty]$) the p-entropy of entanglement varies between 0 and $\log k$. Moreover, a pure state ψ satisfies $E_p(\psi) = 0$ if and only if it is a product vector, and satisfies $E_p(\psi) = \log k$ if and only if it is a maximally entangled vector.

These definitions make sense only in the bipartite case, as they rely on the Schmidt decomposition of a bipartite pure state, which has no canonical analogue for the multipartite case. The limit case $p = \infty$ is different: E_∞ depends only on the largest Schmidt coefficient, which can be defined in a multipartite system as the maximal modulus of inner product (or the maximal overlap) with a product vector (cf. (2.13)). We elaborate on this in Section 8.5.

One of the goals of this chapter is to find subspaces $\mathcal{W} \subset \mathbb{C}^k \otimes \mathbb{C}^d$ which are very entangled, in the sense that the quantity $E(\psi)$ (or $E_p(\psi)$) has a uniform lower bound over all unit vectors $\psi \in \mathcal{W}$.

8.2.2. Channels as subspaces. A crucial insight allowing to relate analysis of quantum channels to high-dimensional convex geometry is the observation that there is an essentially one-to-one correspondence between channels and linear subspaces of composite Hilbert spaces. Specifically, let $\mathcal{W}$ be a subspace of $\mathbb{C}^k \otimes \mathbb{C}^d$ of dimension m. Then $\Phi : B(\mathcal{W}) \to \mathsf{M}_k$ defined by $\Phi(\rho) = \mathrm{Tr}_{\mathbb{C}^d}(\rho)$ is a quantum channel. Alternatively, and perhaps more properly, we could identify $\mathcal{W}$ with $\mathbb{C}^m$ via an isometry $V : \mathbb{C}^m \to \mathbb{C}^k \otimes \mathbb{C}^d$ whose range is $\mathcal{W}$ and define, for $\rho \in \mathsf{M}_m$, the corresponding channel $\Phi : \mathsf{M}_m \to \mathsf{M}_k$ by

$$\Phi(\rho) = \mathrm{Tr}_{\mathbb{C}^d}(V\rho V^\dagger). \tag{8.6}$$

There is no restriction in considering quantum channels of the form (8.6): by Stinespring representation theorem (Theorem 2.24), any quantum channel $\Phi : \mathsf{M}_m \to \mathsf{M}_k$ can be represented via (8.6) for some subspace $\mathcal{W} \subset \mathbb{C}^k \otimes \mathbb{C}^d$, with $d = km$.

It is now easy to define a natural family of random quantum channels. They will be associated, via the above scheme, to random m-dimensional subspaces $\mathcal{W}$ of $\mathbb{C}^k \otimes \mathbb{C}^d$, distributed according to the Haar measure on the corresponding Grassmann manifold (for some fixed positive integers m, d, k that will be specified later). Note that most interesting parameters of a channel defined by (8.6) depend only on the subspace $\mathcal{W} = V(\mathbb{C}^m)$ and not on a particular choice of the isometry V (see, e.g., Lemma 8.2). In this sense, the language of "random m-dimensional subspaces of $\mathbb{C}^k \otimes \mathbb{C}^d$" is equivalent to that of "random isometries from $\mathbb{C}^m$ to $\mathbb{C}^k \otimes \mathbb{C}^d$," with the corresponding mathematical objects being, respectively, the closely related Grassmann manifolds and Stiefel manifolds (see Appendix B.4).

8.2.3. Minimal output entropy and additivity problems. Given a quantum channel $\Phi : \mathsf{M}_m \to \mathsf{M}_k$, we define its *minimum output entropy* as

$$S^{\min}(\Phi) = S_1^{\min}(\Phi) = \min_{\rho \in \mathrm{D}(\mathbb{C}^m)} S(\Phi(\rho)), \tag{8.7}$$

as well as the p-entropy variant for $p \geqslant 0$,

$$S_p^{\min}(\Phi) = \min_{\rho \in \mathrm{D}(\mathbb{C}^m)} S_p(\Phi(\rho)).$$

The following lemma shows that, for channels defined via (8.6), the minimum output entropy depends only on the range of the isometry V.

LEMMA 8.2. *Let $\Phi : \mathsf{M}_m \to \mathsf{M}_k$ a random channel, obtained by* (8.6) *from a Haar-distributed isometry $V : \mathbb{C}^m \to \mathbb{C}^k \otimes \mathbb{C}^d$. Then, for any $0 \leqslant p \leqslant \infty$,*

$$S_p^{\min}(\Phi) = \min_{\psi \in \mathcal{W}, |\psi|=1} E_p(\psi),$$

where $\mathcal{W} \subset \mathbb{C}^k \otimes \mathbb{C}^d$ is the range of V.

PROOF. Since the function S_p is concave (see Section 1.3.3), the minimum is achieved on a pure state (pure states are extreme points of $\mathrm{D}(\mathbb{C}^m)$). Consequently,

$$S_p^{\min}(\Phi) = \min_{\varphi \in S_{\mathbb{C}^m}} S_p(\Phi(|\varphi\rangle\langle\varphi|)) = \min_{\psi \in \mathcal{W} \,:\, |\psi|=1} S_p(\mathrm{Tr}_{\mathbb{C}^d} |\psi\rangle\langle\psi|)$$

and the result follows. □

For some time, an important open problem in quantum information theory was to decide whether the quantity $S^{\min}$ is additive, i.e., whether every pair (Φ, Ψ) of quantum channels satisfies

$$S^{\min}(\Phi \otimes \Psi) \stackrel{?}{=} S^{\min}(\Phi) + S^{\min}(\Psi). \tag{8.8}$$

The problem admits several equivalent formulations with operational meaning, notably whether entangled inputs can increase the capacity of a quantum channel to transmit classical information. (Note that the inequality "$\leqslant$" in (8.8) always holds and is easy, see Exercise 8.2.)

A similar question can be asked for the quantities $S_p^{\min}$, the motivation being that a positive answer to the $p > 1$ question would have implied a positive answer to the (arguably more important) $p = 1$ problem. However, it turns out that all these equalities are do not hold, at least for sufficiently large dimensions.

THEOREM 8.3. *For any $p \geqslant 1$, there exist quantum channels Φ, Ψ such that*

$$S_p^{\min}(\Phi \otimes \Psi) < S_p^{\min}(\Phi) + S_p^{\min}(\Psi). \tag{8.9}$$

Theorem 8.3 will be a consequence of Proposition 8.6 (for $p > 1$) and Proposition 8.24 (for $p = 1$).

EXERCISE 8.2 ($S_p^{\min}$ is always subadditive). Show that the inequality $S_p^{\min}(\Phi \otimes \Psi) \leqslant S_p^{\min}(\Phi) + S_p^{\min}(\Psi)$ is satisfied for any channels Φ, Ψ and any $p \geqslant 0$.

EXERCISE 8.3 (Reduction of the additivity problem to the case $\Phi = \Psi$). A trick based on direct sums (as defined in (2.42)) allows a reduction to the case $\Phi = \Psi$ in questions such as (8.8).
(i) Given quantum channels Φ, Ψ, show that $S_p^{\min}(\Phi \oplus \Psi) = \min(S_p^{\min}(\Phi), S_p^{\min}(\Psi))$.
(ii) Assume that there is a pair of channels Φ, Ψ such that (8.9) holds for some p. Deduce formally the existence of a channel Ξ such that $S_p^{\min}(\Xi \otimes \Xi) < 2S_p^{\min}(\Xi)$.

8.2.4. On the $1 \to p$ norm of quantum channels. The $p > 1$ version of the additivity problem has a nice functional-analytic interpretation. If $p > 1$ and ρ is a state, then $S_p(\rho) = \frac{p}{1-p} \log \|\rho\|_p$, and so the study of $S_p^{\min}(\Phi)$ is replaced by that of $\max\{\|\Phi(\rho)\|_p \ : \ \rho \in \mathrm{D}(\mathbb{C}^m)\}$, or the *maximum output p-norm*. The latter quantity equals $\|\Phi\|_{1\to p}$, i.e., the norm of Φ as an operator from $(\mathsf{M}_m^{\mathrm{sa}}, \|\cdot\|_1)$ to $(\mathsf{M}_k^{\mathrm{sa}}, \|\cdot\|_p)$. Therefore (8.9) is equivalent to

$$\|\Phi \otimes \Psi\|_{1\to p} > \|\Phi\|_{1\to p} \|\Psi\|_{1\to p}. \tag{8.10}$$

A remarkable fact is that for completely positive maps (and even for 2-positive maps), the norm $\|\cdot\|_{1\to p}$ is unchanged if we drop the self-adjointness constraint.

PROPOSITION 8.4. *Let $\Phi : \mathsf{M}_m \to \mathsf{M}_k$ be a 2-positive map, and $p \geqslant 1$. Then*

$$\sup_{X \in \mathsf{M}_m, \|X\|_1 = 1} \|\Phi(X)\|_p = \sup_{X \in \mathsf{M}_m^{\mathrm{sa}}, \|X\|_1 = 1} \|\Phi(X)\|_p \tag{8.11}$$

We first show the following fact

LEMMA 8.5. *If $A, B, C \in \mathsf{M}_k$ are such that the block matrix $M = \begin{bmatrix} A & B \\ B^\dagger & C \end{bmatrix}$ is positive semi-definite, then for every $p \geqslant 1$, $\|B\|_p^2 \leqslant \|A\|_p \|C\|_p$.*

PROOF. From the singular value decomposition, there exist unitary matrices $U, V \in \mathsf{U}(k)$ such that $UBV^\dagger$ is a diagonal matrix with nonnegative diagonal entries. Denote $W = U \oplus V \in \mathsf{U}(2k)$. We have

$$WMW^\dagger = \begin{bmatrix} UAU^\dagger & UBV^\dagger \\ VB^\dagger U^\dagger & VCV^\dagger \end{bmatrix}.$$

Since the Schatten norms are invariant under multiplication by unitaries, this shows that to prove the Lemma it is enough to treat the case when the matrix B is diagonal with nonnegative entries, which we consider now.

We first note that $b_{ii}^2 \leqslant a_{ii}c_{ii}$, which follows from the matrix $\begin{bmatrix} a_{ii} & b_{ii} \\ b_{ii} & c_{ii} \end{bmatrix}$ being positive as a submatrix of M. Consequently, we have

$$\|B\|_p^p = \sum_{i=1}^k b_{ii}^p \leqslant \sum_{i=1}^k a_{ii}^{p/2} c_{ii}^{p/2} \leqslant \left(\sum_{i=1}^k a_{ii}^p\right)^{1/2} \left(\sum_{i=1}^k c_{ii}^p\right)^{1/2} \leqslant \|A\|_p^{p/2} \|C\|_p^{p/2},$$

where the last inequality uses the fact that the diagonal is majorized by the spectrum (Lemma 1.14). □

PROOF OF PROPOSITION 8.4. For $\varphi, \psi \in S_{\mathbb{C}^m}$, consider $u = \varphi \otimes |1\rangle + \psi \otimes |2\rangle \in \mathbb{C}^m \otimes \mathbb{C}^2$. By direct calculation

$$\Phi \otimes \mathrm{Id}_{\mathsf{M}_2}(|u\rangle\langle u|) = \begin{bmatrix} \Phi(|\varphi\rangle\langle\varphi|) & \Phi(|\psi\rangle\langle\varphi|) \\ \Phi(|\varphi\rangle\langle\psi|) & \Phi(|\psi\rangle\langle\psi|) \end{bmatrix}.$$

Since Φ is 2-positive, the resulting matrix is block-positive and thus, by Lemma 8.5,

$$\|\Phi(|\psi\rangle\langle\varphi|)\|_p^2 \leqslant \|\Phi(|\psi\rangle\langle\psi|)\|_p \, \|\Phi(|\varphi\rangle\langle\varphi|)\|_p .$$

Taking supremum over unit vectors gives the required result (recall that extreme points of S_1^d and $S_1^{d,\mathrm{sa}}$ are rank 1 operators). □

EXERCISE 8.4 (The equality (8.11) does not hold always). Define $\Phi : \mathsf{M}_2 \to \mathsf{M}_2$ by $\Phi(X) = X - \mathrm{Tr}(X)\frac{1}{2}$. Show that for $p > 1$, Φ fails to satisfy the equality (8.11). Known examples where (8.11) fails for $p = 1$ are more complicated, see **[Wat05]**.

8.3. Concentration of E_p for $p > 1$ and applications

8.3.1. Counterexamples to the multiplicativity problem. We first consider the case of the p-entropy of entanglement with $p > 1$, and show that the Dvoretzky theorem can be used to produce counterexamples to the multiplicativity problem as announced in Theorem 8.3.

PROPOSITION 8.6. *There is a constant c such that the following holds. Let $p > 1$, and $\Phi : \mathsf{M}_m \to \mathsf{M}_k$ be a random channel, obtained by (8.6) from a Haar-distributed isometry $V : \mathbb{C}^m \to \mathbb{C}^k \otimes \mathbb{C}^d$. Denote $\Psi = \overline{\Phi}$, the channel obtained from $\overline{V}$, the complex conjugate of V. Assume that $k = d$ and that $m = cd^{1+1/p}$. Then, for d large enough, with high probability,*

$$\|\Phi \otimes \Psi\|_{1\to p} > \|\Phi\|_{1\to p} \|\Psi\|_{1\to p}. \tag{8.12}$$

PROOF. Denote by $\mathcal{W} \subset \mathsf{M}_d$ the range of V (we may consider $\mathcal{W}$ as a subspace of M_d after we identify tensors and matrices). From (8.4) and Lemma 8.2, we have

$$\|\Phi\|_{1\to p} = \max_{A \in \mathcal{W} \,:\, \|A\|_{\mathrm{HS}}=1} \|A\|_{2p}^2. \tag{8.13}$$

We remark that $\|\Phi\|_{1\to p} = \|\Psi\|_{1\to p}$ since the Schatten norms are invariant under complex conjugation. We now appeal to Dvoretzky's theorem for the Schatten norm $\|\cdot\|_q$ with $q = 2p$. Provided that $m \leqslant cd^{1+2/q}$ for an appropriate universal constant $c > 0$, it follows from Theorem 7.37 that, with large probability

$$d^{1/q-1/2}\|A\|_{\mathrm{HS}} \leqslant \|A\|_q \leqslant Cd^{1/q-1/2}\|A\|_{\mathrm{HS}}$$

for all $A \in \mathcal{W}$. We have therefore, by (8.13),

$$d^{1/p-1} \leqslant \|\Phi\|_{1\to p} = \|\Psi\|_{1\to p} \leqslant \left(Cd^{1/q-1/2}\right)^2 = C^2 d^{1/p-1}. \tag{8.14}$$

The reason for choosing $\overline{\Phi}$ as a second channel is that the channel $\Phi\otimes\overline{\Phi}$ necessarily has at least one output with at least one large eigenvalue, as shown by the following lemma.

LEMMA 8.7. *Let $\Phi : \mathsf{M}_m \to \mathsf{M}_k$ be a quantum channel obtained from an isometry $V : \mathbb{C}^m \to \mathbb{C}^k \otimes \mathbb{C}^d$, as in* (8.6). *Denote by $\psi \in \mathbb{C}^m \otimes \mathbb{C}^m$ the maximally entangled state*

$$\psi = \frac{1}{\sqrt{m}}\left(|1\rangle\otimes|1\rangle + \cdots + |m\rangle\otimes|m\rangle\right).$$

Then

$$\left\|(\Phi\otimes\overline{\Phi})(|\psi\rangle\langle\psi|)\right\|_\infty \geqslant \frac{m}{dk}$$

and consequently, for any $p > 1$,

$$\|\Phi\otimes\overline{\Phi}\|_{1\to p} \geqslant \|\Phi\otimes\overline{\Phi}\|_{1\to\infty} \geqslant \frac{m}{dk}$$

In our setting, $d = k$ and $m = cd^{1+1/p}$, so we obtain from Lemma 8.7 the lower bound $\|\Phi\otimes\overline{\Phi}\|_{1\to p} = \Omega(d^{1/p-1})$. Since we have, by (8.14),

$$\|\Phi\|_{1\to p}\|\overline{\Phi}\|_{1\to p} = \|\Phi\|^2_{1\to p} = \Theta(d^{2(1/p-1)}),$$

we conclude that the inequality (8.12) holds for d large enough (*a priori* depending on $p > 1$). □

REMARK 8.8. The proof shows that, for any fixed $p > 1$, both the multiplicative violation in (8.10) and the additive violation in (8.9) tend to infinity as the dimension of the problem increases (at the rates $\Omega(d^{1-1/p})$ and $\Omega(\log d)$ respectively).

PROOF OF LEMMA 8.7. We work in the matrix formalism. Identify the range of V with an m-dimensional subspace $\mathcal{W} \subset \mathsf{M}_{k,d}$. Let $(A_1, \ldots, A_m)$ be the orthonormal basis in $\mathcal{W}$ (with respect to the Hilbert–Schmidt inner product) obtained as the image under V of the canonical basis in $\mathbb{C}^m$, and

$$M = \frac{1}{\sqrt{m}}\sum_{i=1}^m A_i \otimes \overline{A_i} \in \mathcal{W}\otimes\overline{\mathcal{W}}.$$

The conclusion of the Lemma is equivalent to the inequality $\|M\|_\infty \geqslant \sqrt{m/kd}$.

Let $(\varphi_j)_{1\leqslant j\leqslant k}$ and $(\psi_{j'})_{1\leqslant j'\leqslant d}$ be orthonormal bases in $\mathbb{C}^k$ and $\mathbb{C}^d$, respectively. We consider the maximally entangled states

$$\varphi = \frac{1}{\sqrt{k}}\sum_{j=1}^k \varphi_j\otimes\overline{\varphi_j}, \quad \psi = \frac{1}{\sqrt{d}}\sum_{j'=1}^d \psi_{j'}\otimes\overline{\psi_{j'}}$$

and compute

$$\|M\|_\infty \quad \geqslant \quad \left|\langle\psi|M|\varphi\rangle\right|$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{mkd}} \sum_{i=1}^{m} \sum_{j=1}^{k} \sum_{j'=1}^{d} \left| \langle \psi_{j'} \otimes \overline{\psi_{j'}} | A_i \otimes \overline{A_i} | \varphi_j \otimes \overline{\varphi_j} \rangle \right| \\
&= \frac{1}{\sqrt{mkd}} \sum_{i=1}^{m} \sum_{j=1}^{k} \sum_{j'=1}^{d} \left| \langle \psi_{j'} | A_i | \varphi_j \rangle \right|^2 \\
&= \frac{\sqrt{m}}{\sqrt{kd}},
\end{aligned}
$$

where we used the fact that $\|X\|_{\mathrm{HS}}^2 = \sum_{j,j'} \left| \langle \psi_{j'} | X | \varphi_j \rangle \right|^2$. □

EXERCISE 8.5 (Non-random counterexamples for $p > 2$). Let $\mathcal{W} \subset \mathsf{M}_d$ the subspace of anti-symmetric matrices, i.e., such that $A^T = -A$.
(i) Show that for any $A \in \mathcal{W}$, $\|A\|_\infty \leqslant \frac{1}{\sqrt{2}} \|A\|_{\mathrm{HS}}$.
(ii) Let Φ be the quantum channel constructed from $\mathcal{W}$ as in (8.6) and fix $p > 2$. Using Lemma 8.7, show that the pair $(\Phi, \overline{\Phi})$ is an example for which (8.10) holds for d large enough.

8.3.2. Almost randomizing channels. A variant of the construction used in the proof of Proposition 8.6 for $p = +\infty$ gives the following: a channel $\Phi : \mathsf{M}_d \to \mathsf{M}_d$ constructed from a generic random embedding $V : \mathbb{C}^d \to \mathbb{C}^d \otimes \mathbb{C}^N$ with $N = O(d)$ has the property that $\|\Phi(\rho)\|_{\mathrm{op}} \leqslant C/d$ for any state $\rho \in \mathrm{D}(\mathbb{C}^d)$. In other words, all output states have small eigenvalues. It is natural to ask whether similar lower bounds of the eigenvalues of output states can also be achieved; showing that this is indeed the case is the content of this section. Recall also (see Section 2.3.3) that the dimension N of the environment in the Stinespring representation is an upper bound on the Kraus rank of Φ.

Let $0 < \varepsilon < 1$. A quantum channel $\Phi : \mathsf{M}_d \to \mathsf{M}_d$ is said to be ε-randomizing if for all states $\rho \in \mathrm{D}(\mathbb{C}^d)$

$$\|\Phi(\rho) - \rho_*\|_{\mathrm{op}} \leqslant \varepsilon/d.$$

Recall that $\rho_* = \mathrm{I}/d$ denotes the maximally mixed state. These channels can be thought as approximations of the completely randomizing channel R, which is defined by the property $R(\rho) = \rho_*$ for any $\rho \in \mathrm{D}(\mathbb{C}^d)$. The completely randomizing channel rank has Kraus rank equal to d^2 (see Exercise 8.6). On the other hand, it turns out that there exist ε-randomizing channels with a substantially smaller Kraus rank, as shown by the following theorem. The dependence on d is optimal since any ε-randomizing channel has Kraus rank at least d, which is due to the fact that rank one states must be mapped to full rank states.

THEOREM 8.9. *Let $(U_i)_{1 \leqslant i \leqslant N}$ be independent random matrices Haar-distributed on the unitary group $\mathsf{U}(d)$. Let $\Phi : \mathsf{M}_d \to \mathsf{M}_d$ be the quantum channel defined by*

$$\Phi(\rho) = \frac{1}{N} \sum_{i=1}^{N} U_i \rho U_i^\dagger.$$

Assume that $0 < \varepsilon < 1$ and $N \geqslant Cd/\varepsilon^2$. Then the channel Φ is ε-randomizing with high probability.

The proof of Theorem 8.9 is based on two lemmas.

LEMMA 8.10. *Let ρ and σ be pure states on $\mathbb{C}^d$ and let $(U_i)_{1\leqslant i\leqslant N}$ be independent Haar-distributed random unitary matrices. Then, for every $0 < \delta < 1$,*

$$\mathbf{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N}\operatorname{Tr}(U_i\rho U_i^\dagger\sigma) - \frac{1}{d}\right| \geqslant \frac{\delta}{d}\right) \leqslant 2\exp(-c\delta^2 N).$$

PROOF. Write $\rho = |\varphi\rangle\langle\varphi|$ and $\sigma = |\psi\rangle\langle\psi|$. Denote $X_i = d\operatorname{Tr}(U_i\rho U_i^\dagger\sigma) = \left|\sqrt{d}\langle\psi|U_i|\varphi\rangle\right|^2$. We know from Lemma 5.57 that this variable is subexponential (as the square of a subgaussian variable) and satisfies $\|X_i\|_{\psi_1} \leqslant C$. The conclusion follows now directly from Bernstein's inequalities (Proposition 5.59). □

LEMMA 8.11. *Let $\Delta : \mathsf{M}_d^{\mathrm{sa}} \to \mathsf{M}_d^{\mathrm{sa}}$ be a linear map. Let A be the quantity*

$$A = \sup_{\rho\in\mathrm{D}(\mathbb{C}^d)} \|\Delta(\rho)\|_{\mathrm{op}} = \sup_{\rho,\sigma\in\mathrm{D}(\mathbb{C}^d)} |\operatorname{Tr}\sigma\Delta(\rho)|$$

Let $0 < \delta < 1/4$ and $\mathcal{N}$ be a δ-net in $(S_{\mathbb{C}^d}, |\cdot|)$. Then $A \leqslant (1-4\delta)^{-1}B$, where

$$B = \sup_{\varphi,\psi\in\mathcal{N}} |\operatorname{Tr}|\psi\rangle\langle\psi|\Delta(|\varphi\rangle\langle\varphi|)|\,.$$

PROOF OF LEMMA 8.11. First note that for any $X, Y \in \mathsf{M}_d^{\mathrm{sa}}$, we have

$$|\operatorname{Tr} Y\Delta(X)| \leqslant A\|X\|_1\|Y\|_1. \tag{8.15}$$

By a convexity argument, the supremum in A can be restricted to pure states. Given unit vectors $\varphi, \psi \in S_{\mathbb{C}^d}$, let $\varphi_0, \psi_0 \in \mathcal{N}$ so that $|\varphi - \varphi_0| \leqslant \delta$ and $|\psi - \psi_0| \leqslant \delta$. Given $\chi \in S_{\mathbb{C}^d}$, we write P_χ for $|\chi\rangle\langle\chi|$. We have

$$\|P_\varphi - P_{\varphi_0}\|_1 \leqslant \|P_\varphi - |\varphi\rangle\langle\varphi_0|\|_1 + \||\varphi\rangle\langle\varphi_0| - P_{\varphi_0}\|_1 \leqslant 2\delta$$

and similarly $\|P_\psi - P_{\psi_0}\|_1 \leqslant 2\delta$ (this simple bound is not optimal). We now write

$$|\operatorname{Tr} P_\psi\Delta(P_\varphi)| \leqslant |\operatorname{Tr}(P_\psi - P_{\psi_0})\Delta(P_\varphi)| + |\operatorname{Tr} P_{\psi_0}\Delta(P_\varphi - P_{\varphi_0})| + |\operatorname{Tr} P_{\psi_0}\Delta(P_{\varphi_0})|\,.$$

Using twice (8.15) and taking supremum over φ, ψ gives $A \leqslant 2\delta A + 2\delta A + B$, hence the result. □

PROOF OF THEOREM 8.9. Fix a $\frac{1}{8}$-net $\mathcal{N} \subset (S_{\mathbb{C}^d}, |\cdot|)$ with $\operatorname{card}\mathcal{N} \leqslant 16^{2d}$, as provided by Lemma 5.3. Let $\Delta = R - \Phi$ and A, B as in Lemma 8.11. Here A and B are random quantities and it follows from Lemma 8.11 that

$$\mathbf{P}\left(A \geqslant \frac{\varepsilon}{d}\right) \leqslant \mathbf{P}\left(B \geqslant \frac{\varepsilon}{2d}\right).$$

Using the union bound and Lemma 8.10, we get

$$\mathbf{P}\left(B \geqslant \frac{\varepsilon}{2d}\right) \leqslant 16^{4d}\cdot 2\exp(-c\varepsilon^2 N/4).$$

This is less than 1 if $N \geqslant Cd/\varepsilon^2$, for some constant C. □

EXERCISE 8.6 (Kraus decomposition of the completely randomizing channel).
(i) Show that the Kraus rank of the completely randomizing channel R is d^2.
(ii) Let $\omega = \exp(2i\pi/d)$ and A, B be the unitary operators defined by their action on the canonical basis by

$$A|j\rangle = |j+1 \bmod d\rangle \qquad B|j\rangle = \omega^j|j\rangle. \tag{8.16}$$

Show that the operators $(B^jA^k)_{1\leqslant j,k\leqslant d}$ give a Kraus decomposition of R. These operators are sometimes called the Heisenberg–Weyl operators.

8.4. Concentration of von Neumann entropy and applications

8.4.1. The basic concentration argument. We now consider the von Neumann entropy (instead of the p-Rényi entropy) as the invariant quantifying entanglement. Since the von Neumann entropy is not naturally associated with a norm, we are going to use the version of Dvoretzky theorem for Lipschitz functions (Theorem 7.15). The relevant function is the entropy of entanglement $\psi \mapsto E(\psi)$, defined (via (8.1)) on the unit sphere in $\mathbb{C}^k \otimes \mathbb{C}^d$. As usual in such situations, we need two pieces of information: the Lipschitz constant of $E(\cdot)$ and a central value. They are provided by the next two lemmas.

LEMMA 8.12. *The Lipschitz constant of the function $\psi \mapsto E(\psi)$, defined on $(S_{\mathbb{C}^k \otimes \mathbb{C}^d}, |\cdot|)$ is bounded from above by $C \log k$ for some absolute constant C.*

This is clearly optimal up to the value of the constant C, since the function E maps $S_{\mathbb{C}^k \otimes \mathbb{C}^d}$ (which has diameter π, or $\pi/2$ if we consider E as a function on $\mathsf{P}(\mathbb{C}^k \otimes \mathbb{C}^d)$) onto the segment $[0, \log k]$. (Remember that in this chapter we always assume $k \leqslant d$.) Note that, in view of (B.1), it doesn't matter—apart from the value of the constant—whether we use the geodesic distance or the extrinsic distance. For a discussion of the optimal values of the constants see Exercise 8.7.

PROOF. We first check the commutative case by considering the function $f : S^{k-1} \to [0, \log k]$ defined by

$$f(x) = -\sum x_i^2 \log(x_i^2), \tag{8.17}$$

i.e., the Shannon entropy of the probability distribution $(x_i^2) \in \Delta_k$. In the terminology of (8.2), this is equivalent to restricting attention to vectors ψ whose Schmidt decompositions use fixed sequences (φ_i), (χ_i). One computes

$$|\nabla f(x)|^2 = 4 \sum_{i=1}^{k} x_i^2 (1 + \log(x_i^2))^2 \leqslant C \log^2 k, \tag{8.18}$$

where the last inequality can be obtained by observing that the function $t \mapsto t(1 + \log t)^2$ is concave on $[0, e^{-2}]$, and so the quantity $|\nabla f(x)|$ increases when we replace the coordinates of x smaller than e^{-1} by their ℓ_2 average. It follows that if L is the Lipschitz constant of f with respect to the geodesic distance on S^{k-1}, then $L \leqslant C^{1/2} \log k$. Our objective is to show is that the same constant works for the function $\psi \mapsto E(\psi)$.

To that end, we will consider an auxiliary function which is defined as follows. Let $(u_i)_{1 \leqslant i \leqslant k}$ be an orthonormal basis of $\mathbb{C}^k$. If $\psi \in S_{\mathbb{C}^k \otimes \mathbb{C}^d}$, set $\rho = \mathrm{Tr}_{\mathbb{C}^d} |\psi\rangle\langle\psi|$ and let

$$\tilde{f}(\psi) = -\sum_{i=1}^{k} \langle u_i|\rho|u_i\rangle \log(\langle u_i|\rho|u_i\rangle). \tag{8.19}$$

In other words, $\tilde{f}(\psi)$ is the entropy of the diagonal part of ρ, calculated in the basis (u_i). An important property of $\tilde{f}$ is that $\tilde{f}(\psi) = S(\rho)$ if (u_i) a basis which diagonalizes ρ (which is obvious from the definitions) and $\tilde{f}(\psi) \leqslant S(\rho)$ in general (which is a consequence of concavity of S and is the content of Exercise 1.50). Next, one verifies that $\langle u_i|\rho|u_i\rangle = |P_i\psi|$, where P_i is the orthogonal projection onto the subspace $u_i \otimes \mathbb{C}^d \subset \mathbb{C}^k \otimes \mathbb{C}^d$. Since the map $\psi \mapsto (|P_1\psi|, \ldots, |P_k\psi|)$ is a contraction,

it follows that the Lipschitz constant of $\tilde{f}$ (with respect to g, the geodesic distance on $S_{\mathbb{C}^k\otimes\mathbb{C}^d}$) is at most L.

We now return to the original question. Let $\psi_1, \psi_2 \in S_{\mathbb{C}^k\otimes\mathbb{C}^d}$; set $\rho_k = \operatorname{Tr}_{\mathbb{C}^d} |\psi_k\rangle\langle\psi_k|$ and let $\tilde{f}$ be defined by (8.19) using a basis (u_i) which diagonalizes ρ_1. Then

$$E(\psi_1) - E(\psi_2) = S(\rho_1) - S(\rho_2) = \tilde{f}(\psi_1) - S(\rho_2) \leqslant \tilde{f}(\psi_1) - \tilde{f}(\psi_2) \leqslant L\, g(\psi_1, \psi_2).$$

Since the roles of ψ_1 and ψ_2 can be reversed, it follows that the Lipschitz constant of E with respect to g is at most L (and hence exactly L), as claimed. □

LEMMA 8.13 (not proved here, but see Remark 8.14). *For $k \leqslant d$, the expectation of the function $\psi \mapsto E(\psi)$ (with respect to the uniform measure on the unit sphere in $\mathbb{C}^k \otimes \mathbb{C}^d$) satisfies*

$$\mathbf{E}\, E(\psi) = \left(\sum_{j=d+1}^{kd} \frac{1}{j}\right) - \frac{k-1}{2d} \geqslant \log k - \frac{1}{2}\frac{k}{d}. \tag{8.20}$$

REMARK 8.14 (An easy bound on the entropy of entanglement). An inequality slightly weaker than (8.20) follows readily from Proposition 6.36 (or Exercise 6.43, which is even more elementary). First, with large probability, all Schmidt coefficients of ψ belong to the interval

$$\left[\frac{1}{\sqrt{k}} - \frac{C}{\sqrt{d}}, \frac{1}{\sqrt{k}} + \frac{C}{\sqrt{d}}\right]$$

for some constant C. It follows that all the eigenvalues of the $\operatorname{Tr}_{\mathbb{C}^d} |\psi\rangle\langle\psi|$ lie then in an interval $\left[\frac{1-\varepsilon}{k}, \frac{1+\varepsilon}{k}\right]$ for some $\varepsilon = O\big(\sqrt{k/d}\big)$, and Lemma 1.20 yields the bound $E(\psi) = S(\rho) \geqslant \log k - C'k/d$. (The use of Lemma 1.20 requires $\varepsilon \leqslant 1$, for larger ε we may use the simpler bound $S(\rho) \geqslant S_\infty(\rho) = -\log \|\rho\|_\infty$.)

An immediate consequence of Dvoretzky's theorem (in the form from Theorem 7.15) is now:

THEOREM 8.15. *Let $\varepsilon > 0$ and $m \leqslant c\varepsilon^2 kd/\log^2 k$. Then most m-dimensional subspaces $\mathcal{W} \subset \mathbb{C}^k \otimes \mathbb{C}^d$ have the property that any unit vector $x \in \mathcal{W}$ satisfies*

$$E(x) \geqslant \log k - \frac{1}{2}\frac{k}{d} - \varepsilon.$$

In some cases the result given by Theorem 8.15 can be improved. In particular, in order to obtain violations for the additivity of $S_{\min}$ we will need to produce "extremely entangled subspaces," in which every state has entropy $\log(k) - o(1)$ (see Section 8.4.3).

In the opposite direction, Exercise 8.9 shows an upper bound on the minimal entropy inside *any* subspace of given dimension.

EXERCISE 8.7 (Sharp bounds for the Lipschitz constant of E). In the notation of Lemma 8.12, assume $k \leqslant d$ and let $L = L_k$ be the Lipschitz constant of the function $\psi \mapsto E(\psi)$, calculated with respect to the geodesic distance on $S_{\mathbb{C}^k\otimes\mathbb{C}^d}$ (or on $\mathsf{P}(\mathbb{C}^k \otimes \mathbb{C}^d)$). Show that $L_k \sim \log k$.

EXERCISE 8.8. Show that any s-dimensional subspace $F \subset \mathbb{C}^n$ contains a unit vector x satisfying $\|x\|_\infty \geqslant \sqrt{s/n}$.

EXERCISE 8.9 (An upper bound on the minimal entropy for general subspaces). Let $\mathcal{W} \subset \mathbb{C}^k \otimes \mathbb{C}^d$ be a subspace of dimension αkd, with $\alpha \geqslant 1/k$. (i) Using the previous exercise, show that $\mathcal{W}$ contains a unit vector ψ satisfying $E(\psi) \leqslant h(\alpha) + (1-\alpha)\log(k-1)$, where $h(t) = -t\log t - (1-t)\log(1-t) \leqslant \log 2$ is the binary entropy function. (ii) Conclude that if $\lambda \geqslant 1$ and $E(\psi) \geqslant \log k - \lambda/k$ for all $\psi \in \mathcal{W}$, then $\dim \mathcal{W} = O\big(\lambda d/(1+\log\lambda)\big)$.

8.4.2. Entangled subspaces of small codimension. The argument from the previous section gives nothing for subspaces of dimension cdk or larger: if $\varepsilon = \log d$, the conclusion of Theorem 8.15 does not even imply nonnegativity of $E(x)$. However, in view of Theorem 8.1, it seems plausible to quantify entanglement on subspaces of larger dimension. This can be achieved provided we use a suitable measure of entanglement.

One possibility is to use the p-Rényi entropy for $p = 1/2$. Recall from (8.5) that if we identify a unit vector $x \in \mathbb{C}^k \otimes \mathbb{C}^d$ with $A \in \mathsf{M}_{k,d}$, then

$$E_{1/2}(x) = 2\log \|A\|_1,$$

and our problem becomes a question about the behavior of $\|\cdot\|_1$ vs. $\|\cdot\|_2$ on subspaces of $\mathsf{M}_{k,d}$.

THEOREM 8.16. *Let $k \leqslant d$, and $\mathcal{W} \subset \mathbb{C}^k \otimes \mathbb{C}^d$ be a random subspace of dimension m. The following holds with large probability: for every unit vector $x \in \mathcal{W}$,*

$$E_{1/2}(x) \geqslant \log(k - m/d) - C.$$

The conclusion of Theorem 8.16 yields nontrivial quantitative information for subspaces of codimension larger than $C_1 d$, for some constant C_1. This compares well with Theorem 8.1, which asserts that subspaces of codimension smaller than $d + k - 1$ are *never* fully entangled.

PROOF. We identify $\mathbb{C}^k \otimes \mathbb{C}^d$ with $\mathsf{M}_{k,d}$, and apply the low M^*-estimate (Theorem 7.45) to the norm $\|\cdot\|_1$. One needs the value of $M^* := \mathbf{E}\,\|X\|_{\mathrm{op}}$, where X is uniformly distributed on the Hilbert–Schmidt sphere in $\mathsf{M}_{k,d}$. The inequality $M^* \leqslant C/\sqrt{k}$ follows Proposition 6.36. Denoting $\alpha = 1 - m/kd$, we conclude that for every $A \in \mathcal{W}$,

$$\|A\|_1 \geqslant c\sqrt{k}\sqrt{\alpha}\|A\|_{\mathrm{HS}},$$

and therefore, for every unit vector $x \in \mathcal{W}$ (now seen as a subspace of $\mathbb{C}^k \otimes \mathbb{C}^d$),

$$E_{1/2}(x) = 2\log\|A\|_1 \geqslant \log(k - m/d) - C. \qquad \square$$

8.4.3. Extremely entangled subspaces. In a different direction, we might seek for subspaces of not-so-large dimension, but with near-maximal entropy of entanglement, say $\log k - o(1)$ for example. In view of Lemma 8.13, this requires $k = o(d)$. For simplicity, we will focus on the case $d = k^2$. This choice of dimensions allows us to produce an example of a pair of channels violating the additivity relation (8.8), although the method is applicable to a wider range of parameters.

PROPOSITION 8.17. *There are absolute constants c, C such that the following holds. Let k be an integer and set $d = k^2$, $m = ck^2$. With large probability, a random m-dimensional subspace $\mathcal{W} \subset \mathbb{C}^k \otimes \mathbb{C}^d$ has the property that any unit vector $\psi \in \mathcal{W}$ satisfies*

$$E(\psi) \geqslant \log k - \frac{C}{k}.$$

REMARK 8.18. Proposition 8.17 is optimal in the following sense. First, we cannot hope for larger values of $E(\psi)$ on a random subspace since (by Lemma 8.13) the *global* average value is precisely of order $\log k - \frac{C}{k}$. Second, subspaces of dimension larger than Ck^2 cannot have this property, as shown by Exercise 8.9 (ii).

We start by relating the entropy of very mixed states to their Hilbert–Schmidt distance to the maximally mixed state ρ_* (cf. Lemma 1.20, which leads to a slightly stronger conclusion under stronger hypothesis).

LEMMA 8.19. *If ρ is any state on $\mathbb{C}^k$, then*

$$S(\rho) \geqslant \log k - k \, \|\rho - \rho_*\|_{\mathrm{HS}}^2 .$$

PROOF. The following inequality compares the entropy with its second order approximation: for every $x, t \in [0,1]$,

$$-x \log x \geqslant -t \log t - (1 + \log t)(x - t) - \frac{1}{t}(x - t)^2 . \tag{8.21}$$

To check inequality (8.21), notice that it can be rewritten as $\log(y) \leqslant y - 1$ with $y = x/t$. Given a state $\rho \in \mathrm{D}(\mathbb{C}^k)$ with eigenvalues $(p_i)_{1 \leqslant i \leqslant k}$, we apply (8.21) with $x = p_i$ and $t = 1/k$. Summing over i, we obtain the announced inequality. □

It will be more convenient to work with a random matrix $\mathsf{M} \in \mathsf{M}_{k,d}$ of Hilbert–Schmidt norm 1, rather than with a random unit vector $\psi \in \mathbb{C}^k \otimes \mathbb{C}^d$ (both approaches are equivalent, see Section 0.8). Also recall that when a vector ψ is identified with a matrix M, we have $\mathrm{Tr}_{\mathbb{C}^d} |\psi\rangle\langle\psi| = MM^\dagger$, see (2.12).

Here is a proposition which (via Lemma 8.19) immediately implies Proposition 8.17.

PROPOSITION 8.20. *There are absolute constants c, C such that the following holds. Let k be an integer, $d = k^2$, $m = ck^2$ and let S_{HS} be the Hilbert–Schmidt sphere in $\mathsf{M}_{k,d}$. Consider the function $g : S_{\mathrm{HS}} \to \mathbb{R}$ defined by*

$$g(M) = \left\| MM^\dagger - \frac{\mathrm{I}}{k} \right\|_{\mathrm{HS}} .$$

With large probability, a random m-dimensional subspace $\mathcal{W} \subset \mathsf{M}_{k,d}$ has the property that

$$\sup_{M \in S_{\mathrm{HS}} \cap \mathcal{W}} g(M) \leqslant C/k. \tag{8.22}$$

REMARK 8.21. We wish to point out that while Proposition 8.20 will be *derived from* Dvoretzky for Lipschitz functions, it can be *rephrased* in the language of the standard Dvoretzky's theorem. Indeed, its assertion says that for every $M \in \mathcal{W}$ with $\|M\|_{\mathrm{HS}} = 1$ we have

$$\frac{C^2}{k^2} \geqslant \left\| MM^\dagger - \frac{\mathrm{I}}{k} \right\|_{\mathrm{HS}}^2 = \mathrm{Tr}\,|M|^4 - \frac{2\,\mathrm{Tr}\,MM^\dagger}{k} + \frac{\mathrm{Tr}\,\mathrm{I}}{k^2} = \mathrm{Tr}\,|M|^4 - \frac{1}{k} \geqslant 0. \tag{8.23}$$

Consequently,

$$k^{-1/4}\|M\|_{\mathrm{HS}} \leqslant \|M\|_4 \leqslant k^{-1/4}\Big(1 + \frac{C^2}{k}\Big)^{1/4}\|M\|_{\mathrm{HS}} \leqslant k^{-1/4}\Big(1 + \frac{C^2}{4k}\Big)\|M\|_{\mathrm{HS}} \tag{8.24}$$

for all $M \in \mathcal{W}$. In other words, $\mathcal{W}$ is $(1+\delta)$-Euclidean, with $\delta = \frac{C^2}{4k}$, when considered as a subspace of the Schatten normed space $\big(\mathsf{M}_{k,d}, \|\cdot\|_4\big)$. On the other

hand, the Dvoretzky dimension of $\big(\mathsf{M}_{k,d}, \|\cdot\|_4\big)$ equals $k^{1/2}d$ (see Theorem 7.37) and therefore the general theory (such as Theorem 7.19) gives only $\delta = O(k^{-1/4})$ for m-dimensional subspaces. Although the Dvoretzky dimension is sharp for the size of isomorphically Euclidean subspaces (in the sense exemplified in Exercises 7.12 and 7.25), (8.24) supplies an instance where it can be beaten for almost isometrically Euclidean subspaces.

Before embarking on the proof of Proposition 8.20 we offer some preliminary remarks. We know from Proposition 6.36 (the elementary argument from Exercise 6.43 would actually be sufficient) that all singular values of a typical $M \in S_{\mathrm{HS}}$ belong to the interval

$$\left[\frac{1}{\sqrt{k}} - \frac{C}{\sqrt{d}}, \frac{1}{\sqrt{k}} + \frac{C}{\sqrt{d}}\right]. \tag{8.25}$$

It follows that $\|MM^\dagger - \mathrm{I}/k\|_\infty = O(k^{-3/2})$ and thus the median M_g of g satisfies $M_g \leqslant C/k$. We next estimate the Lipschitz constant of g. The inequality

$$\|MM^\dagger - NN^\dagger\|_{\mathrm{HS}} \leqslant \|M(M^\dagger - N^\dagger) + (M-N)N^\dagger\|_{\mathrm{HS}} \leqslant (\|M\|_{\mathrm{op}} + \|N\|_{\mathrm{op}})\|M-N\|_{\mathrm{HS}}$$

has the following immediate consequence.

LEMMA 8.22. *Let $\Omega_t = \{M \in S_{\mathrm{HS}} \,:\, \|M\|_{\mathrm{op}} \leqslant t\}$ for some $t \geqslant 0$. The function defined on Ω_t by $M \mapsto MM^\dagger$ is $2t$-Lipschitz with respect to the Hilbert–Schmidt norm.*

In particular, the function g is 2-Lipschitz on $\Omega_1 = S_{\mathrm{HS}}$. However, a direct application of Theorem 7.15 yields only a bound of order $1/\sqrt{k}$ in (8.22). (This calculation parallels the one from Remark 8.21 that was expressed in the alternative language of the Dvoretzky dimension.) The trick is to apply concentration of measure twice: to the function g itself, and to the function $f : M \mapsto \|M\|_{\mathrm{op}}$, which is used to control the Lipschitz constant of g.

The function f is 1-Lipschitz on S_{HS}. By (8.25), its median equals $1/\sqrt{k} + O(1/k)$; in particular it is bounded by $2/\sqrt{k}$ for k large enough. Consequently, Lévy's lemma (Corollary 5.17) implies that

$$\mathbf{P}\left(f(M) \geqslant 3/\sqrt{k}\right) \leqslant \frac{1}{2}\exp(-k^2). \tag{8.26}$$

Similarly, an application of the standard Dvoretzky's theorem (Theorem 7.19) to the norm $\|\cdot\|_\infty$ with $\varepsilon = 1/\sqrt{k}$ (note that the dimension of the ambient space is $n = kd$ and that the Dvoretzky dimension is of order d, see Theorem 7.37) shows that a random ck^2-dimensional subspace $\mathcal{W}$ satisfies $S_{\mathrm{HS}} \cap \mathcal{W} \subset \Omega_{3/\sqrt{k}}$ with high probability.

Starting from this point, we will present two possible paths to complete the proof of Proposition 8.20. The first argument uses twice the general Dvoretzky theorem for Lipschitz functions (Theorem 7.15) with the optimal dependence on ε. The second argument is based on a trick due to Fukuda making the overall argument more elementary. In terms of the hierarchy discussed at the beginning of Section 6.1, the first proof we give uses principles from level (ii), namely the Dudley inequality, whereas the second argument uses a single ε-net, staying at level (i).

PROOF #1 OF PROPOSITION 8.20. We know from Lemma 8.22 that the function g is $2t$-Lipschitz on Ω_t. Let $\tilde{g}$ be a $2t$-Lipschitz extension of $g_{|\Omega}$ to S_{HS}. Note

that, in any metric space X, it is possible to extend any L-Lipschitz function h defined on a subset Y without increasing the Lipschitz constant; use, e.g., the formula

$$\tilde{h}(x) = \inf_{y\in Y} \left[h(y) + L\,\mathrm{dist}(x,y)\right].$$

This formula also guarantees that the extended function $\tilde{g}$ is circled. Since $\tilde{g} = g$ on most of S_{HS}, the median of g (resp., $\tilde{g}$) is a central value of $\tilde{g}$ (resp., g). We apply Theorem 7.19 to $\tilde{g}$ with $\varepsilon = 1/k$, $\mu = M_g$ and $L = 2t = 6k^{-1/2}$ to get

$$\sup_{S_{\mathrm{HS}}\cap\mathcal{W}} |\tilde{g} - \mu| \leqslant 1/k$$

on a random subspace $\mathcal{W} \subset \mathsf{M}_{k,d}$ of dimension $m = c_0 \cdot kd \cdot (k^{-1}/(6k^{-1/2}))^2 = cd$. We then have

$$\sup_{S_{\mathrm{HS}}\cap\mathcal{W}} \tilde{g} \leqslant \mu + \frac{1}{k} \leqslant \frac{C'}{k}.$$

If $S_{\mathrm{HS}} \cap \mathcal{W} \subset \Omega$ (which, as noticed before, holds with large probability), g and $\tilde{g}$ coincide on $S_{\mathrm{HS}} \cap \mathcal{W}$ and therefore $g \leqslant C'/k$ on $S_{\mathrm{HS}} \cap \mathcal{W}$, proving (8.22). □

Proof #2 of Proposition 8.20. We use the following lemma which allows to discretize the supremum in (8.22).

Lemma 8.23. *Let $\mathcal{N}$ be an ε-net in $(S_{\mathrm{HS}} \cap \mathcal{W}, |\cdot|)$ with $\varepsilon < \sqrt{2} - 1$. Then*

$$\sup_{M\in S_{\mathrm{HS}}\cap\mathcal{W}} g(M) \leqslant \frac{1}{1-\varepsilon^2-2\varepsilon} \sup_{M\in\mathcal{N}} g(M)$$

Proof of Lemma 8.23. Let $M \in S_{\mathrm{HS}} \cap \mathcal{W}$. There exists $M_0 \in \mathcal{N}$ such that $\delta := \|M - M_0\|_{\mathrm{HS}} \leqslant \varepsilon$. We write $M = M_0 + \delta N$ with $N \in S_{\mathrm{HS}}$, and consider also $A = M_0 + N$ and $B = M_0 - N$ (note that the operators N, A and B all belong to $\mathcal{W}$). One checks that $\|A\|_{\mathrm{HS}}^2 = 2 - \delta$ and $\|B\|_{\mathrm{HS}}^2 = 2 + \delta$. We then set

$$\Delta := MM^\dagger - MM_0^\dagger = \frac{\delta}{2}\left(AA^\dagger - BB^\dagger + 2\delta NN^\dagger\right),$$

and the triangle inequality implies

$$\|\Delta\|_{\mathrm{HS}} \leqslant \frac{\delta}{2}\left(\|AA^\dagger - (2-\delta)\rho_*\|_{\mathrm{HS}} - \|BB^\dagger - (2+\delta)\rho_*\|_{\mathrm{HS}} + \|2\delta NN^\dagger - 2\delta\rho_*\|_{\mathrm{HS}}\right).$$

We can thus estimate

$$\begin{aligned} g(M) &\leqslant g(M_0) + \|MM^\dagger - M_0M_0^\dagger\|_{\mathrm{HS}} \\ &\leqslant g(M_0) + \frac{\delta}{2}\left((2-\delta)g(A/\|A\|_{\mathrm{HS}}) + (2+\delta)g(B/\|B\|_{\mathrm{HS}}) + 2\delta g(N)\right) \\ &\leqslant g(M_0) + (2\delta+\delta^2) \sup_{X\in S_{\mathrm{HS}}\cap\mathcal{W}} g(X). \\ &\leqslant g(M_0) + (2\varepsilon+\varepsilon^2) \sup_{X\in S_{\mathrm{HS}}\cap\mathcal{W}} g(X) \end{aligned}$$

and taking supremum over $M \in S_{\mathrm{HS}}$ gives the result. □

We now return to the proof of the Proposition. The random subspace is realized as $\mathcal{W} = V(\mathbb{C}^m)$ where $V : \mathbb{C}^m \to \mathsf{M}_{k,d}$ is a Haar-distributed isometry. If $\mathcal{M}$ is an ε-net in $(S_{\mathbb{C}^m}, |\cdot|)$, then $\mathcal{N} = V(\mathcal{M})$ is an ε-net in $(S_{\mathrm{HS}} \cap \mathcal{W}, |\cdot|)$. Let us choose (for example) $\varepsilon = 1/3$; by Lemma 5.3, we can ensure that $\operatorname{card}\mathcal{N} \leqslant 36^m$.

We apply the "local Lévy lemma" (Corollary 5.35) to the function g with the subset $\Omega = \Omega_{3/\sqrt{k}} \subset S_{\mathrm{HS}}$ and $\varepsilon = 1/k$. The function $g_{|\Omega}$ is $6/\sqrt{k}$-Lipschitz, and therefore, using (8.26)

$$\mathbf{P}(\{g > M_g + 1/k\}) \leqslant \mathbf{P}(S_{\mathrm{HS}} \subset \Omega) + 2\exp(-d/36) \leqslant C\exp(-cd).$$

Using the union bound and Lemma 8.23, this gives

$$\mathbf{P}\left(\sup_{M \in S_{\mathrm{HS}} \cap \mathcal{W}} g(M) \geqslant \frac{9}{2}(M_g + 1/k)\right) \leqslant 36^m C\exp(-cd)$$

and this quantity is (much) smaller than 1 provided $m \leqslant c'd$, for sufficiently small $c' > 0$. Since $M_g = O(1/k)$, this concludes the proof. □

8.4.4. Counterexamples to the additivity problem. Using Proposition 8.17 and the approach used in Proposition 8.6 for the p-Rényi entropy, we can show the following.

PROPOSITION 8.24. *There is a constant c such that the following holds. Let $d = k^2$, $m = ck^2$ and $\Phi : \mathsf{M}_m \to \mathsf{M}_d$ be a random channel, obtained by (8.6) from a Haar-distributed isometry $V : \mathbb{C}^m \to \mathbb{C}^d \otimes \mathbb{C}^d$. Set $\Psi = \overline{\Phi}$, the channel obtained from $\overline{V}$, the complex conjugate of V. If k is large enough, then with large probability,*

$$S^{\min}(\Phi \otimes \Psi) < S^{\min}(\Phi) + S^{\min}(\Psi).$$

PROOF. Denote by $\mathcal{W} \subset \mathbb{C}^k \otimes \mathbb{C}^d$ the range of V. From Lemma 8.2, we have

$$S_{\min}(\Phi) = \min_{\psi \in \mathcal{W}, |\psi|=1} E(\psi).$$

Note that $S_{\min}(\Phi) = S_{\min}(\Psi)$. From Proposition 8.17, we have with large probability

$$S_{\min}(\Phi) \geqslant \log k - \frac{C}{k}.$$

On the other hand, we know from Lemma 8.7 that applying $\Phi \otimes \overline{\Phi}$ to the maximally entangled state yields an output state with an eigenvalue greater than or equal to $\frac{\dim \mathcal{W}}{\dim \mathsf{M}_{k,d}} = \frac{m}{kd} = \frac{c}{k}$. Then, a simple argument using just concavity of S (see Proposition 1.19) reduces the problem to calculating the entropy of the state with one eigenvalue *equal* to $\frac{c}{k}$ and all the remaining ones identical, which yields

$$S_{\min}(\Phi \otimes \Psi) \leqslant 2\log k - \frac{c\log k}{k} + \frac{1}{k}.$$

We have therefore $S^{\min}(\Phi \otimes \Psi) < S^{\min}(\Phi) + S^{\min}(\Psi)$ provided k is large enough. □

8.5. Entangled pure states in multipartite systems

8.5.1. Geometric measure of entanglement. The definition of the p-entropy of entanglement relies on the Schmidt decomposition, which is specific to the bipartite case. However, the case $p = \infty$ is different since its definition only involves the largest Schmidt coefficient, and this quantity can be defined in a multipartite setting as the square of the maximal overlap with a product vector. In the multipartite setting, the corresponding "∞-entropy of entanglement" has been introduced in the QIT literature via the *geometric measure of entanglement*.

Let $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$ be a multipartite real or complex Hilbert space. Given a unit vector $\psi \in \mathcal{H}$, the geometric measure of entanglement of ψ is defined as

$$g(\psi) = \max\left\{\left|\langle \psi, \psi_1 \otimes \cdots \otimes \psi_k\rangle\right| \ : \ \psi_i \text{ unit vector in } \mathcal{H}_i, 1 \leqslant i \leqslant k\right\} \tag{8.27}$$

(cf. (2.13)) and the ∞-entropy of entanglement is

$$E_\infty(\psi) = -2\log g(\psi). \tag{8.28}$$

We always have $E_\infty(\psi) \geqslant 0$, and $E_\infty(\psi)$ is equal to 0 if and only if ψ is a product vector. Therefore, it makes sense to call unit vectors ψ which maximize $E_\infty(\psi)$ "maximally entangled" vectors. In the bipartite case $\mathbb{C}^d \otimes \mathbb{C}^d$, one recovers the usual notion of a maximally entangled state (see Section 2.2.4). However, in the multipartite case it seems hard to describe the maximally entangled vectors. The problem has an immediate geometric reformulation.

PROPOSITION 8.25 (easy). *Let $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$. The following numbers are equal*
(i) The minimal value of $g(\psi)$ over all unit vectors $\psi \in \mathcal{H}$.
(ii) The inradius of $B_{\mathcal{H}_1} \mathbin{\widehat{\otimes}} \cdots \mathbin{\widehat{\otimes}} B_{\mathcal{H}_k}$, where $B_{\mathcal{H}_i}$ denotes the unit ball in $\mathcal{H}_i$.
(iii) The largest constant c such that any k-linear map $\phi : \mathcal{H}_1 \times \cdots \times \mathcal{H}_k \to \mathbb{C}$ satisfies

$$c|||\phi||| \leqslant \max\{|\phi(x_1, \ldots, x_k)| \ : \ |x_1| \leqslant 1, \ldots, |x_k| \leqslant 1\},$$

where $|||\cdot|||$ denotes the norm

$$|||\phi|||^2 = \sum_{x_1 \in \mathscr{B}_1} \cdots \sum_{x_k \in \mathscr{B}_k} |\phi(x_1, \ldots, x_k)|^2$$

with $\mathscr{B}_i$ an orthonormal basis in $\mathcal{H}_i$ (the value of $|||\cdot|||$ does not depend on the choice of the bases).

Denote by $g_{\min}(\mathcal{H})$ be the common value of the numbers appearing in Proposition 8.25. There is a simple lower bound on $g_{\min}(\mathcal{H})$.

LEMMA 8.26. *If $\mathcal{H} = \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_k}$ or $\mathcal{H} = \mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_k}$ with $d_1 \leqslant \cdots \leqslant d_k$, then $g_{\min}(\mathcal{H}) \geqslant 1/\sqrt{d_1 \cdots d_{k-1}}$. Equivalently, for every unit vector $\psi \in \mathcal{H}$,*

$$E_\infty(\psi) \leqslant \log(d_1) + \cdots \log(d_{k-1}).$$

PROOF OF LEMMA 8.26. The same argument works for the real case and the complex case; we prove the Lemma by induction on k. For $k = 2$, we have

$$g_{\min}(\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}) = \frac{1}{\min(\sqrt{d_1}, \sqrt{d_2})}$$

which is a restatement on the inequalities between the trace norm and the Hilbert–Schmidt norm on the space of $d_1 \times d_2$ matrices. For the induction step, we use the bound (which is again the $k = 2$ case)

$$g_{\min}(\mathbb{C}^{d_1} \otimes \mathcal{H}) \geqslant \frac{1}{\sqrt{d_1}} g_{\min}(\mathcal{H}). \qquad \square$$

8.5.2. The case of many qubits. We will now focus, for simplicity, on the particular case of k qubits, i.e., $d_1 = d_2 = \cdots = d_k = 2$ in the complex case.

In this section it is convenient to define entropy via logarithm to the base 2 and so we will exceptionally use $E_\infty^{(2)}(\psi) := -2\log_2 g(\psi)$ (cf. (8.28)). In this notation, the conclusion of Lemma 8.26 can be rewritten as follows: for any pure state $\psi \in (\mathbb{C}^2)^{\otimes k}$, we have $E_\infty^{(2)}(\psi) \leqslant k - 1$. The following seems to be unknown.

PROBLEM 8.27. *Does there exist a constant C, and for each k a unit vector $\psi \in (\mathbb{C}^2)^{\otimes k}$, such that*

$$E_\infty^{(2)}(\psi) \geqslant k - C \ ?$$

The next proposition shows that random states are typically very entangled, but not entangled enough to give a positive answer to Problem 8.27.

PROPOSITION 8.28. *There exist absolute constants c, C such that a uniformly distributed random unit vector $\psi \in (\mathbb{C}^2)^{\otimes k}$ satisfies with high probability*

$$c\frac{\sqrt{k\log k}}{2^{k/2}} \leqslant g(\psi) \leqslant C\frac{\sqrt{k\log k}}{2^{k/2}}.$$

The conclusion of Proposition 8.28 can be equivalently rewritten as

$$k - \log(k) - \log\log(k) - C' \leqslant E_\infty^{(2)}(\psi) \leqslant k - \log(k) - \log\log(k) + C'.$$

PROOF OF PROPOSITION 8.28. The average of g over the unit sphere is exactly the mean width of $K = (B_{\mathbb{C}^2})^{\hat{\otimes} k}$ (we think of $(\mathbb{C}^2)^{\otimes k}$ as a 2^{k+1}-dimensional real space). The concentration of the functional g around its mean follows from Lévy's lemma (see Table 5.2). Indeed, since K is contained in the unit ball, the functional $g = w(K, \cdot)$ is 1-Lipschitz and therefore

$$\mathbf{P}(|g(\psi) - w(K)| > t) \leqslant 2\exp(-2^k t^2).$$

It remains to show that $w(K) = \Theta(\sqrt{k\log k}\, 2^{-k/2})$, or equivalently that $w_G(K) = \Theta(\sqrt{k\log k})$. The upper bound follows from a standard ε-net argument: let $\mathcal{N}$ be an ε-net in $(S_{\mathbb{C}^2}, |\cdot|)$ with $\operatorname{card}\mathcal{N} \leqslant (2/\varepsilon)^4$ (see Lemma 5.3). From Exercise 5.7 (the weaker result from Lemma 5.9 would be enough here), it follows that $\operatorname{conv}\mathcal{N} \supset (1 - \varepsilon^2/2)B_{\mathbb{C}^2}$. Consequently, denoting by $\mathcal{N}^{\otimes k}$ the set

$$\mathcal{N}^{\otimes k} = \{\psi_1 \otimes \cdots \otimes \psi_k \ : \ \psi_i \in \mathcal{N} \text{ for } 1 \leqslant i \leqslant k\},$$

we have

$$\operatorname{conv}(\mathcal{N}^{\otimes k}) \supset (1 - \varepsilon^2/2)^k K.$$

Using Lemma 6.1, we conclude that

$$w_G(\operatorname{conv}(\mathcal{N}^{\otimes k})) \leqslant \sqrt{2\operatorname{card}(\mathcal{N}^{\otimes k})} \leqslant \sqrt{8k\log(2/\varepsilon)}.$$

Choosing $\varepsilon = 1/\sqrt{k}$ gives the upper bound $w_G(K) = O(\sqrt{k\log k})$.

To show that this argument is sharp, we are going to construct large separated sets in K. Start with a set $\mathcal{M} = \{x_1, \ldots, x_N\}$ which is $1/\sqrt{k}$-separated in the projective space over $\mathbb{C}^2$, with $N = \operatorname{card}(\mathcal{M}) \geqslant ck$. (The estimate on the size of separated sets in $\mathsf{P}(\mathbb{C}^2)$ is an elementary special case of Theorem 5.11 or Exercise 5.10; note that $\mathsf{P}(\mathbb{C}^2)$ identifies with the Bloch sphere, a 2-dimensional Euclidean sphere of radius 1/2, if we use the metric (B.5).) This means that for $i \neq j$, we have $|\langle x_i, x_j\rangle| \leqslant 1 - 1/2k$.

We claim that a large subset of $\mathcal{M}^{\otimes k}$ is separated. To construct it, introduce $Q = \{1, \ldots, N\}^k$, equipped with the normalized Hamming metric, defined for $\alpha, \beta \in Q$ by

$$d(\alpha, \beta) = \frac{1}{k}\operatorname{card}\{i \ : \ \alpha_i \neq \beta_i\}.$$

To each element $\alpha = (\alpha_1, \ldots, \alpha_k) \in Q$ we associate the vector

$$x_\alpha = x_{\alpha_1} \otimes \cdots \otimes x_{\alpha_k} \in K.$$

When $\alpha, \beta \in Q$ are such that $d(\alpha, \beta) \geqslant k/10$, we have

$$|\langle x_\alpha, x_\beta \rangle| = \prod_{j=1}^{k} |\langle x_{\alpha_j}, x_{\beta_j} \rangle| \leqslant (1 - 1/2k)^{k/10} \leqslant c$$

for some constant $c < 1$. We then have $|x_\alpha - x_\beta| \geqslant c' := \sqrt{2 - 2c} > 0$. If we start from a subset $\mathcal{Q} \subset Q$ which is $k/10$-separated, the set $\{x_\alpha \ : \ \alpha \in \mathcal{Q}\}$ is c'-separated in $(\mathbb{C}^2)^{\otimes k}$. By the Sudakov inequality (Proposition 6.10), we have then

$$w_G(K) \geqslant c\sqrt{\log \operatorname{card} \mathcal{Q}}.$$

It remains to give a lower bound on the size of $\mathcal{Q}$. Using the inequality (5.17) from Chapter 5 (which was obtained by the greedy packing algorithm), we obtain $\operatorname{card} \mathcal{Q} \geqslant N^{k(1-H_N(1/5))} \geqslant N^{c''k}$ for some constant $c'' > 0$. It follows that $w_G(K) \geqslant c\sqrt{k \log k}$. □

8.5.3. Multipartite entanglement in real Hilbert spaces. It turns out that in the real case, Lemma 8.26 is surprisingly sharp, so that the real version of Problem 8.27 has a positive answer with $C = 1$. The construction from Proposition 8.29 seems to be specific to the real case. For variants related to Clifford algebras, see Exercise 8.10.

Proposition 8.29. *For any integers $k \geqslant 1$, we have*

$$g_{\min}((\mathbb{R}^2)^{\otimes k}) = 2^{-(k-1)/2}.$$

Proof of Proposition 8.29. The inequality $g_{\min}((\mathbb{R}^2)^{\otimes k}) \geqslant 2^{-(k-1)/2}$ is a consequence of Lemma 8.26. Using Proposition 8.25(iii), the converse inequality will follow provided we show the existence of a k-linear form $\phi : (\mathbb{R}^2)^k \to \mathbb{R}$ such that $|\phi(x_1, \dots, x_k)| \leqslant 1$ for unit vectors $x_1, \dots, x_k$, and $|||\phi||| = 2^{(k-1)/2}$. Let $\theta : \mathbb{R}^2 \to \mathbb{C}$ the canonical isomorphism. It is easily verified that

$$\phi : (x_1, \dots, x_k) \mapsto \operatorname{Re} \prod_{i=1}^{k} \theta(x_i)$$

(where $\prod$ means complex multiplication) satisfies the desired conclusion. □

Exercise 8.10 (Clifford matrices and multipartite maximally entangled states). Given $d \geqslant 2$, let N such that $\mathsf{M}_N(\mathbb{R})$ contains a d-dimensional subspace E in which every matrix is a multiple of an isometry (the smallest possible N is described in Theorem 11.4). Show that

$$g_{\min}((\mathbb{R}^d)^{\otimes k}) \leqslant \frac{\sqrt{N}}{d^{k/2}}. \tag{8.29}$$

When $d \in \{2, 4, 8\}$, one can achieve $N = d$ and the upper bound (8.29) matches the lower bound from Lemma 8.26.

Notes and Remarks

Section 8.1. Theorem 8.1 was proved in [**Wal02, Par04, WS08**]. The statement from Exercise 8.1 is taken from [**CDJ$^+$08**].

Section 8.2. There are multiple operational motivations to use the von Neumann entropy when defining the entropy of entanglement in (8.1). Given a bipartite state ρ, there are several ways to quantify how much entanglement it contains. Two approaches that are in some sense extremal and dual to each other are the entanglement of distillation (the rate at which one can LOCC-transform copies of ρ into Bell states, see also Chapter 12) and the entanglement cost (the rate at which one can LOCC-transform Bell states into copies of ρ). For a general survey on entanglement measures we refer to [**PV07**]. If we restrict ourselves to **pure** states as we do in this chapter, all these entanglement measures coincide with the entropy of entanglement (see Chapter 12.5.2 in [**NC00**].)

The "additivity conjecture" (8.8) has been a major open problem in QIT, particularly since work by Shor [**Sho04**], who showed that the additivity of the minimum output von Neumann entropy was equivalent to the additivity of several other quantities, including the capacity of quantum channels to carry classical information and the entanglement of formation (defined later in Section 10.3.1). For example, the entire ICM 2006 talk by A. Holevo [**Hol06**] was devoted to this circle of ideas. A positive answer would have greatly simplified the theory, leading to a "single letter" formula for the aforementioned capacity, see, e.g., [**Hol06**]. However, the answer to the conjecture was shown to be negative by Hastings [**Has09**].

Exercise 8.3 is based on [**FW07**].

Proposition 8.4 was proved in [**Wat05, Aud09, Sza10**]. We follow here the argument from [**Sza10**].

Our presentation in this chapter barely scratches the surface of the topic of quantum channel capacities. In the quantum context, there are many notions of capacity (see, e.g., [**Wil17**]) and each of them leads to its own class of mathematical questions. For a recent overview of applications of operator space theory to the problem of estimating quantum capacity (i.e., the capacity to carry quantum information) see [**LJL15**].

Section 8.3. The question of the multiplicativity of $\|\cdot\|_{1\to p}$ (8.10) has been considered in [**WH02**] and solved in [**HW08**]. The presentation in the text is based on [**ASW10**], where the connection to Dvoretzky's theorem was noticed. It is also known that $\|\cdot\|_{1\to p}$ is not multiplicative for p close to 0 [**CHL$^+$08**], but part of the range $0 \leqslant p < 1$ is not covered by any approach. The explicit example from Exercise 8.5 comes from [**GHP10**].

Modulo the optimal dependence on the dimension, Theorem 8.9 concerning ε-randomizing channels has been proved in [**HLSW04**]; the parasitic logarithmic factor has been removed in [**Aub09**]. A step towards derandomization has also been made in [**Aub09**], where it was shown that the unitaries in question can be sampled from any Kraus decomposition of the completely randomizing channel.

Section 8.4. Lemma 8.12 appears in [**HLW06**] with the value $C = \sqrt{8}/\log 2$. The argument leading to a better constant ($C_k \sim 1$) in Lemma 8.12 that is sketched in Exercise 8.7 was an unpublished byproduct of the work on [**ASW11**]. For various aspects of continuity of the von Neumann entropy, see [**Win16**].

The exact formula (8.20) from Lemma 8.13 has been conjectured in [**Pag93**] and proved in [**FK94, SR95, Sen96**]. Having the precise form (as opposed to the weaker version stated in Remark 8.14) results in better constants in Theorem 10.16 in Section 10.3.1.

Theorem 8.16 appears to be new.

After Hastings's counterexample to the additivity conjecture [**Has09**] appeared, several papers tried to simplify and extend the original approach, including [**BH10, FKM10, FK10, ASW11, Fuk14**]. We follow mostly [**ASW11**]; Lemma 8.23 and the second proof of Proposition 8.20 are from [**Fuk14**].

A completely different strategy was used in a series of papers initiated by Collins–Nechita [**CN10, CN11**] via free probability and allows to derive results which are more precise in some regimes. Here is a sample theorem from [**BCN12, CFN15**]. *Fix an integer k and $t \in (0,1)$. There is a deterministic convex set $K_{k,t} \subset \mathrm{D}(\mathbb{C}^k)$ with the following property: if $\Phi : \mathsf{M}_m \to \mathsf{M}_k$ is a quantum channel obtained from a random embedding $V : \mathbb{C}^m \to \mathbb{C}^k \otimes \mathbb{C}^d$ with $m = tkd$, then, almost surely as $d \to \infty$, the set $\Phi(\mathrm{D}(\mathbb{C}^m))$ converges to $K_{k,t}$.* This allows, at least in principle, to answer any question about minimal output entropies in this range of parameters. It was subsequently shown in [**BCN16**] that generic channels violating additivity can be obtained by following this strategy if and only if $k \geqslant 183$. Moreover, the defect of non-additivity, i.e., the difference between the two sides of (8.9) is generically almost $\log 2$ for large k (or 1 bit if we use $\log_2$ to define entropy). This improves on the preceding arguments—including the one presented in the text—which showed a violation that was minuscule. Still, in contrast with the Hayden–Winter example [**HW08**] (cf. Remark 8.8), the demonstrated violation does not go to infinity as the dimensions increase. A drawback of the free probability-based method is that the results are valid only when the environment dimension d goes to infinity, and obtaining explicit values of d, for which these asymptotic phenomena hold, requires extra analysis, which is not supplied in [**BCN16**]. For more information on this approach we refer to the survey [**CN16**]. Still another approach, due to Collins [**Col16**] and perhaps more conceptual, relies on the Haagerup inequality about the norms of convolutions on the free group.

In the opposite direction, it is proved in [**Mon13**] that random quantum channels satisfy a weak form of multiplicativity.

Section 8.5. The geometric measure of entanglement was considered under a different terminology in [**Shi95, BL01**] ; see also [**WG03**]. Lemma 8.26 is well-known and appears for example in [**AS06, JHK$^+$08, Arv09**].

We could not locate Problem 8.27 in the literature although it seems a very natural question. It is known that $E_\infty(\psi) < k-1$ for any unit vector $\psi \in (\mathbb{C}^2)^{\otimes k}$ whenever $k \geqslant 3$ (see [**JHK$^+$08**]). The fact that random states are very entangled (the upper bound from Proposition 8.28) has been noticed and used in [**GFE09, BMW09**].

The argument behind Proposition 8.29 and Exercise 8.10 has been communicated to us by Mikael de la Salle (see also Theorem 3.3 in [**Hil07a**]). The papers [**Hil06, Hil07a**] compute also the exact values $g_{\min}((\mathbb{R}^3)^{\otimes 4}) = 1/\sqrt{7}$ and $g_{\min}((\mathbb{R}^3)^{\otimes 4}) = 1/\sqrt{21}$.