The Church of the Symmetric Subspace

Aram W. Harrow[*]

September 2, 2013

Abstract

The symmetric subpace has many applications in quantum information theory. This review article begins by explaining key background facts about the symmetric subspace from a quantum information perspective. Then we review, and in some places extend, work of Werner and Chiribella that connects the symmetric subspace to state estimation, optimal cloning, the de Finetti theorem and other topics. In the third and final section, we discuss how the symmetric subspace can yield concentration-of-measure results via the calculation of higher moments of random quantum states.

There are no new results in this article, but only some new proofs of existing results, such as a variant of the exponential de Finetti theorem. The purpose of the article is (a) pedagogical, and (b) to collect in one place many, if not all, of the quantum information applications of the symmetric subspace.

Contents

[*]Center for Theoretical Physics, MIT. email: `aram@mit.edu`

Variable	Definition
d	local dimension of each subsystem
$[d]$	the set $\{1, \ldots, d\}$
$\vee^n \mathbb{C}^d$	the symmetric subspace of $(\mathbb{C}^d)^{\otimes n}$
$P_{\text{sym}}^{d,n}$	the orthogonal projector onto $\vee^n \mathbb{C}^d$
$d[n]$	$\binom{d+n-1}{n} = \dim \vee^n \mathbb{C}^d = \operatorname{tr} P_{\text{sym}}^{d,n}$
$P_d(\pi)$	$\sum_{i_1,\ldots,i_n \in [d]} \left| i_{\pi^{-1}(1)}, \ldots, i_{\pi^{-1}(n)} \right\rangle \left\langle i_1, \ldots, i_n \right|$
$\mathcal{S}_n$	the symmetric group on n objects
$\mathcal{U}_d$	the group of $d \times d$ unitary matrices
$\mathbb{Z}_+$	nonnegative integers
$\mathcal{I}_{d,n}$	$\{(t_1, \ldots, t_d) : t_1, \ldots, t_d \in \mathbb{Z}_+, t_1 + \ldots + t_d = n\}$
$\binom{n}{t}$	$\frac{n!}{t_1! \ldots t_d!}$
L(V)	linear operators on a vector space V
H(V)	Hermitian operators on V
φ	$\left| \varphi \right\rangle \left\langle \varphi \right|$ (convention used for all pure states)

Table 1: Here is a table of notation, used throughout the notes. For now, you should skip it and go straight to Section 1.

Schur-Weyl duality between the unitary and symmetric groups is a powerful and useful tool in quantum information. But some aspects of it are unsatisfactory. The proofs are rarely fully self-contained, and require excursions into other Lie algebras. At the same time, they involve irreducible representations (irreps) that lack simple, explicit, constructions, making the theory less useful for calculations than one would like. But in many cases, the symmetric subspace is the only necessary piece that needs to be understood. The symmetric subspace is the simplest component of Schur-Weyl duality (with the antisymmetric subspace a close second) and often can be used effectively without the need to ever explicitly invoke representation theory.

In Section 1 of these notes, I will give a self-contained review of the properties of the symmetric subspace. Some applications of the symmetric subspace involve cloning, state estimation and the de Finetti theorem. These are discussed in a unified way by [10], and in Section 2, I will give a brief review of that paper. Another reason to study the symmetric subspace is that it is a way of looking at higher moments of quantum states. In Section 3, I'll explain how this can be used to give alternate and unified derivations of many concentration-of-measure results in quantum information theory. This is the only part of the paper to mostly consist of original work, although even here this consists mostly of new proofs of previously known theorems.

1 The symmetric subspace

One motivation for writing these notes is that there is no comprehensive treatment of the symmetric subspace from the quantum information viewpoint. Ref. [11] covers some of it, Ref. [23] a little less, and Refs. [18, 33] are excellent, but approach the subject respectively from the Lie-algebraic or combinatorial perspective, rather than in terms of quantum information. All of these are really more focused on Schur-Weyl duality in general than the symmetric subspace specifically. Some exceptions are Ref. [5, 27], which are good, but present only as much of the theory as they need for

for their applications. Koenraad Audenaert has also written some nice notes on the representation theory of the symmetric group [4].

Let $\mathcal{S}_n$ be the symmetric group on n letters. For $\pi \in \mathcal{S}_n$, define

$$P_d(\pi) = \sum_{i_1,\ldots,i_n \in [d]} \left| i_{\pi^{-1}(1)}, \ldots, i_{\pi^{-1}(n)} \right\rangle \left\langle i_1, \ldots, i_n \right|.$$

Note that $P_d(\pi_1 \pi_2) = P_d(\pi_1)P_d(\pi_2)$. In other words, P_d is a representation of $\mathcal{S}_n$ on $(\mathbb{C}^d)^{\otimes n}$.

The *symmetric subspace* of $(\mathbb{C}^d)^{\otimes n}$ is denoted $\vee^n \mathbb{C}^d$ and is defined to be

$$\vee^n \mathbb{C}^d = \{ |\psi\rangle \in (\mathbb{C}^d)^{\otimes n} : P_d(\pi) |\psi\rangle = |\psi\rangle \ \forall \, |\psi\rangle \in \mathcal{S}_n \}. \tag{1}$$

(The $\vee$ denotes the symmetric product, by contrast with $\wedge$ which stands for the antisymmetric product, and which we will not discuss here.)

Define

$$P_{\text{sym}}^{d,n} = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} P_d(\pi). \tag{2}$$

Proposition 1. $P_{sym}^{d,n}$ *is the orthogonal projector onto* $\vee^n \mathbb{C}^d$.

Proof. Since group multiplication is invertible, we have that for any $\pi \in \mathcal{S}_n$

$$
\begin{aligned}
P_d(\pi)P_{\text{sym}}^{d,n} &= P_d(\pi)\frac{1}{n!} \sum_{\pi' \in \mathcal{S}_n} P_d(\pi') \\
&= \frac{1}{n!} \sum_{\pi' \in \mathcal{S}_n} P_d(\pi\pi') \\
&= \frac{1}{n!} \sum_{(\pi^{-1}\pi') \in \mathcal{S}_n} P_d(\pi') \\
&= \frac{1}{n!} \sum_{\pi' \in \mathcal{S}_n} P_d(\pi') = P_{\text{sym}}^{d,n}.
\end{aligned}
\tag{3}
$$

Similarly, we have $P_{\text{sym}}^{d,n} P_d(\pi) = P_{\text{sym}}^{d,n}$.

This implies that

$$(P_{\text{sym}}^{d,n})^\dagger P_{\text{sym}}^{d,n} = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} P_d(\pi^{-1}) P_{\text{sym}}^{d,n} = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} P_{\text{sym}}^{d,n} = P_{\text{sym}}^{d,n}.$$

Therefore $P_{\text{sym}}^{d,n}$ is an orthogonal projector, since $\Pi^\dagger \Pi = \Pi$ is a necessary and sufficient condition for an operator Π to be an orthogonal projector.

We also use (3) to show that for any $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$,

$$P_d(\pi)P_{\text{sym}}^{d,n} |\psi\rangle = P_{\text{sym}}^{d,n} |\psi\rangle .$$

Thus $P_{\text{sym}}^{d,n} |\psi\rangle \in \vee^n \mathbb{C}^d$, and we have that $\text{Im}\, P_{\text{sym}}^{d,n} \subseteq \vee^n \mathbb{C}^d$.

To show that $\vee^n \mathbb{C}^d \subseteq \text{Im}\, P_{\text{sym}}^{d,n}$, we observe that if $|\psi\rangle \in \vee^n \mathbb{C}^d$ then $P_{\text{sym}}^{d,n} |\psi\rangle = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} P_d(\pi) |\psi\rangle = |\psi\rangle$. $\qquad \square$

The alert reader will notice that almost no properties of $\mathcal{S}_n$ were used in the above proof. We can generalize Proposition 1 to a large class of groups. The necessary condition is that a group G should have an invariant measure μ. That is, for any integrable function $f : G \to \mathbb{C}$, and any $g \in G$, we have $\int_{x \in G} \mu(x) f(x) \mathrm{d}x = \int_{x \in G} \mu(x) f(gx) \mathrm{d}x$. Such measures exist for all finite groups (take $\mu(x) = 1/|G|$ and replace the integral by a sum) and for all compact Lie groups, such as the unitary group. In the latter case, there is a unique measure (up to normalization) called the Haar measure.

For a vector space V, define $L(V)$ to be the set of linear operators on V.

Proposition 2. *Let G be a group with an invariant measure μ, and a representation $R : G \to L(V)$. Define*

$$V^G := \{|\psi\rangle \in V : R(g)|\psi\rangle = |\psi\rangle \, \forall g \in G\} \tag{4}$$

$$\Pi := \int_{x \in G} \mathrm{d}x \mu(x) R(x) \tag{5}$$

Then Π is an orthogonal projector onto V^G.

We omit the proof, as it follows the same lines as that of Proposition 1.

We now return to our discussion of the symmetric subspace, and give two equivalent characterizations of $\vee^n \mathbb{C}^d$. First define

$$A = \mathrm{span}\{|\varphi\rangle^{\otimes n} : |\varphi\rangle \in \mathbb{C}^d\}.$$

Second, let $\mathbb{Z}_+$ denote nonnegative integers. Let $\mathcal{I}_{d,n} = \{(t_1, \ldots, t_d) : t_1, \ldots, t_d \in \mathbb{Z}_+, t_1 + \ldots + t_d = n\}$. For $\vec{t} \in \mathcal{I}_{d,n}$ we abbreviate the multinomial coefficient $\frac{n!}{t_1! \ldots t_d!}$ by $\binom{n}{\vec{t}}$. For $\vec{i} = (i_1, \ldots, i_n) \in [d]^n$, the *type* of $\vec{i}$ is denoted $T(\vec{i})$ and defined to be the vector in $\mathcal{I}_{d,n}$ whose j^{th} entry is the number of times that j appears in the string $(i_1, \ldots, i_n)$. Note that $|T^{-1}(\vec{t})| = \binom{n}{\vec{t}}$. Now define

$$|s_{\vec{t}}\rangle := \sqrt{\binom{n}{\vec{t}}} \sum_{\vec{i} : T(\vec{i}) = \vec{t}} |i_1, \ldots, i_n\rangle.$$

Finally we can define the subspace

$$B = \mathrm{span}\{|s_{\vec{t}}\rangle : \vec{t} \in \mathcal{I}_{d,n}\}.$$

We can now state our main theorem about the structure of $\vee^n \mathbb{C}^d$.

Theorem 3.
$$\vee^n \mathbb{C}^d = A = B.$$

Proof. Since A and B are both spanned by sets of vectors that are individually invariant under $P_d(\pi)$, it follows that $A \subseteq \vee^n \mathbb{C}^d$ and $B \subseteq \vee^n \mathbb{C}^d$.

To show that $\vee^n \mathbb{C}^d \subseteq B$, we note that $P_{\text{sym}}^{d,n} |i_1, \ldots, i_n\rangle = \binom{n}{T(\vec{i})}^{-1/2} |s_{T(\vec{i})}\rangle$ and therefore $\mathrm{Im}\, P_{\text{sym}}^{d,n} \subseteq B$. Since $\vee^n \mathbb{C}^d = \mathrm{Im}\, P_{\text{sym}}^{d,n}$, we conclude that $\vee^n \mathbb{C}^d = B$.

The last step is to show that $B \subseteq A$. Here we use polynomials in a clever way. Suppose $p(x) = v_0 + x v_1 + \ldots + x^d v_d$, for $v_0, \ldots, v_d$ vectors in a finite-dimensional space V. For some subspace $W \subset V$, suppose that $p(x) \in W$ for all x. Then I claim that $v_0, \ldots, v_d \in W$. The proof is

that derivatives of $p(x)$ can be expressed as limits of linear combinations of $p(x)$ for different values of x, and therefore are all contained in W. Then we use the fact that $v_k = \frac{1}{k!} \frac{\partial^k}{\partial x^k} p(x)|_{x=0}$.

By induction on the number of variables, we can extend this to argue that if $p(x_1, \ldots, x_d) \in W$ for all $x_1, \ldots, x_d$ then the coefficient of $x_1^{t_1} \ldots x_d^{t_d}$ must be in W for each $t_1, \ldots, t_d \in \mathbb{Z}_+^d$.

Now, consider the polynomial $|p(x_1, \ldots, x_d)\rangle := (\sum_{i=1}^d x_i |i\rangle)^{\otimes n}$. Since $|p(x_1, \ldots, x_d)\rangle$ is a tensor power state, it belongs to A. The coefficient of $x_1^{t_1} \ldots x_d^{t_d}$ in $|p(x_1, \ldots, x_d)\rangle$ is proportional to $|s_{\vec{t}}\rangle$. Therefore $|s_{\vec{t}}\rangle \in A$ for all $\vec{t} \in \mathcal{I}_{d,n}$. We conclude that $B \subseteq A$. $\qquad\square$

One interesting consequence of Theorem 3 is obtained by replacing $\mathbb{C}^d$ with M_d, the vector space of $d \times d$ matrices. The symmetric group acts on $M_d^{\otimes n}$ by conjugation, with π sending $M \in M_d^{\otimes n}$ to $P_d(\pi) M P_d(\pi)^\dagger$.

Corollary 4. *If $M \in M_d^{\otimes n}$, then $[M, P_d(\pi)] = 0$ for all $\pi \in \mathcal{S}_n$ if and only if $M \in \text{span}\{X^{\otimes n} : X \in M_d\}$.*

Corollary 4 is like a baby de Finetti theorem (of which we will discuss more in Section 2). It says that a permutation-invariant state ρ can be written as $\sum a_i X_i^{\otimes n}$. Unfortunately, a_i and X_i do not have to be positive, making this decomposition less useful.

Another natural way to understand $\vee^n \mathbb{C}^d$ is in terms of representation theory.

Theorem 5. $\vee^n \mathbb{C}^d$ *is an irreducible represention of $\mathcal{U}_d$ under the action $U \mapsto U^{\otimes n}$.*

This fact is often proved using facts about the irreducible representations of Lie algebras. To keep things self-contained, I will give an elementary proof using ideas familiar to quantum information.

Proof. Consider an arbitrary pair of unit vectors $|\psi_1\rangle, |\psi_2\rangle \in \vee^n \mathbb{C}^d$. We will demonstrate the existence of a $U \in \mathcal{U}_d$ such that $\langle \psi_1| U^{\otimes n} |\psi_2\rangle \neq 0$. Equivalently, we will choose a probability distribution over U satisfying

$$\mathbb{E}_U \langle \psi_1| U^{\otimes n} |\psi_2\rangle \neq 0. \tag{6}$$

To this end, choose unit vectors $|\varphi_1\rangle, |\varphi_2\rangle \in \mathbb{C}^d$ such that $\langle \psi_i| \cdot |\varphi_i\rangle^{\otimes n} \neq 0$ for $i = 1, 2$. By Theorem 3, these vectors must exist. Then, for $i = 1, 2$, choose V_i uniformly at random from the set of unitaries satisfying $V_i |\varphi_i\rangle = |\varphi_i\rangle$. Such unitaries can be constructed by choosing a random element of $\mathcal{U}_{d-1}$ and embedding it in the space orthogonal to $|\varphi_i\rangle$. Since $\mathbb{E}_{V_i} V_i^{\otimes n}$ is an average over a group action, Proposition 2 implies that it is a projector onto the set of vectors fixed by each $V_i^{\otimes n}$; in particular,

$$\mathbb{E}_{V_i} V_i^{\otimes n} = |\varphi_i\rangle \langle \varphi_i|^{\otimes n}.$$

Finally, choose $W \in \mathcal{U}_d$ to be a unitary satisfying $W |\varphi_2\rangle = |\varphi_1\rangle$, and let $U = V_1^\dagger W V_2$. Then $\mathbb{E}_U U^{\otimes n} = |\varphi_1\rangle \langle \varphi_2|^{\otimes n}$, and since we have assumed $|\psi_i\rangle$ has nonzero overlap with $|\varphi_i\rangle^{\otimes n}$ for $i = 1, 2$, we obtain (6). $\qquad\square$

Using Theorem 5 and Schur's Lemma gives us another characterization of $P_{\text{sym}}^{d,n}$.

Proposition 6.

$$\mathbb{E}_\varphi \varphi^{\otimes n} = \frac{P_{\text{sym}}^{d,n}}{\text{tr } P_{\text{sym}}^{d,n}} = \frac{P_{\text{sym}}^{d,n}}{d[n]} = \frac{\sum_{\pi \in \mathcal{S}_n} P_d(\pi)}{d(d+1)\cdots(d+n-1)}. \tag{7}$$

5

Here $\mathbb{E}_\varphi$ means that we average over a randomly chosen unit vector $|\varphi\rangle \in \mathbb{C}^d$.

Proof. Observe that $\rho := \mathbb{E}_\varphi \, \varphi^{\otimes n}$ commutes with all $U^{\otimes n}$. Thus, by Schur's Lemma and Theorem 5, ρ must be proportional to the identity operator on that space, which is $P_{\text{sym}}^{d,n}$. To find the normalization, we observe that $\text{tr}\,\rho = 1$. □

One consequence of Proposition 6 is that

$$\{d[n] \, |\varphi\rangle \, \langle\varphi|^{\otimes n} \, d\varphi\} \tag{8}$$

forms a continuous POVM.

Another consequence is that averaging $\hat{\varphi}^{\otimes n}$ over Gaussian vectors $|\hat{\varphi}\rangle$ gives an operator proportional to a projector. If the normalization (and covariance) of the Gaussian is chosen so that $\mathbb{E}_{\hat{\varphi}} \, \hat{\varphi} = I/d$, then we can show that

$$\mathbb{E}_{\hat{\varphi}} \, \hat{\varphi}^{\otimes n} = \frac{n!}{d^n} P_{\text{sym}}^{d,n} = d^{-n} \sum_{\pi \in \mathcal{S}_n} P_d(\pi). \tag{9}$$

How? Well, let $|\hat{\varphi}\rangle = x \, |\varphi\rangle$, for $x \in \mathbb{R}_+$ and $|\varphi\rangle$ a unit vector. Because of the rotational invariance of the Gaussian distribution, it follows that x and $|\varphi\rangle$ are independent random variables. Thus $\mathbb{E}_{\hat{\varphi}} \, \hat{\varphi}^{\otimes n} = \mathbb{E}_x \, |x|^{2n} \, \mathbb{E}_\varphi \, \varphi^{\otimes n}$. It remains only to compute $\mathbb{E}_x \, |x|^{2n}$. To do so, let $a_j := \langle j|\hat{\varphi}\rangle$ so that $|x|^2 = |a_1|^2 + \cdots + |a_d|^2$. Next, we recall the formula for a Gaussian integral:

$$\int_{a \in \mathbb{C}^d} \mathrm{d}a \, e^{-\alpha(|a_1|^2 + \cdots + |a_d|^2)} = (\pi/\alpha)^d. \tag{10}$$

We can use this to calculate $\mathbb{E}_x \, |x|^{2n}$ by differentiating (10) with respect to $-\alpha$ n times, then dividing by the normalization $(\pi/\alpha)^d$, and finally setting $\alpha = d$. This yields $\mathbb{E}_x \, |x|^{2n} = (1 + \frac{1}{d}) \cdots (1 + \frac{n-1}{d})$. Combining with (7), we obtain (9).

Alternatively, (9) can be derived directly (and is sometimes called Wick's theorem[1]), and then used to obtain (7). See [17] for a nice exposition of this approach.

1.1 Operators on the symmetric subspace

For a complex vector space V, let $L(V)$ denote the space of operators on V and $H(V)$ the space of Hermitian operators on V. Given that $\vee^n \mathbb{C}^d$ is spanned by vectors of the form $|\varphi\rangle^{\otimes n}$, can we say something similar about $L(\vee^n \mathbb{C}^d)$ and $H(\vee^n \mathbb{C}^d)$?

Happily, in this case the matrices $\varphi^{\otimes n}$ play the same role, and we have

$$L(\vee^n \mathbb{C}^d) = \text{span}_{\mathbb{C}}\{\varphi^{\otimes n} : |\varphi\rangle \in \mathbb{C}^d\} \tag{11a}$$

$$H(\vee^n \mathbb{C}^d) = \text{span}_{\mathbb{R}}\{\varphi^{\otimes n} : |\varphi\rangle \in \mathbb{C}^d\}. \tag{11b}$$

In both cases, the RHS is trivially contained in the LHS. Conversely, the LHS of (11a) can be expressed as a span of operators of the form $(|\alpha\rangle \langle\beta|)^{\otimes n}$. To express $(|\alpha\rangle \langle\beta|)^{\otimes n}$ as a linear combination of terms on the RHS of (11a), define $|v_{x,y}\rangle = e^{ix} |\alpha\rangle + e^{iy} |\beta\rangle$ and note

$$(|\alpha\rangle \langle\beta|)^{\otimes n} = \frac{1}{(2\pi)^2} \int_0^{2\pi} \mathrm{d}x \int_0^{2\pi} \mathrm{d}y \, e^{in(y-x)} (|v_{x,y}\rangle \langle v_{x,y}|)^{\otimes n} \tag{12}$$

---

[1] The way to do this calculation is to calculate the integral of $\exp(-\sum_{i=1}^d \alpha_i |a_i|^2)$ and differentiate with respect to various α_i.

Similarly for $H(\vee^n \mathbb{C}^d)$, the LHS is spanned by the operators $(|\alpha\rangle\langle\beta|)^{\otimes n} + (|\beta\rangle\langle\alpha|)^{\otimes n}$. We write this operator as a real linear combination of terms from the RHS of (11b) as follows:

$$(|\alpha\rangle\langle\beta|)^{\otimes n} + (|\beta\rangle\langle\alpha|)^{\otimes n} = \frac{1}{(2\pi)^2} \int_0^{2\pi} dx \int_0^{2\pi} dy\, (e^{in(y-x)} + e^{-in(y-x)})(|v_{x,y}\rangle\langle v_{x,y}|)^{\otimes n} \qquad (13)$$

1.2 The real case

What about $\vee^n \mathbb{R}^d$? Things are now totally different. Let $|\gamma\rangle \in \mathbb{R}^d$ be a random unit vector and $|\hat{\gamma}\rangle \in \mathbb{R}^d$ be a random Gaussian vector with $\mathbb{E}\langle\hat{\gamma}|\hat{\gamma}\rangle = 1$. To describe $\mathbb{E}\,\hat{\gamma}^{\otimes n}$, we introduce some more notation. Let $\mathcal{M}_{2n}$ be the set of perfect matchings on $[2n]$; i.e. n disjoint subsets of $[2n]$, each containing two elements. Say that a string $i_1, \ldots, i_{2n} \in [d]$ is compatible with $M \in \mathcal{M}_{2n}$ if $i_j = i_k$ for each $\{j, k\} \in M$. Let S_M denote the set of $i_1, \ldots, i_{2n}$ that are compatible with M, and define $\sigma_M := \sum_{(i_1,\ldots,i_{2n}) \in S_M} |i_1, \ldots, i_n\rangle\langle i_{n+1}, \ldots, i_{2n}|$. Note that $P_d(\pi) = \sigma_{M_\pi}$ where we define $M_\pi := \{(1, n+\pi(1)), \ldots, (n, n+\pi(n))\}$; however, other matchings do not correspond to any permutation. The moments of $\hat{\gamma}$ are then given by

$$\mathbb{E}\,\hat{\gamma}^{\otimes n} = d^{-n} \sum_{M \in \mathcal{M}_{2n}} \sigma_M. \qquad (14)$$

I'll skip the derivation, as it's similar to the complex case.

As an example, when $n = 2$, then $\mathbb{E}\,\hat{\gamma}^{\otimes 2} = \frac{I + \text{SWAP}}{d^2} + \frac{\Phi}{d}$, where $|\Phi\rangle = d^{-1/2} \sum_{i=1}^d |i, i\rangle$ is a maximally entangled state. The $|\Phi\rangle$ term is new, and dramatically increases the largest eigenvalue of resulting matrix. To see why it appears, consider a simple (univariate) Gaussian variable x. If x is a complex Gaussian, then $\mathbb{E}\,x^2 = \mathbb{E}\,\bar{x}^2 = 0$, and only $\mathbb{E}\,x\bar{x}$ is nonzero. However, if x is a real Gaussian, then $\mathbb{E}\,x^2$ is nonzero. This means that there additional terms, corresponding to matchings not of the form M_π, that contribute to terms like Φ. The reason that these terms lead to higher eigenvalues is that they don't distinguish between row and column indices, and so are not adapted to the matrix structure.

For unit vectors, the situation is similar except for the overall normalization, which is described by a higher moment of a χ^2-distribution. After a (skipped) calculation, one obtains

$$\mathbb{E}\,\gamma^{\otimes n} = \left(\frac{d}{2}\right)^n \frac{\Gamma(d/2)}{\Gamma(n+d/2)} \mathbb{E}\,\hat{\gamma}^{\otimes n} = \frac{1}{2^n} \frac{\Gamma(d/2)}{\Gamma(n+d/2)} \sum_{M \in \mathcal{M}_{2n}} \sigma_M. \qquad (15)$$

Here $\Gamma(z)$ is the gamma function, equal to $z-1!$ for integer z, and $\frac{(2z-1)!}{4^z(z-1/2)!}\sqrt{4\pi}$ for half-integer z. As in the complex case, random unit vectors resemble random Gaussians when n is small relative to d.

2 Estimation, cloning and the de Finetti theorem

This section discusses three important applications of the symmetric subspace, essentially following the treatment of [10], but with some of the material on cloning from [36].

Consider the following three problems.

1. *State estimation:* Measure $|\varphi\rangle^{\otimes n}$ to obtain an estimate $|\hat{\varphi}\rangle$. Try to maximize $\mathbb{E}_\varphi\,|\langle\varphi|\hat{\varphi}\rangle|^{2k}$ for some k.

2. *Cloning:* Construct a map T from n qudits to $n+k$ qudits that maximizes $\mathbb{E}_\varphi \operatorname{tr} \varphi^{\otimes n+k} T(\varphi^{\otimes n})$.

3. *de Finetti:* Given $|\psi\rangle \in \vee^n \mathbb{C}^d$, how well is $\operatorname{tr}_{n-k} \psi$ approximiated by a mixture of tensor power states?

It turns out that there are close relations between these problems.

2.1 Estimation and Measure-and-prepare channels

Start with state estimation. Note that $|\langle\varphi|\hat\varphi\rangle|^{2k} = \operatorname{tr} \varphi^{\otimes k} \hat\varphi^{\otimes k}$. The most general strategy possible (more or less) is to perform the POVM with measurement operators $M_1, \ldots, M_\ell$ (with $M_1 + \cdots + M_\ell = P_{\text{sym}}^{d,n}$), and upon outcome i, to output the estimate $|\hat\varphi_i\rangle$. Let $\rho_i = \hat\varphi_i^{\otimes k}$. Then

$$F_{\text{estimate}} = \mathbb{E}_\varphi \sum_{i=1}^\ell \operatorname{tr}(\varphi^{\otimes n} M_i) \operatorname{tr}(\varphi^{\otimes k} \rho_i) \tag{16}$$

$$= \mathbb{E}_\varphi \sum_{i=1}^\ell \operatorname{tr} \varphi^{\otimes n+k} (M_i \otimes \rho_i) \tag{17}$$

$$= \operatorname{tr} \frac{P_{\text{sym}}^{d,n+k}}{d[n+k]} \sum_{i=1}^\ell (M_i \otimes \rho_i) \tag{18}$$

$$\leq \frac{\sum_{i=1}^\ell \operatorname{tr}(M_i \otimes \rho_i)}{d[n+k]} \tag{19}$$

$$= \frac{d[n]}{d[n+k]} \tag{20}$$

On the other hand, (20) is achieved by using the continuous POVM from (8). Why? We replace (18) with

$$\operatorname{tr} \frac{P_{\text{sym}}^{d,n+k}}{d[n+k]} \mathbb{E}_{\hat\varphi} d[n] \hat\varphi^{\otimes n} \otimes \hat\varphi^{\otimes k} = \frac{d[n]}{d[n+k]}. \tag{21}$$

This analysis has also yielded the solution to a related problem, which is to find the optimal "measure-and-prepare" channel mapping $|\varphi\rangle^{\otimes n}$ to an approximation of $|\varphi\rangle^{\otimes k}$. Measure-and-prepare channels are of the form

$$T(\sigma) = \sum_i \operatorname{tr}(M_i \sigma) \rho_i \tag{22}$$

and are also called "entanglement-breaking" channels [25], because it turns out that the form (22) is equivalent to the condition that $(T \otimes I)$ maps all states to separable states. The optimal measure-and-prepare channel is denoted $\text{MP}_{n \to k}$ and is

$$\text{MP}_{n \to k}(\rho) = \operatorname{tr}_n \mathbb{E}_\varphi d[n] \varphi^{\otimes n+k} (\rho \otimes I^{\otimes k}) = \operatorname{tr}_n \frac{d[n]}{d[n+k]} P_{\text{sym}}^{d,n+k} (\rho \otimes I^{\otimes k}). \tag{23}$$

2.2 Optimal cloning

The no-cloning theorem says that $|\varphi\rangle \to |\varphi\rangle \otimes |\varphi\rangle$ is impossible. But in fact $|\varphi\rangle^{\otimes n} \to |\varphi\rangle^{\otimes n+k}$ is also impossible for any $n, k > 0$. Still, we can try to approximate this map. It turns out that the

optimal cloning map (due to [36]) is

$$\text{Clone}_{n\to n+k}(\rho) = P_{\text{sym}}^{d,n+k}(\rho \otimes I^{\otimes k})P_{\text{sym}}^{d,n+k}\frac{d[n]}{d[n+k]}. \tag{24}$$

Note that normally $\rho = \varphi^{\otimes n}$.

(24) is rather remarkable. At first, it's not even obvious that Clone_n^{n+k} is trace-preserving, but this can be deduced with the help of Corollary 4. Optimality takes more work, but is similar in spirit to the optimality of the above estimation procedure.

The main theorem of [10] gives a relation between MP and Clone. Specifically

Theorem 7 (Chiribella's theorem[10]).

$$\text{MP}_{n\to k}(\rho) = \sum_{s=0}^{k} \frac{\binom{n}{s}\binom{d+k-1}{k-s}}{\binom{d+n+k-1}{k}} \text{Clone}_{s\to k}(\text{tr}_{n-s}\rho) \tag{25}$$

We give a slightly simpler proof than the one in [10].

Proof. First, we observe (following [10]) that the space of density matrices on $\vee^n \mathbb{C}^d$ is spanned by vectors of the form $|\varphi\rangle\langle\varphi|^{\otimes n}$. Thus, to compute the action of $\text{MP}_{n\to k}$, it suffices to calculate

$$f(\alpha, \beta) := \text{tr}\,\beta^{\otimes k}\,\text{MP}_{n\to k}(\alpha^{\otimes n})$$

for all unit vectors $|\alpha\rangle, |\beta\rangle \in \mathbb{C}^d$. Further, the unitary covariance of quantities involved means that $f(\alpha, \beta)$ depends only on the scalar $x := \text{tr}\,\alpha\beta$. Let $f(x) := f(\alpha, \beta)$.

Using the definition in (23), we see that

$$f(x) = \frac{d[n]}{d[n+k]}\text{tr}(I^{\otimes n} \otimes \beta^{\otimes k})P_{\text{sym}}^{d,n+k}(\alpha^{\otimes n} \otimes I^{\otimes k}) \tag{26a}$$

$$= \frac{d[n]}{d[n+k]}\text{tr}\,P_{\text{sym}}^{d,n+k}(\alpha^{\otimes n} \otimes \beta^{\otimes k}) \tag{26b}$$

$$= \frac{d[n]}{d[n+k]}\sum_{s=0}^{k}\frac{\binom{k}{s}\binom{n}{s}}{\binom{n+k}{k}}x^s \tag{26c}$$

The term on the last line is the probability that a random $\pi \in \mathcal{S}_{n+k}$ satisfies $|\pi([n]) \cap [n]| = s$. This is a hypergeometric distribution, equivalent to the probability that when n balls are drawn without replacement from a bucket of n white balls and k black balls, that the resulting sample contains $n-s$ white balls and s black balls.

On the other hand, to analyze the RHS of (25), we calculate

$$\text{tr}\,\beta^{\otimes k}\,\text{Clone}_{s\to k}(\text{tr}_{n-s}\,\alpha^{\otimes n}) = \text{tr}\,\beta^{\otimes k}\,\text{Clone}_{s\to k}(\alpha^{\otimes s}) \tag{27a}$$

$$= \text{tr}\,\beta^{\otimes k}P_{\text{sym}}^{d,k}(\alpha^{\otimes s} \otimes I^{\otimes k-s})P_{\text{sym}}^{d,k}\frac{d[s]}{d[k]} \tag{27b}$$

$$= \frac{d[s]}{d[k]}x^s \tag{27c}$$

9

Finally we calculate

$$\frac{d[n]d[k]}{d[n+k]d[s]} \frac{\binom{k}{s}\binom{n}{s}}{\binom{n+k}{k}} = \frac{\binom{d+k-1}{k}\binom{k}{s}}{\binom{d+s-1}{s}} \cdot \binom{n}{s} \cdot \frac{\binom{d+n-1}{n}}{\binom{d+n+k-1}{n+k}\binom{n+k}{k}} \tag{28a}$$

$$= \frac{\binom{d+k-1}{k-s}\binom{n}{s}}{\binom{d+n+k-1}{k}} \tag{28b}$$

Combining (26), (27) and (28), we obtain (25). $\qquad\square$

Inspired by Chiribella's theorem, we define the polynomials

$$M_k^{(d,n)}(x) = \sum_{s=0}^{k} \frac{\binom{n}{s}\binom{d+k-1}{k-s}}{\binom{d+n+k-1}{k}} x^s = \sum_{s=0}^{k} M_{k,s}^{(d,n)} x^s \tag{29}$$

The coefficients $M_{k,s}^{(d,n)}$ correspond to a hypergeometric distribution whose moment-generating function is given by $M_k^{(d,n)}(e^t)$.

We observe that these polynomials can be described in terms of Jacobi polynomials as

$$M_k^{(d,n)}(x) = \frac{(x-1)^k}{\binom{d+n+k-1}{k}} P_k^{(n-k,d-1)}\left(\frac{x+1}{x-1}\right), \tag{30}$$

and so are orthogonal with respect to the weight $(1-y)^\alpha(1+y)^\beta dy$ over $y \in [-1,1]$, where $y = (x+1)/(x-1)$, $\alpha = n-k$ and $\beta = d-1$. Unfortunately, if $x \in [0,1]$, then $y \le -1$, so this standard interpretation of the Jacobi polynomials appears not to apply. Similarly, the interpretation of Jacobi polynomials as matrix elements of irreps of $\mathcal{U}_2$ only applies directly when the argument is in the range $[-1,1]$. Jacobi polynomials have previously appeared in analysis of de Finetti errors in [30], and it is possible that an alternate derivation of (29) might proceed via the representation theory of $\mathcal{U}_d$ rather than $\mathcal{S}_n$. Similar polynomials have also been analyzed in terms of functions of two variables [21].

2.3   de Finetti theorem

What's the point of all these expansions? Who cares if we can shuffle a bunch of permutations around and relate one thing that we didn't care that much about (the $n \to k$ measure-and-prepare channel) to the far more obscure task of choosing s from a hypergeometric distribution, tracing out all but s subsystems, and cloning back up to k?

One application mentioned in [10] is to give an alternate proof of the de Finetti theorem. Observe that

$$M_{k,k}^{(d,n)} = \frac{\binom{n}{k}}{\binom{d+n+k-1}{k}} = \frac{n!d+n-1!}{n-k!d+n+k-1!} \ge \left(1 - \frac{d+k}{n+d}\right)^k \ge 1 - \frac{k(d+k)}{n+d} \tag{31}$$

Thus, (25) implies that

$$\mathrm{MP}_{n\to k} = (1-\epsilon)\operatorname{tr}_{n-k} + \epsilon\mathcal{N}, \tag{32}$$

where $\epsilon \le k(d+k)/(n+d)$ (assuming this quantity is ≤ 1) and $\mathcal{N}$ is a trace-preserving quantum operation. Thus

$$\|\mathrm{MP}_{n\to k} - \operatorname{tr}_{n-k}\|_\diamond \le 2\epsilon. \tag{33}$$

This establishes the de Finetti theorem in an elegant form: given a symmetric state on n qudits, tracing out $n - k$ qudits yields a state that is within 2ϵ of a mixture of tensor powers. The advantages of this formula are that it is concise, it naturally handles the case of symmetric states that are entangled with reference systems and it gives an explicit description of how to produce the approximation.

In fact, this approach can also yield the so-called exponential de Finetti theorem of [31, 28]. This is the only original result in this section.

To introduce the exponential de Finetti theorem, we need the idea of an "almost-product state" introduced in [31] (see also [32]). Define the (k, r, d)-almost product states to be

$$
\bigcup_{|\varphi\rangle \in \mathbb{C}^d} \mathrm{span}\{P_{\mathrm{sym}}^{d,k} |\varphi\rangle^{\otimes k-r} \otimes |\psi\rangle : |\psi\rangle \in (\mathbb{C}^d)^{\otimes r}\}.
$$

This set is *not* a linear subspace, but see [28] for a discussion of almost-product states from a representation-theoretic perspective. Note that the set of almost-product states has no real classical analogue, and indeed the exponential de Finetti theorem (stated below) fails in the classical case.[2] Observe that $\mathrm{Clone}_{k-s\to k}$ maps product states to (k, s, d)-almost-product states, and thus $\mathrm{Clone}_{k-s\to k} \circ \mathrm{MP}_{n\to k-s}$ maps symmetric states to (k, s, d)-almost-product states.

Theorem 8 (Exponential de Finetti theorem[32, 28]). *For any $0 \le r \le k$, there exist $x_0, \ldots, x_r \in \mathbb{R}$ such that $|x_s| \le (2\delta)^s/(1-\delta)$, $\epsilon := \delta^r/(1-3\delta)$, $\delta := k(d+k)/n$ and*

$$
\left\| \mathrm{tr}_{n-k} - \sum_{s=0}^{r} x_s \, \mathrm{Clone}_{k-s\to k} \circ \mathrm{MP}_{n\to k-s} \right\|_\diamond \le \epsilon,, \tag{34}
$$

where the maps in (34) are restricted to act on $\vee^n \mathbb{C}^d$.

Perhaps a more natural formulation comes from taking $r = k$, so that

$$
\mathrm{tr}_{n-k} = \sum_{s=0}^{k} x_s \, \mathrm{Clone}_{k-s\to k} \circ \mathrm{MP}_{n\to k-s}, \tag{35}
$$

with again the bound $|x_s| \le (2\delta)^s/(1-\delta)$ for each s.

By contrast, the error in [32] is $\le 3(n-k)^d \exp(-\frac{(r+1)(n-k)}{n})$, and [28] has a similar bound. Our result is thus weaker when $n - k$ is small (say $\sim n^{2/3}$), but stronger when r is small and d is large. The likely culprit for this disadvantage is the fact that we upper-bound an alternating sum by taking the absolute value of each term.

Proof. The idea is to write tr_{n-k} as a linear combination of $\mathrm{Clone}_{k-s\to k} \circ \mathrm{MP}_{n\to k-s}$ by inverting the formula (25). For brevity, fix n, k, d, let A_s denote $\mathrm{Clone}_{k-s\to k} \circ \mathrm{tr}_{n-(k-s)}$ and let B_s denote $\mathrm{Clone}_{k-s\to k} \circ \mathrm{MP}_{n\to k-s}$. In this notation, we have $B_0 = \sum_{s=0}^{k} M_{k,k-s}^{(d,n)} A_s$. Observe also that $\mathrm{Clone}_{b\to c} \circ \mathrm{Clone}_{a\to b} = \mathrm{Clone}_{a\to c}$.

---

[2]I am grateful to Matthias Christandl and Ben Toner for sharing with me their unpublished manuscript which proves this point. The idea of their proof is to compare variances. If we choose a random sample of $(1 - o(1))n$ positions from $0^{n/2}1^{n/2}$, then the resulting distribution of Hamming weights will have $o(n)$ variance. However, any almost-product distribution that is approximately balanced must have $\Omega(n)$ variance. With some technical effort, they then translate this into a lower bound on the trace distance between the resulting distributions.

We now rearrange (25) to obtain

$$A_0 = \frac{B_0}{M_{k,k}^{(d,n)}} - \sum_{s=1}^{k} \frac{M_{k,k-s}^{(d,n)}}{M_{k,k}^{(d,n)}} A_s \tag{36}$$

From (31) we have that

$$\frac{1}{M_{k,k}^{(d,n)}} \leq \left(1 - \frac{k(d+k)}{n+d}\right)^{-1} \leq (1-\delta)^{-1} \tag{37}$$

Similarly,

$$\frac{M_{k,k-s}^{(d,n)}}{M_{k,k}^{(d,n)}} = \frac{\binom{n}{k-s}\binom{d+k-1}{s}}{\binom{n}{k}} = \binom{k}{s}\frac{(d+k-s)\cdots(d+k-1)}{(n-k)\cdots(n-k+s-1)} \leq \left(\frac{k(d+k)}{n}\right)^s = \delta^s \tag{38}$$

We now claim that for each r, there exists $x_0, \ldots, x_r, y_{r+1}^{(r)}, \ldots, y_k^{(r)} \in \mathbb{R}$ such that

$$A_0 = \sum_{s=0}^{r-1} x_s B_s + \sum_{s=r}^{k} y_s^{(r)} A_s \tag{39a}$$

and the coefficients satisfy

$$|y_s^{(r)}| \leq 2^r \delta^s \qquad \text{and} \qquad |x_s| \leq \frac{|y_s^{(s)}|}{1-\delta} \leq \frac{(2\delta)^s}{1-\delta} \tag{39b}$$

We prove (39) by induction. The $r=0$ case is trivial; we simply have $y_0^{(0)} = 1$. Next, for $r \geq 0$, we assume that (39) holds for r and attempt to prove it for $r+1$. First we replace the $y_r^{(r)} A_r$ term in (39a) with the linear combination of B_r and $A_{r+1}, \ldots, A_k$ given by (36), to obtain

$$x_r = \frac{y_r}{M_{r,r}^{(d,n)}} \qquad \text{and} \qquad y_s^{(r+1)} = y_s^{(r)} - \frac{M_{k-r,k-s}^{(d,n)}}{M_{k-r,k-r}^{(d,n)}} y_r^{(r)} \tag{40}$$

Using (37) we obtain the claimed bound on $|x_r|$ in (39b). To obtain the claimed bound on $|y_s^{(r+1)}|$, we use induction and (38) to argue that $|y_s^{(r+1)}| \leq 2^r \delta^s + \delta^{s-r} \cdot 2^r \delta^r = 2^{r+1}\delta^s$. $\qquad \square$

2.3.1 Applications of the de Finetti theorem

The de Finetti theorem has an amazing array of applications, but these are not entirely obvious upon first inspection. We wil avoid delving into them deeply here, but single out only two.

1. *Extensive quantities.* Often we are interested in *extensive* properties of a state, such as energy or entropy, that scale linearly with the number of copies of a state. In other words, they satisfy $f(\rho^{\otimes n}) = nf(\rho)$. In this case, the de Finetti approximation provides a way to understand extensive properties of symmetric states by reducing to the case of density matrices on single systems. See [32] for more discussion of this point, [31] for an application to quantum key distribution or [16] for an application to mean-field Hamiltonians.

12

2. *Approximating separable states.* The set of separable density matrices (i.e. the convex hull of $|\alpha\rangle\langle\alpha| \otimes |\beta\rangle\langle\beta|$) is notoriously hard to approximate [22, 6], so we are forced to use heuristics and relaxations. One of the leading relaxations comes from the de Finetti theorem. We say that ρ^{AB} is k-extendible if there exists a state $\sigma^{AB_1\cdots B_k}$ such that $\operatorname{supp}\sigma^{B_1\cdots B_k} \subseteq \vee^k B$ and $\rho^{AB} = \sigma^{AB_1}$. All separable states are clearly k-extendible for all k, and it can also be shown that all non-separable (i.e. entangled) states *fail* to be k-extendible for some, perhaps large, k. Thus, the set of k-extendible states comprises a hierarchy of relaxations of the set of separable states [13]. This can be understood in terms of the fact that by the de Finetti theorem, tracing out $B_2 \ldots B_k$ is similar to applying the entanglement-breaking channel $\mathrm{MP}_{k\to 1}$ to $B_1 \ldots B_k$. (Alternate intuition comes from the idea of "monogamy of entanglement," which states that A cannot simultaneously be highly entangled with all of the B_i; e.g. see [37].) An intriguing open problem is to understand how k-extendability combines with the PPT condition; see [30] for some work along these lines.

Research on the de Finetti theorem continues, and the interested reader is referred to [28] for a far-reaching representation-theoretic generalization or [9] for a powerful variant that uses a different norm than the trace distance (a line of work continued in [8]).

3 Netless concentration of measure

In this section, we show how the symmetric subspace can be a way to prove large-deviation bounds in a manner analogous to controlling higher moments of real random variables. The techniques are (to my knowledge) new, but the results obtained are not substantially different from previous results. One appealing feature of these results, though, is the unified derivation of previously unrelated statements about the minimum entanglement of random subspaces.

3.1 Introduction

A useful trick in high-dimensional geometry is to combine a concentration-of-measure bound with a union bound over an epsilon net. The idea is that we have a metric space S, a random function $f : S \to \mathbb{R}$ (usually with the Lipschitz property $|f(x) - f(y)| \le d(x, y)$) and an ϵ-net $N \subset S$. The concentration-of-measure bound states that for any x, $f(x)$ has extremely low probability, say $\le \eta$, of deviating from its mean value μ by more than δ. This implies that with probability $\ge 1 - |N|\eta$ (and often we only need that this probability is > 0), we have $|f(x) - \mu| \le \delta$ for all $x \in N$, and by the Lipschitz property, $|f(x) - \mu| \le \epsilon + \delta$ everywhere. In some cases, we can do better. For example, if f is a seminorm (and S satisfies some more conditions, maybe having diameter 1) then we can obtain the often stronger bound $|f(x) - \mu| \le \delta/(1 - \epsilon)$.

There is an alternate way to view the first sort of bound. For an event E, define $[E]$ to be the random variable that is 1 if E is true and 0 if E is false. Then define

$$g(x) := [f(x) \ge \mu + \delta].$$

By our concentration-of-measure assumption, for any x, $\mathbb{E}_f[g(x)] \le \eta$. Now fix a normalized measure on S. Then $\mathbb{E}_x \mathbb{E}_f[g(x)] \le \delta$ as well. For any x, let $B(x, \epsilon)$ denote the ball of radius ϵ around x. Assume further that our measure on S has the property that $|B(x, \epsilon)| = V(\epsilon)$, i.e. is independent of x. Now let

$$p = \Pr\{\exists \hat{x} \in S : f(\hat{x}) \ge \mu + \epsilon + \delta\}.$$

In case such an $\hat{x}$ exists, the Lipschitz condition on f guarantees that $f(x) \geq \mu+\delta$ for all $x \in B(\hat{x}, \epsilon)$. Therefore we have $\mathbb{E}_x \mathbb{E}_f[g(x)] \geq pV(\epsilon)$. Combining our two bounds on $\mathbb{E}_x \mathbb{E}_f[g(x)]$, we find that $p \leq \eta/V(\epsilon)$. Since $V(\epsilon)^{-1} \leq |N| \leq V(\epsilon/2)^{-1}$ for a minimal ϵ-net N, this yields bounds that are at least as strong as the ϵ-net-based approach, although not dramatically better.

However, this approach can be further improved by different choices of function g. Indeed, we need only that $\mathbb{E}_x \mathbb{E}_f[g(x)]$ is extremely small, and that conditioned on $f(x)$ being large for some value of x, $\mathbb{E}_f[g(x)]$ is also large. This is the idea behind Chebyshev's inequality and the Bernstein trick, in which $g(x)$ is taken to be either $(f(x) - \mu)^2$ or $e^{yf(x)}$, respectively. The idea of such choices of g is to amplify large deviations of f so that they make have a greater effect on the expectation. These techniques have been useful in quantum information theory for proving concentration bounds, for example in [7], where a moment generating function was used, and in [1], where bounding the second moment was sufficient to produce powerful results.

One reason that $g(x) = e^{yf(x)}$ is an appealing choice is Cramer's theorem [12], which, up to technical caveats, is as follows. When x is of the form $(x_1, \ldots, x_n)$ for i.i.d. $x_1, \ldots, x_n$ and with $f(x) := \frac{1}{n} \sum_{i=1}^n F(x_i)$, then (i) $\Pr\{f(x) \geq a\} \sim \exp(-ns(a) - o(n))$ for some $s(a)$, and (ii) optimizing over y can yield the nearly-optimal bound of $\exp(-ns(a))$. (To relate to the earlier discussion, we have $a = \mu + \delta$.) Indeed, $s(a) = \sup_y(ya - \ln \mathbb{E}_{x_1}[e^{yF(x_1)}])$.

However, it turns out that taking $g(x) = x^p$ and optimizing over p always yields a bound that is at least as powerful than when $g(x)$ is of the form $e^{yf(x)}$ [14]. For this to work, we need that $f(x) \geq 0$ with probability 1, but no longer need the i.i.d. assumption. To see why $g(x) = x^p$ is at least as good a choice, let $\gamma(a) = \min_{p \in \mathbb{N}} \mathbb{E}[f(x)^p]/a^p$ be the optimal bound obtainable by optimizing over p. Then $\mathbb{E}[f(x)^p] \geq \gamma(a) \cdot a^p$ for all nonnegative integers p, and thus $\mathbb{E}[e^{yx}] \geq \sum_{p \geq 0} \gamma(a) y^p a^p/p! = \gamma(a) e^{ya}$ and finally $e^{-s(a)} \geq \gamma(a)$.

In this section, we will focus on showing that random subspaces are likely to contain only highly entangled states. Thus our results will be similar in many ways to those of [24], which used the more conventional methods of ϵ-nets and Levy's Lemma (which is based on Gaussian concentration, which in turn can be derived from moment-generating functions). The advantages of this approach is that the proof is somewhat more self-contained and the resulting bounds are now strong ehough to unify several different previous results. The main disadvantage compared with Levy's Lemma is a loss in flexibility, a limitation whose consequences we will return to below.

3.2 Statement of results

Define S^d to be the set of unit vectors in $\mathbb{C}^d$. All expectations are taken with respect to unitarily invariant measures. In this part of the paper, we will always consider the following scenario. There are k quantum systems of dimensions $d_1, d_2, \ldots d_k$, with $D := d_1 d_2 \cdots d_k$. For any Hermitian operator Π acting on $\mathbb{C}^D$, we will define

$$\nu(\Pi) := \max \left\{ \mathrm{tr}(\varphi_1 \otimes \cdots \otimes \varphi_k)\Pi : |\varphi_1\rangle \in S^{d_1}, \ldots, |\varphi_k\rangle \in S^{d_k} \right\}.$$

We will generally consider the case when Π is a random orthogonal projector of rank r.

Theorem 9. *Let Π be a random rank-r orthogonal projector acting on $\bigotimes_{i=1}^k \mathbb{C}^{d_i}$. Then for any $\gamma > 0$,*

$$\Pr_{\Pi} [\nu(\Pi) \geq \gamma] \leq \inf_n \frac{\binom{r+n-1}{n} \prod_{i=1}^k \binom{d_i+n-1}{n}}{\gamma^n \binom{D+n-1}{n}} \tag{41}$$

Before presenting the proof, we examine three corollaries of Theorem 9 corresponding to different special cases.

3.2.1   Large subspaces

One limit is the case of subspaces with small codimension, where the minimal entanglement is small or zero.

Corollary 10. *Let $D = \prod_{i=1}^{k} d_i$ and let V be a uniformly random projector in $\bigotimes_{i=1}^{k} \mathbb{C}^{d_i}$ of rank r such that $D > r + \sum_{i=1}^{k}(d_i - 1)$. Then the probability that V contains a product state is zero. Equivalently, if we take Π to be the orthogonal projector onto V, then $\nu(\Pi) < 1$ with probability 1.*

This can be proven by standard algebraic-geometric arguments [15, 26]; a more explicit argument for this fact was given recently by Walgate and Scott [35]. These works also proved the optimality of Corollary 10, meaning that if $D \le r + \sum_{i=1}^{k}(d_i - 1)$ then any subspace of dimension r must contain at least one product state.

Proof. Set $\gamma = 1$. Then the RHS of (41) is

$$\Pr_{\Pi}\left[\nu(\Pi) \ge \gamma\right] \le \inf_{n} \frac{\binom{r+n-1}{n} \prod_{i=1}^{k} \binom{d_i+n-1}{n}}{\binom{D+n-1}{n}} \tag{42}$$

Note that when d is fixed and n is large $\binom{d+n-1}{n} = O(n^{d-1})$. Thus as $n \to \infty$ (42) is

$$O(n^{\sum_{i=1}^{k}(d_i-1)+r-1-D+1})$$

which tends to zero if $D > r + \sum_{i=1}^{k}(d_i - 1)$. $\qquad\square$

One difference between our proof and those based on algebraic geometry is that ours degrades smoothly when we take γ to be slightly smaller than one. Indeed, we can prove a nonzero, but weak, lower bound on the minimum entanglement of vector spaces meeting the conditions of Corollary 10. For simplicity, we consider the case of $k = 2$ and $d_1 = d_2 = d$, although the general case poses no additional difficulties.

Proposition 11. *Let Π be the projectors onto a random subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$ of dimension $r = d^2 - 2(d-1) - x$ for some positive integer x. Then*

$$\Pr_{\Pi}[\nu(\Pi) \ge 1 - d^{-2-2d/x}] \le d^{-d} \tag{43}$$

Proof. Set $n = d^{2+2d/x}$ so that $\gamma = 1 - 1/n$. Observe that

$$\frac{n^{a-1}}{a-1!} \le \binom{n+a-1}{n} \le \frac{n^{a-1}}{a-1!} e^{a^2/2n}. \tag{44}$$

Applying this to (41) with $\gamma =$, we obtain

$$\Pr_{\Pi}[\nu(\Pi) \ge \gamma] \le \frac{d^2 - 1!}{d - 1!^2(d^2 - 2(d-1) - x)!} \frac{1}{\gamma^n n^x} e^{d^4/n}$$

$$\le \frac{d^{2(d-1)+2x}}{\gamma^n n^x}$$

Substituting $\gamma = 1 - 1/n$ yields the desired bound. $\qquad\square$

3.2.2 Entanglement of random pure bipartite states

Second, consider the case when $r = 1$, and so $\Pi = |\psi\rangle \langle \psi|$. When $k = 2$, $\nu(\psi)$ is simply the largest Schmidt value of a random state, and should be roughly $(1/\sqrt{d_1} + 1/\sqrt{d_2})^2$, according to the Marčenko-Pastur law [19]. Unfortunately, Theorem 9 is not quite strong enough to prove this, and can only obtain a bound of $1/d_1 + \frac{O(1 + \log(d_2/d_1))}{\sqrt{d_1 d_2}}$ when $d_1 \gg d_2$. To illustrate the technique, we consider the case of $d_1 = d_2 = d$, when the true value of $\nu(\Pi)$ is $\approx 4/d$, and we achieve a result that is weaker by a constant factor.

Corollary 12. *Let $|\psi\rangle$ be drawn uniformly at random from $\mathbb{C}^d \otimes \mathbb{C}^d$, Let $\gamma_0 = \frac{16}{ed}$. Then*

$$\Pr_\psi \left\{ \|\psi^A\|_\infty \geq \gamma_0 e^\epsilon \right\} \leq e^{-d\epsilon} \tag{45}$$

Proof. Let $n = d$. Then $\binom{d+n-1}{n} \leq 4^d$ and $\binom{d^2+n-1}{n} \geq d^{2d}/d! \geq (ed)^d$. We now substitute into (41) and obtain (45). $\qquad \square$

3.2.3 Multi-qubit states

We also recover another corollary in the case of many qubits that slightly sharpens the main technical result of [20].

Corollary 13. *Let $d_1 = \cdots = d_k = 2$ and $r = 1$. Choose some $\epsilon > 0$. Then a random k-qubit pure state has probability*

$$\left(\frac{1}{\epsilon^{(1+1/\epsilon)} k} \right)^k \sim k^{-k} \tag{46}$$

of having overlap $\geq \gamma := k^{1+2\epsilon} 2^{-k}/e$ with any k-qubit product state.

Proof. Plugging $d_1 = \cdots = d_k = 2$ and $r = 1$ into (41) yields an upper bound of

$$\frac{(n+1)^k n!}{\gamma^n 2^k (2^k + 1) \cdots (2^k + n - 1)} \tag{47}$$

We then choose $n = k/\epsilon$ and can bound (47) with

$$\leq \frac{(n+1)^k n!}{(k^{1+2\epsilon}/e)^n} \leq \frac{(k/\epsilon)^k (k/e\epsilon)^{k/\epsilon}}{k^{k(2+1/\epsilon)} e^{k/\epsilon}} = \left(\frac{1}{\epsilon^{(1+1/\epsilon)} k} \right)^k.$$

$\qquad \square$

By contrast, Gross, Flammia and Eisert [20] prove that the probability of $\nu(\Pi) \geq 8k^2 2^{-k}$ is $\leq e^{-k^2}$.

3.3 Proof of the main result

To prove Theorem 9, we will relate the maximum overlap of Π with a product state to the nth moment of its overlap, which we define as

$$\mu_k^n(\Pi) := \mathbb{E}_{\varphi_1,\ldots,\varphi_k} (\text{tr}(\Pi \bigotimes \varphi_i))^n \tag{48}$$

16

Naturally $\nu(\Pi) = \lim_{n\to\infty} \mu_k^n(\Pi)^{1/n}$. But we will see that more precise quantitative estimates are possible.

It will be convenient to let (48) be defined for any positive semidefinite operator Π.

Lemma 14.

- μ_k^n *is* homogenous. *That is, for any $x > 0$ and $\Pi \geq 0$, $\mu_k^n(x\Pi) = x^n \mu_k^n(\Pi)$.*

- μ_k^n *is* non-decreasing. *That is, if $0 \leq A \leq B$, then $0 \leq \mu_k^n(A) \leq \mu_k^n(B)$.*

The proof is omitted. (In fact, the μ_k^n are norms. See [29] for more discussions of their properties and relations between μ_k^n for different values of n.)

The strategy of our proof is to calculate $\mathbb{E}_\Pi[\mu_k^n(\Pi)]$ in two different ways. On the one hand, it can be evaluated exactly, as we will discuss below. On the other hand, for any fixed Π, this expression can be lower-bounded in terms of $\nu(\Pi)$ as follows:

Lemma 15. *For any Π and any $n > 0$,*

$$\mu_k^n(\Pi) \geq \frac{\nu(\Pi)^n}{\prod_{i=1}^k \binom{d_i+n-1}{n}}. \tag{49}$$

Proof of Lemma 15. Let $|\hat{\varphi}\rangle = |\hat{\varphi}_1\rangle \otimes \cdots \otimes |\hat{\varphi}_k\rangle$ be a product state maximizing $\operatorname{tr}\Pi\hat{\varphi}$; i.e. such that $\operatorname{tr}\Pi\hat{\varphi} = \nu(\Pi)$. Define $p := \operatorname{tr}\Pi\hat{\varphi}$ and $|\psi\rangle := p^{-1/2}\Pi\hat{\varphi}$. Note that $\Pi \geq \psi$ and that $\nu(\psi) = \nu(\Pi)$. Thus, it suffices to prove the lemma in the case when $\Pi = \psi$.

Let $|\psi_k\rangle = |\psi\rangle$. We now iteratively define $p_{k-1}, \ldots, p_1$ and $|\psi_{k-1}\rangle, \ldots, |\psi_1\rangle$ as follows. For $j = k-1, \ldots, 1$, choose $p_j > 0$ and $|\psi_j\rangle \in S^{d_1\cdots d_j}$ to satisfy

$$I_{d_1} \otimes \cdots \otimes I_{d_{j-1}} \otimes \langle\hat{\varphi}_j| \cdot |\psi_j\rangle = \sqrt{p_j}\,|\psi_j\rangle. \tag{50}$$

Observe that $p = p_1 \cdots p_{k-1}$.

We will show that

$$\mu_k^n(\psi) \geq \frac{p_{k-1}^n}{\binom{d_k+n-1}{n}} \mu_{k-1}^n(\psi_{k-1}). \tag{51}$$

This can then be applied inductively to establish the lemma.

Now we calculate

$$\mu_k^n(\Pi) = \mathop{\mathbb{E}}_{\varphi_1,\ldots,\varphi_k} \operatorname{tr} \psi^{\otimes n} \bigotimes_{i=1}^{k} \varphi_i^{\otimes n} \tag{52}$$

$$= \mathop{\mathbb{E}}_{\varphi_1,\ldots,\varphi_{k-1}} \sum_{\pi \in \mathcal{S}_n} \operatorname{tr} \psi^{\otimes n} \left(\bigotimes_{i=1}^{k-1} \varphi_i^{\otimes n} \otimes \frac{P_d(\pi)}{d_k^{\bar{n}}} \right) \qquad \text{by Proposition 6} \tag{53}$$

$$= \mathop{\mathbb{E}}_{\varphi_1,\ldots,\varphi_{k-1}} \sum_{\pi \in \mathcal{S}_n} \operatorname{tr} \psi^{\otimes n} P_d(\pi)^{\otimes k} P_d(\pi^{-1})^{\otimes k} \left(\bigotimes_{i=1}^{k-1} \varphi_i^{\otimes n} \otimes \frac{P_d(\pi)}{d_k^{\bar{n}}} \right) \tag{54}$$

$$= \mathop{\mathbb{E}}_{\varphi_1,\ldots,\varphi_{k-1}} \sum_{\pi \in \mathcal{S}_n} \operatorname{tr} \psi^{\otimes n} \left(\bigotimes_{i=1}^{k-1} \varphi_i^{\otimes n} \otimes \frac{I_d^{\otimes n}}{d_k^{\bar{n}}} \right) \qquad \text{since } \langle\psi|^{\otimes n} \text{ and } |\varphi_i\rangle^{\otimes n} \text{ are symmetric} \tag{55}$$

$$= \frac{1}{\binom{d_k+n-1}{n}} \mathop{\mathbb{E}}_{\varphi_1,\ldots,\varphi_{k-1}} \operatorname{tr}(\operatorname{tr}_k \psi)^{\otimes n} \bigotimes_{i=1}^{k-1} \varphi_i^{\otimes n} \tag{56}$$

$$= \frac{1}{\binom{d_k+n-1}{n}} \mu_{k-1}^n(\operatorname{tr}_k \psi) \tag{57}$$

$$\geq \frac{1}{\binom{d_k+n-1}{n}} \mu_{k-1}^n(p_{k-1}\psi_{k-1}) \qquad \text{since } \operatorname{tr}_k \psi \geq p_{k-1}\psi_{k-1} \tag{58}$$

$$= \frac{p_{k-1}^n}{\binom{d_k+n-1}{n}} \mu_{k-1}^n(\psi_{k-1}) \qquad \text{by homogeneity (Lemma 14)} \tag{59}$$

This concludes the proof of the Lemma. $\qquad\qquad\square$

Remark: Lemma 15 has the following alternate interpretation (which we will make use of).

$$\max\{ |\langle\psi|\hat{\varphi}_1,\ldots,\hat{\varphi}_k\rangle|^{2n} : |\hat{\varphi}_1\rangle \in S^{d_1},\ldots,|\hat{\varphi}_k\rangle \in S^{d_k} \} \cdot \mathop{\mathbb{E}}_{|\varphi_1\rangle \in S^{d_1},\ldots,|\varphi_k\rangle \in S^{d_k}} [|\langle\hat{\varphi}_1,\ldots,\hat{\varphi}_k|\varphi_1,\ldots,\varphi_k\rangle|^{2n}]$$
$$\leq \mathop{\mathbb{E}}_{|\varphi_1\rangle \in S^{d_1},\ldots,|\varphi_k\rangle \in S^{d_k}} [|\langle\psi|\hat{\varphi}_1,\ldots,\hat{\varphi}_k\rangle|^{2n}]. \tag{60}$$

$$\max\{ |\langle\psi|\hat{\varphi}_1,\ldots,\hat{\varphi}_k\rangle|^{2n} : |\hat{\varphi}_1\rangle \in S^{d_1},\ldots,|\hat{\varphi}_k\rangle \in S^{d_k} \}$$
$$\leq \frac{\mathbb{E}_{|\varphi_1\rangle \in S^{d_1},\ldots,|\varphi_k\rangle \in S^{d_k}}[|\langle\psi|\hat{\varphi}_1,\ldots,\hat{\varphi}_k\rangle|^{2n}]}{\mathbb{E}_{|\varphi_1\rangle \in S^{d_1},\ldots,|\varphi_k\rangle \in S^{d_k}}[|\langle\varphi_1,\ldots,\varphi_k|\hat{\varphi}_1,\ldots,\hat{\varphi}_k\rangle|^{2n}]} \tag{61}$$

Proof of Theorem 9. Let $S_\gamma := \{\Pi : \nu(\Pi) \geq \gamma\}$ and let $p := \Pr_\Pi\{\Pi \in S_\gamma\}$. Our goal is to upper bound p. We will do this by computing

$$\mathop{\mathbb{E}}_\Pi \mu_k^n(\Pi) \tag{62}$$

in two different ways.

Since $\operatorname{tr}\Pi\varphi$ is always ≥ 0, we can lower bound the expectation over all Π by considering the contribution only from $\Pi \in S_\gamma$. By Lemma 15 this gives us the lower bound

$$\mathbb{E}_\Pi \mu_k^n(\Pi) \geq p \frac{\nu(\Pi)^n}{\prod_{i=1}^k \binom{d_i+n-1}{n}}. \tag{63}$$

On the other hand, we can also calculate (62) exactly. Indeed, $\mathbb{E}_\Pi \mu_k^n(\Pi) = \mathbb{E}_{\Pi,\varphi}(\operatorname{tr}\Pi\varphi)^n$ and it turns out that this expectation is independent of φ. To see this, let $\Pi = U^\dagger \Pi_0 U$ for Π_0 a fixed rank-r projector and U drawn uniformly randomly from $U(D)$.

$$\begin{aligned}
\mathbb{E}_\Pi (\operatorname{tr}\Pi\varphi)^{\otimes n} &= \mathbb{E}_U (\operatorname{tr} U\varphi U^\dagger \Pi_0)^n \\
&= \mathbb{E}_U \operatorname{tr}(U\varphi U^\dagger)^{\otimes n} \Pi_0^{\otimes n} \\
&= \operatorname{tr} \frac{P_{\text{sym}}^{D,n}}{\operatorname{tr} P_{\text{sym}}^{D,n}} \Pi_0^{\otimes n} \\
&= \frac{\operatorname{tr} P_{\text{sym}}^{r,n}}{\operatorname{tr} P_{\text{sym}}^{D,n}} = \frac{\binom{r+n-1}{r-1}}{\binom{D+n-1}{D-1}}
\end{aligned} \tag{64}$$

Since (64) holds for all φ, it also equals the expectation and in turn equals $\mathbb{E}_\Pi \mu_k^n(\Pi)$. Finally, we combine (63) and (64) to obtain the desired bound on p. $\qquad\square$

3.4 Discussion

This approach has its strengths, but is also more limited in scope than techniques based on Levy's Lemma. For example, replacing the maximum overlap with product states with some other measure of entanglement would require more effort. Even showing the concentration of the *smallest* Schmidt value of all pure states in a random substate appears to require some additional ideas, although this is not completely hopeless.

We remark that these techniques have some significant overlap with the classic *method of moments* from random matrix theory (see [34, 3] for reviews, or [2] for a quantum example).

Acknowledgments

References

[1] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: Restructuring quantum information's family tree. *Proc. Roc. Soc. A*, 465(2108):2537–2563, 2009, arXiv:quant-ph/0606225.

[2] A. Ambainis, A. W. Harrow, and M. Hastings. Random tensor theory: extending random matrix theory to random product states. *Commun. Math. Phys.*, 310(1):25–74, 2012, arXiv:0910.0472.

19

[3] G. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[4] K. M. Audenaert. A digest on representation theory of the symmetric group, 2006. `http://personal.rhul.ac.uk/usah/080/QITNotes_files/Irreps_v06.pdf`.

[5] A. Barenco, B. André, D. Deutsch, A. Ekert, R. Jozsa, and C. Macchiavello. Stabilization of quantum computations by symmetrization. *SIAM J. Comput.*, 26:1541–1557, 1997, `arXiv:quant-ph/9604028`.

[6] S. Beigi and P. W. Shor. Approximating the set of separable states using the positive partial transpose test. *J. Math. Phys.*, 51(4):042202, 2010, `arXiv:0902.1806`.

[7] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. J. Winter. Remote preparation of quantum states. *IEEE Trans. Inf. Theory*, 51(1):56–74, 2005, `arXiv:quant-ph/0307100`.

[8] F. G. Brandao and A. W. Harrow. Quantum de Finetti theorems under local measurements with applications. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, STOC '13, pages 861–870, 2013, `arXiv:1210.6367`.

[9] F. G. S. L. Brandão, M. Christandl, and J. Yard. Faithful squashed entanglement. *Commun. Math. Phys.*, 306(3):805–830, 2011, `arXiv:1010.1750`.

[10] G. Chiribella. On quantum estimation, quantum cloning and finite quantum de Finetti theorems. In *Proceedings of the 5th conference on Theory of quantum computation, communication, and cryptography*, TQC'10, pages 9–25, Berlin, Heidelberg, 2011. Springer-Verlag, `arXiv:1010.1875`.

[11] M. Christandl. *The structure of bipartite quantum states: Insights from group theory and cryptography*. PhD thesis, University of Cambridge, 2006, `arXiv:quant-ph/0604183`.

[12] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer, 2009.

[13] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Complete family of separability criteria. *Phys. Rev. A*, 69:022308, Feb 2004, `arXiv:quant-ph/0308032`.

[14] D. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[15] D. Eisenbud. Linear sections of determinantal varieties. *Amer. J. Math*, 110(3):541–575, 1988.

[16] M. Fannes and C. Vandenplas. Finite size mean-field models. *J. Phys. A*, 39(45):13843, 2006, `arXiv:quant-ph/0605216`.

[17] G. B. Folland. How to integrate a polynomial over a sphere. *The American Mathematical Monthly*, 108(5):446–448, May 2001.

[18] R. Goodman and N. Wallach. *Representations and Invariants of the Classical Groups*. Cambridge University Press, 1998.

[19] F. Götze and A. Tikhomirov. Rate of convergence in probability to the Marčenko-Pastur law. *Bernoulli*, 10(3):503–548, 2004, `arXiv:1110.1284`.

[20] D. Gross, S. T. Flammia, and J. Eisert. Most quantum states are too entangled to be useful as computational resources. *Phys. Rev. Lett.*, 102:190501, May 2009, `arXiv:0810.4331`.

[21] F. A. Grünbaum, L. Vinet, and A. Zhedanov. Linear operator pencils on lie algebras and laurent biorthogonal polynomials. *Journal of Physics A: Mathematical and General*, 37(31):7711, 2004.

[22] L. Gurvits. Classical deterministic complexity of Edmonds' problem and quantum entanglement. In *Proc. 35th Annual ACM Symp. Theory of Computing*, pages 10–19, 2003. arXiv:quant-ph/0303055.

[23] A. W. Harrow. *Applications of coherent classical communication and Schur duality to quantum information theory.* PhD thesis, M.I.T., Cambridge, MA, 2005, `arXiv:quant-ph/0512255`.

[24] P. Hayden, D. W. Leung, and A. Winter. Aspects of generic entanglement. *Commun. Math. Phys.*, 265:95, 2006, `arXiv:quant-ph/0407049`.

[25] M. Horodecki, P. W. Shor, and M. B. Ruskai. General entanglement breaking channels. *Reviews in Mathematical Physics*, 15:629–641, 2003, `arXiv:quant-ph/0302031`.

[26] B. Ilic and J. M. Landsberg. On symmetric degeneracy loci, spaces of symmetric matrices of constant rank and dual varieties. *Math. Ann.*, 314:159–174, 1999.

[27] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki. Universal quantum information compression. *Phys. Rev. Lett.*, 81:1714–1717, 1998, `arXiv:quant-ph/9805017`.

[28] R. Koenig and G. Mitchison. A most compendious and facile quantum de Finetti theorem. *J. Math. Phys.*, 50:012105, 2009, `arXiv:quant-ph/0703210`.

[29] A. Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *Journal of Mathematical Physics*, 53(12):122206, 2012, `arXiv:1208.0161`.

[30] M. Navascués, M. Owari, and M. B. Plenio. Power of symmetric extensions for entanglement detection. *Phys. Rev. A*, 80:052306, Nov 2009, `arXiv:0906.2731`.

[31] R. Renner. *Security of quantum key distribution.* PhD thesis, ETHZ, Zurich, 2005, `arXiv:quant-ph/0512258`.

[32] R. Renner. Symmetry of large physical systems implies independence of subsystems. *Nature Physics*, 3:645–649, 2007, `arXiv:quant-ph/0703069`.

[33] R. P. Stanley. *Enumerative Combinatorics, vol. 2.* Cambridge University Press, 1999.

[34] T. Tao. *Topics in Random Matrix Theory.* Graduate Studies in Mathematics. American Mathematical Society, 2012.

[35] J. Walgate and A. J. Scott. Generic local distinguishability and completely entangled subspaces, 2007, `arXiv:0709.4238`.

[36] R. F. Werner. Optimal cloning of pure states. *Phys. Rev. A*, 58(3):1827–1832, Sep 1998, arXiv:quant-ph/9804001.

[37] D. Yang. A simple proof of monogamy of entanglement. *Physics Letters A*, 360(2):249–250, 2006, arXiv:quant-ph/0604168.