

Chapter 1

Concentration of Measure And Quantum Entanglement

Abstract

The concentration of measure phenomenon has been applied to the study of non-commutative probability theory. Basically, the non-trivial observation, citing from Gromov's work [Gro81], states that an arbitrary 1-Lipschitz function $f : S^n \rightarrow \mathbb{R}$ concentrates near a single value $a_0 \in \mathbb{R}$ as strongly as the distance function does. That is,

$$\mu\{x \in S^n : |f(x) - a_0| \geq \epsilon\} < \kappa_n(\epsilon) \leq 2 \exp\left(-\frac{(n-1)\epsilon^2}{2}\right)$$

is applied to computing the probability that, given a bipartite system $A \otimes B$, assume $\dim(B) \geq \dim(A) \geq 3$, as the dimension of the smaller system A increases, with very high probability, a random pure state $\sigma = |\psi\rangle\langle\psi|$ selected from $A \otimes B$ is almost as good as the maximally entangled state.

Mathematically, that is:

Let $\psi \in \mathcal{P}(A \otimes B)$ be a random pure state on $A \otimes B$.

If we define $\beta = \frac{1}{\ln(2)} \frac{d_A}{d_B}$, then we have

$$\Pr[H(\psi_A) < \log_2(d_A) - \alpha - \beta] \leq \exp\left(-\frac{1}{8\pi^2 \ln(2)} \frac{(d_A d_B - 1)\alpha^2}{(\log_2(d_A))^2}\right)$$

where $d_B \geq d_A \geq 3$ [HLW06].

In this report, we will show the process of my exploration of the concentration of measure phenomenon in the context of non-commutative probability theory.

First, we will build the mathematical model describing the behavior of quantum system and why they makes sense for physicists and meaningful for general publics.

1.1 Motivation

First, we introduce a motivation for introducing non-commutative probability theory to the study of quantum mechanics. This section is mainly based on the book [KM].

1.1.1 Light polarization and the violation of Bell's inequality

The light which comes through a polarizer is polarized in a certain direction. If we fix the first filter and rotate the second filter, we will observe the intensity of the light will change.

The light intensity decreases with α (the angle between the two filters). The light should vanish when $\alpha = \pi/2$.

However, for a system of 3 polarizing filters F_1, F_2, F_3 , having directions $\alpha_1, \alpha_2, \alpha_3$, if we put them on the optical bench in pairs, then we will have three random variables P_1, P_2, P_3 .

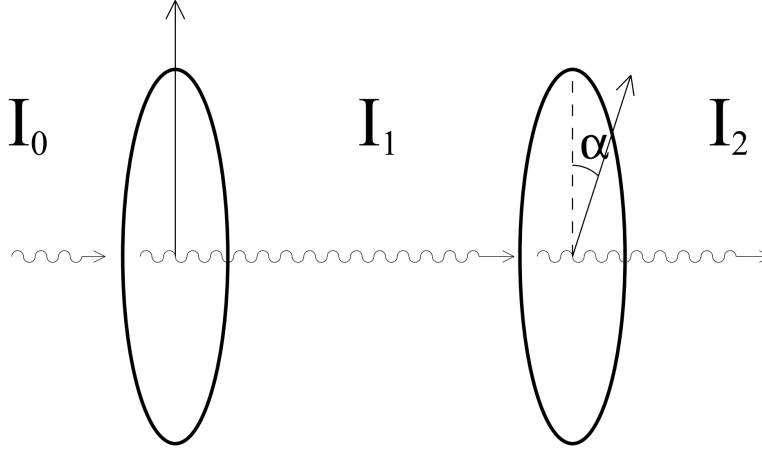


FIG. 1

Figure 1.1: The light polarization experiment, image from [KM]

Theorem 1. *Bell's 3 variable inequality:*

For any three random variables P_1, P_2, P_3 in a classical probability space, we have

$$\text{Prob}(P_1 = 1, P_3 = 0) \leq \text{Prob}(P_1 = 1, P_2 = 0) + \text{Prob}(P_2 = 1, P_3 = 0)$$

Proof. By the law of total probability there are only two possibility if we don't observe any light passing the filter pair F_i, F_j , it means the photon is either blocked by F_i or F_j , it means

$$\begin{aligned} \text{Prob}(P_1 = 1, P_3 = 0) &= \text{Prob}(P_1 = 1, P_2 = 0, P_3 = 0) \\ &\quad + \text{Prob}(P_1 = 1, P_2 = 1, P_3 = 0) \\ &\leq \text{Prob}(P_1 = 1, P_2 = 0) + \text{Prob}(P_2 = 1, P_3 = 0) \end{aligned}$$

□

However, according to our experimental measurement, for any pair of polarizers F_i, F_j , by the

complement rule, we have

$$\begin{aligned}
\text{Prob}(P_i = 1, P_j = 0) &= \text{Prob}(P_i = 1) - \text{Prob}(P_i = 1, P_j = 1) \\
&= \frac{1}{2} - \frac{1}{2} \cos^2(\alpha_i - \alpha_j) \\
&= \frac{1}{2} \sin^2(\alpha_i - \alpha_j)
\end{aligned}$$

This leads to a contradiction if we apply the inequality to the experimental data.

$$\frac{1}{2} \sin^2(\alpha_1 - \alpha_3) \leq \frac{1}{2} \sin^2(\alpha_1 - \alpha_2) + \frac{1}{2} \sin^2(\alpha_2 - \alpha_3)$$

If $\alpha_1 = 0, \alpha_2 = \frac{\pi}{6}, \alpha_3 = \frac{\pi}{3}$, then

$$\begin{aligned}
\frac{1}{2} \sin^2\left(-\frac{\pi}{3}\right) &\leq \frac{1}{2} \sin^2\left(-\frac{\pi}{6}\right) + \frac{1}{2} \sin^2\left(\frac{\pi}{6} - \frac{\pi}{3}\right) \\
\frac{3}{8} &\leq \frac{1}{8} + \frac{1}{8} \\
\frac{3}{8} &\leq \frac{1}{4}
\end{aligned}$$

Other revised experiments (e.g., Aspect's experiment, calcium entangled photon experiment) are also conducted and the inequality is still violated.

1.1.2 The true model of light polarization

The contradiction above marks the point where classical probability stops being adequate. To continue, the sample-space picture must be replaced by states in a Hilbert space and by projections representing measurements. This operator model keeps the experimental probabilities but no longer forces incompatible measurements into a single classical joint distribution.

The full description of the light polarization is given below:

State of polarization of a photon: $\psi = \alpha|0\rangle + \beta|1\rangle$, where $|0\rangle$ and $|1\rangle$ are the two orthogonal polarization states in $\mathbb{C}^2$.

Polarization filter (generalized 0,1 valued random variable): orthogonal projection P_α on $\mathbb{C}^2$ corresponding to the direction α (operator satisfies $P_\alpha^* = P_\alpha = P_\alpha^2$).

The matrix representation of P_α is given by

$$P_\alpha = \begin{pmatrix} \cos^2(\alpha) & \cos(\alpha)\sin(\alpha) \\ \cos(\alpha)\sin(\alpha) & \sin^2(\alpha) \end{pmatrix}$$

Probability of a photon passing through the filter P_α is given by $\langle P_\alpha \psi, \psi \rangle$; this is $\cos^2(\alpha)$ if we set $\psi = |0\rangle$.

Since the probability of a photon passing through the three filters is not commutative, it is impossible to discuss $\text{Prob}(P_1 = 1, P_3 = 0)$ in the classical setting.

We now show how the experimentally observed probability

$$\frac{1}{2} \sin^2(\alpha_i - \alpha_j)$$

arises from the operator model.

Assume the incoming light is *unpolarized*. It is therefore described by the density matrix

$$\rho = \frac{1}{2}I.$$

Let P_{α_i} and P_{α_j} be the orthogonal projections corresponding to the two polarization filters with angles α_i and α_j .

The probability that a photon passes the first filter P_{α_i} is given by the Born rule:

$$\text{Prob}(P_i = 1) = \text{Tr}(\rho P_{\alpha_i}) = \frac{1}{2} \text{Tr}(P_{\alpha_i}) = \frac{1}{2}$$

If the photon passes the first filter, the post-measurement state is given by the Lüders rule:

$$\rho \longmapsto \rho_i = \frac{P_{\alpha_i} \rho P_{\alpha_i}}{\text{Tr}(\rho P_{\alpha_i})} = P_{\alpha_i}.$$

The probability that the photon then passes the second filter is

$$\text{Prob}(P_j = 1 \mid P_i = 1) = \text{Tr}(P_{\alpha_j} P_{\alpha_i}) = \cos^2(\alpha_i - \alpha_j).$$

Hence, the probability that the photon passes P_{α_i} and is then blocked by P_{α_j} is

$$\begin{aligned} \text{Prob}(P_i = 1, P_j = 0) &= \text{Prob}(P_i = 1) - \text{Prob}(P_i = 1, P_j = 1) \\ &= \frac{1}{2} - \frac{1}{2} \cos^2(\alpha_i - \alpha_j) \\ &= \frac{1}{2} \sin^2(\alpha_i - \alpha_j). \end{aligned}$$

This agrees with the experimentally observed transmission probabilities, but it should be emphasized that this quantity corresponds to a *sequential measurement* rather than a joint probability in the classical sense.

1.2 Concentration of measure phenomenon

The operator model explains why entanglement is a meaningful observable, but it does not yet explain why large random systems are typically highly entangled. That is the role of concentration of measure. The next section moves from quantum motivation back to geometry and probability, where high-dimensional spheres already exhibit the same kind of rigidity that later reappears in the entropy of random bipartite states.

Definition 2. *η -Lipschitz function*

Let (X, dist_X) and (Y, dist_Y) be two metric spaces. A function $f : X \rightarrow Y$ is said to be η -Lipschitz if there exists a constant $L \in \mathbb{R}$ such that

$$\text{dist}_Y(f(x), f(y)) \leq L \text{dist}_X(x, y)$$

for all $x, y \in X$. And $\eta = \|f\|_{\text{Lip}} = \inf_{L \in \mathbb{R}} L$.

That basically means that the function f should not change the distance between any two pairs of points in X by more than a factor of L .

This is a stronger condition than continuity, every Lipschitz function is continuous, but not every continuous function is Lipschitz.

Lemma 3. *Isoperimetric inequality on the sphere:*

Let $\sigma_n(A)$ denote the normalized area of A on the n -dimensional sphere S^n . That is, $\sigma_n(A) := \frac{\text{Area}(A)}{\text{Area}(S^n)}$.

Let $\epsilon > 0$. Then for any subset $A \subset S^n$, given the area $\sigma_n(A)$, the spherical caps minimize the volume of the ϵ -neighborhood of A .

Suppose $\sigma^n(\cdot)$ is the normalized volume measure on the sphere $S^n(1)$, then for any closed subset $\Omega \subset S^n(1)$, we take a metric ball B_Ω of $S^n(1)$ with $\sigma^n(B_\Omega) = \sigma^n(\Omega)$. Then we have

$$\sigma^n(U_r(\Omega)) \geq \sigma^n(U_r(B_\Omega))$$

where $U_r(A) = \{x \in X : d(x, A) < r\}$

Intuitively, the lemma means that the spherical caps are the most efficient way to cover the sphere.

Here, the efficiency is measured by the epsilon-neighborhood of the boundary of the spherical cap.

To prove the lemma, we need to have a good understanding of the Riemannian geometry of the sphere. For now, let's just take the lemma for granted.

1.2.1 Levy's concentration theorem

Theorem 4. *Levy's concentration theorem:*

An arbitrary 1-Lipschitz function $f : S^n \rightarrow \mathbb{R}$ concentrates near a single value $a_0 \in \mathbb{R}$ as strongly as the distance function does.

That is,

$$\mu\{x \in S^n : |f(x) - a_0| \geq \epsilon\} < \kappa_n(\epsilon) \leq 2 \exp\left(-\frac{(n-1)\epsilon^2}{2}\right)$$

where

$$\kappa_n(\epsilon) = \frac{\int_{\epsilon}^{\frac{\pi}{2}} \cos^{n-1}(t) dt}{\int_0^{\frac{\pi}{2}} \cos^{n-1}(t) dt}$$

a_0 is a **median** of f , characterized by the following inequalities:

$$\mu(f^{-1}((-\infty, a_0])) \geq \frac{1}{2} \text{ and } \mu(f^{-1}([a_0, \infty))) \geq \frac{1}{2}$$

We will prove the theorem via the Maxwell-Boltzmann distribution law in this section for simplicity. [Shi14] The theorem will be discussed later in more general cases.

Definition 5. *Gaussian measure:*

We denote the Gaussian measure on $\mathbb{R}^k$ as γ^k .

$$d\gamma^k(x) := \frac{1}{\sqrt{2\pi}^k} \exp\left(-\frac{1}{2}\|x\|^2\right) dx$$

$x \in \mathbb{R}^k$, $\|x\|^2 = \sum_{i=1}^k x_i^2$ is the Euclidean norm, and dx is the Lebesgue measure on $\mathbb{R}^k$.

Basically, you can consider the Gaussian measure as the normalized Lebesgue measure on $\mathbb{R}^k$ with standard deviation 1.

It also has another name, the Poincaré limit theorem. [Ver18]

If $X \sim \text{Unif}(S^n(\sqrt{n}))$, then for any fixed unit vector x we have $\langle X, x \rangle \rightarrow N(0, 1)$ in distribution as $n \rightarrow \infty$.

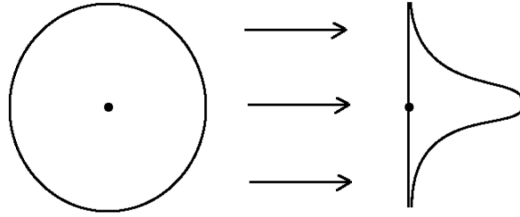


Figure 3.9 The projective central limit theorem: the projection of the uniform distribution on the sphere of radius $\sqrt{n}$ onto a line converges to the normal distribution $N(0, 1)$ as $n \rightarrow \infty$.

Figure 1.2: Maxwell-Boltzmann distribution law, image from [Ver18]

Lemma 6. *Maxwell-Boltzmann distribution law:*

For any natural number k ,

$$\frac{d(\pi_{n,k})_* \sigma^n(x)}{dx} \rightarrow \frac{d\gamma^k(x)}{dx}$$

where $(\pi_{n,k})_*\sigma^n$ is the push-forward measure of σ^n by $\pi_{n,k}$.

In other words,

$$(\pi_{n,k})_*\sigma^n \rightarrow \gamma^k \text{ weakly as } n \rightarrow \infty$$

Proof. We denote the n -dimensional volume measure on $\mathbb{R}^k$ as vol_k .

Observe that $\pi_{n,k}^{-1}(x), x \in \mathbb{R}^k$ is isometric to $S^{n-k}(\sqrt{n - \|x\|^2})$, that is, for any $x \in \mathbb{R}^k$, $\pi_{n,k}^{-1}(x)$ is a sphere with radius $\sqrt{n - \|x\|^2}$ (by the definition of $\pi_{n,k}$).

So,

$$\begin{aligned} \frac{d(\pi_{n,k})_*\sigma^n(x)}{dx} &= \frac{\text{vol}_{n-k}(\pi_{n,k}^{-1}(x))}{\text{vol}_k(S^n(\sqrt{n}))} \\ &= \frac{(n - \|x\|^2)^{\frac{n-k}{2}}}{\int_{\|x\| \leq \sqrt{n}} (n - \|x\|^2)^{\frac{n-k}{2}} dx} \end{aligned}$$

as $n \rightarrow \infty$.

Note that $\lim_{n \rightarrow \infty} (1 - \frac{a}{n})^n = e^{-a}$ for any $a > 0$.

$$(n - \|x\|^2)^{\frac{n-k}{2}} = \left(n(1 - \frac{\|x\|^2}{n}) \right)^{\frac{n-k}{2}} \rightarrow n^{\frac{n-k}{2}} \exp(-\frac{\|x\|^2}{2})$$

So

$$\begin{aligned} \frac{(n - \|x\|^2)^{\frac{n-k}{2}}}{\int_{\|x\| \leq \sqrt{n}} (n - \|x\|^2)^{\frac{n-k}{2}} dx} &= \frac{e^{-\frac{\|x\|^2}{2}}}{\int_{x \in \mathbb{R}^k} e^{-\frac{\|x\|^2}{2}} dx} \\ &= \frac{1}{(2\pi)^{\frac{k}{2}}} e^{-\frac{\|x\|^2}{2}} \\ &= \frac{d\gamma^k(x)}{dx} \end{aligned}$$

□

Now we can prove Levy's concentration theorem, the proof is from [Shi14].

Proof. Let $f_n : S^n(\sqrt{n}) \rightarrow \mathbb{R}$, $n = 1, 2, \dots$, be 1-Lipschitz functions.

Let x and x' be two given real numbers and $\gamma^1(-\infty, x] = \bar{\sigma}_\infty[-\infty, x']$, suppose $\sigma_\infty\{x'\} = 0$, where $\{\sigma_i\}$ is a sequence of Borel probability measures on $\mathbb{R}$.

We want to show that, for all non-negative real numbers ϵ_1 and ϵ_2 .

$$\sigma_\infty[x' - \epsilon_1, x' + \epsilon_2] \geq \gamma^1[x - \epsilon_1, x + \epsilon_2]$$

Consider the two spherical cap $\Omega_+ := \{f_{n_i} \geq x'\}$ and $\Omega_- := \{f_{n_i} \leq x\}$. Note that $\Omega_+ \cup \Omega_- = S^{n_i}(\sqrt{n_i})$.

It is sufficient to show that,

$$U_{\epsilon_1}(\Omega_+) \cap U_{\epsilon_2}(\Omega_-) \subset \{x' - \epsilon_1 \leq f_{n_i} \leq x' + \epsilon_2\}$$

By 1-Lipschitz continuity of f_{n_i} , we have for all $\zeta \in U_{\epsilon_1}(\Omega_+)$, there is a point $\xi \in \Omega_+$ such that $d(\zeta, \xi) \leq \epsilon_1$. So $U_{\epsilon_1}(\Omega_+) \subset \{f_{n_i} \geq x' - \epsilon_1\}$. With the same argument, we have $U_{\epsilon_2}(\Omega_-) \subset \{f_{n_i} \leq x' + \epsilon_2\}$.

So the push-forward measure of $(f_{n_i})_*\sigma^{n_i}$ of $[x' - \epsilon_1, x' + \epsilon_2]$ is

$$\begin{aligned} (f_{n_i})_*\sigma^{n_i}[x' - \epsilon_1, x' + \epsilon_2] &= \sigma^{n_i}(x' - \epsilon_1 \leq f_{n_i} \leq x' + \epsilon_2) \\ &\geq \sigma^{n_i}(U_{\epsilon_1}(\Omega_+) \cap U_{\epsilon_2}(\Omega_-)) \\ &\geq \sigma^{n_i}(U_{\epsilon_1}(\Omega_+)) + \sigma^{n_i}(U_{\epsilon_2}(\Omega_-)) - 1 \end{aligned}$$

By the lemma 3, we have

$$\sigma^{n_i}(U_{\epsilon_1}(\Omega_+)) \geq \sigma^{n_i}(U_{\epsilon_1}(B_{\Omega_+})) \quad \text{and} \quad \sigma^{n_i}(U_{\epsilon_2}(\Omega_-)) \geq \sigma^{n_i}(U_{\epsilon_2}(B_{\Omega_-}))$$

By the lemma 6, we have

$$\sigma^{n_i}(U_{\epsilon_1}(\Omega_+)) + \sigma^{n_i}(U_{\epsilon_2}(\Omega_-)) \rightarrow \gamma^1[x' - \epsilon_1, x' + \epsilon_2] + \gamma^1[x - \epsilon_1, x + \epsilon_2]$$

Therefore,

$$\begin{aligned} \sigma_\infty[x' - \epsilon_1, x' + \epsilon_2] &\geq \liminf_{i \rightarrow \infty} (f_{n_i})_*\sigma^{n_i}[x' - \epsilon_1, x' + \epsilon_2] \\ &\geq \gamma^1[x' - \epsilon_1, \infty) + \gamma^1(-\infty, x + \epsilon_2] - 1 \\ &= \gamma^1[x - \epsilon_1, x + \epsilon_2] \end{aligned}$$

□

The full proof of Levy's concentration theorem requires more digestion for cases where $\bar{\sigma}_\infty \neq \delta_{\pm\infty}$ but I don't have enough time to do so. This section may be filled in the next semester.

1.3 The application of the concentration of measure phenomenon in non-commutative probability theory

Having established concentration for Lipschitz observables on high-dimensional spheres, we can now return to quantum information. The remaining step is to identify a physically meaningful observable on pure states whose geometry is controlled well enough for Levy-type bounds to apply. In this thesis that observable is entanglement entropy, viewed after partial trace.

In quantum communication, we can pass classical bits by sending quantum states. However, by the indistinguishability (Proposition ??) of quantum states, we cannot send an infinite number of classical bits over a single qubit. There exists a bound for zero-error classical communication rate over a quantum channel.

Theorem 7. *Holevo bound:*

The maximal amount of classical information that can be transmitted by a quantum system is given by the Holevo bound. $\log_2(d)$ is the maximum amount of classical information that can be transmitted by a quantum system with d levels (that is, basically, the number of qubits).

The proof of the Holevo bound can be found in [NC10]. In current state of the project, this theorem is not heavily used so we will not make annotated proof here.

1.3.1 Quantum communication

To surpass the Holevo bound, we need to use the entanglement of quantum states.

Definition 8. *Bell state:*

The Bell states are the following four states:

$$\begin{aligned} |\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), & |\Phi^-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\ |\Psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), & |\Psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \end{aligned}$$

These are a basis of the 2-qubit Hilbert space.

1.3.2 Superdense coding and entanglement

Superdense coding is the operational reason entanglement matters in this chapter. It shows that entangled states are not merely algebraically interesting: they change communication capacity. Once that point is clear, the natural probabilistic question is whether highly entangled states are rare or typical, which leads directly to Hayden's concentration theorem.

The description of the superdense coding can be found in [GMS15] and [Hay10].

Suppose A and B share a Bell state (or other maximally entangled state) $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, where A holds the first part and B holds the second part.

A wishes to send 2 **classical bits** to B .

A performs one of four Pauli unitaries (some fancy quantum gates named X, Y, Z, I) on the combined state of entangled qubits $\otimes$ one qubit. Then A sends the resulting one qubit to B .

This operation extends the initial one entangled qubit to a system of one of four orthogonal Bell states.

B performs a measurement on the combined state of the one qubit and the entangled qubits he holds.

B decodes the result and obtains the 2 classical bits sent by A .

Note that superdense coding is a way to send 2 classical bits of information by sending 1 qubit with 1 entangled qubit. **The role of the entangled qubit** is to help them to distinguish the 4 possible states of the total 3 qubits system where 2 of them (the pair of entangled qubits) are mathematically the same.

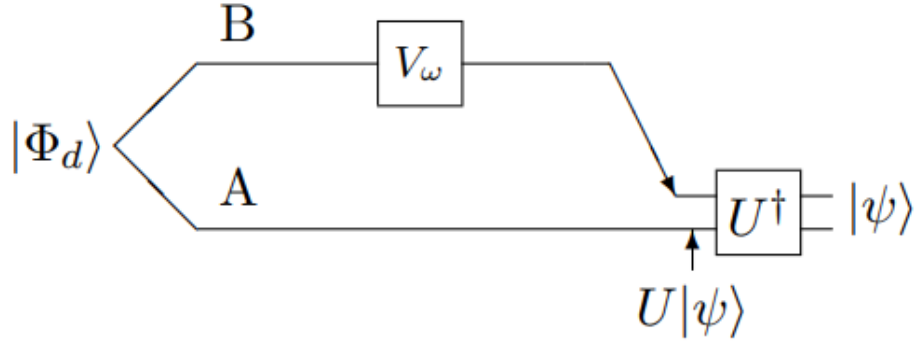


Figure 1.3: Superdense coding, image from [Hay10]

Additionally, no information can be gained by measuring a pair of entangled qubits. To send information from A to B , we need to physically send the qubits from A to B . That means, we cannot send information faster than the speed of light.

1.3.3 Hayden's concentration of measure phenomenon

The geometric and communication-theoretic threads now meet. Random pure states live on a projective state space, partial trace sends them to mixed states, and entropy turns those mixed states into real numbers. Hayden's theorem is precisely the statement that this entropy observable is concentrated when the ambient dimension is large.

The application of the concentration of measure phenomenon in the superdense coding can be realized in random sampling the entangled qubits [Hay10]:

It is a theorem connecting the following mathematical structure:

$$\begin{array}{ccc}
 \mathcal{P}(A \otimes B) & \longleftrightarrow & \mathbb{C}P^{d_A d_B - 1} \\
 \downarrow \text{Tr}_B & \searrow f & \\
 \mathcal{S}(A) & \xrightarrow{H(\psi_A)} & [0, \log_2(d_A)] \subset \mathbb{R}
 \end{array}$$

Figure 1.4: Mathematical structure for Hayden's concentration of measure phenomenon

- The red arrow is the concentration of measure effect. $f = H(\text{Tr}_B(\psi))$.
- $\mathcal{S}(A)$ denotes the mixed states on A .

To prove the concentration of measure phenomenon, we need to analyze the following elements involved in figure 1.4:

The existence and uniqueness of the Haar measure is a theorem in compact lie group theory. For this research topic, we will not prove it.

Due to time constraints of the project, the following lemma is demonstrated but not investigated thoroughly through the research:

Lemma 9. *Page's lemma for expected entropy of mixed states*

Choose a random pure state $\sigma = |\psi\rangle\langle\psi|$ from $A \otimes B$.

The expected value of the entropy of entanglement is known and is given by Page's formula [Pag; San95; BZ17][15.72].

$$\mathbb{E}[H(\psi_A)] = \frac{1}{\ln(2)} \left(\sum_{j=d_B+1}^{d_A d_B} \frac{1}{j} - \frac{d_A - 1}{2d_B} \right) \geq \log_2(d_A) - \frac{1}{2\ln(2)} \frac{d_A}{d_B}$$

It basically provides a lower bound for the expected entropy of entanglement. Experimentally, we can have the following result (see Figure 1.5):

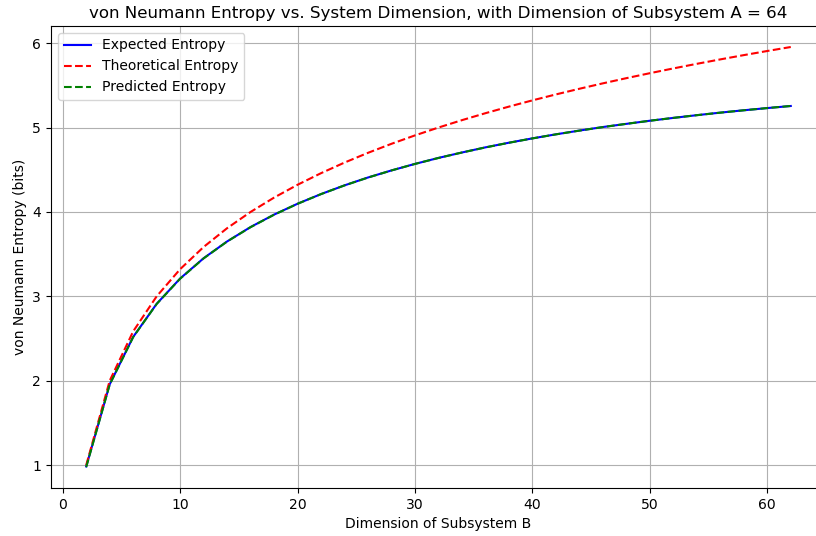


Figure 1.5: Entropy vs dimension

Then we have bound for Lipschitz constant η of the map $S(\varphi_A) : \mathcal{P}(A \otimes B) \rightarrow \mathbb{R}$

Lemma 10. *The Lipschitz constant η of $S(\varphi_A)$ is upper bounded by $\sqrt{8} \log_2(d_A)$ for $d_A \geq 3$.*

Proof. Consider the Lipschitz constant of the function $g : A \otimes B \rightarrow \mathbb{R}$ defined as $g(\varphi) = H(M(\varphi_A))$, where $M : A \otimes B \rightarrow \mathcal{P}(A)$ is any fixed complete von Neumann measurement and $H : \mathcal{P}(A) \otimes \mathcal{P}(B) \rightarrow \mathbb{R}$ is the Shannon entropy.

Let $\{|e_j\rangle_A\}$ be the orthonormal basis for A and $\{|f_k\rangle_B\}$ be the orthonormal basis for B . Then we decompose the state as spectral form $|\varphi\rangle = \sum_{j=1}^{d_A} \sum_{k=1}^{d_B} \varphi_{jk} |e_j\rangle_A |f_k\rangle_B$.

By unitary invariance, suppose $M_j = |e_j\rangle\langle e_j|_A$, and define

$$p_j(\varphi) = \langle e_j | \varphi_A | e_j \rangle = \sum_{k=1}^{d_B} |\varphi_{jk}|^2$$

Then

$$g(\varphi) = H(M(\varphi_A)) = - \sum_{j=1}^{d_A} p_j(\varphi) \log_2(p_j(\varphi))$$

Let $h(p) = -p \log_2(p)$, $h(p) = -\frac{p \ln p}{\ln 2}$, and $h'(p) = -\frac{\ln p + 1}{\ln 2}$. Let $\varphi_{jk} = x_{jk} + iy_{jk}$, then $p_j(\varphi) = \sum_{k=1}^{d_B} (x_{jk}^2 + y_{jk}^2)$, $\frac{\partial p_j}{\partial x_{jk}} = 2x_{jk}$, $\frac{\partial p_j}{\partial y_{jk}} = 2y_{jk}$.

Therefore

$$\frac{\partial g}{\partial x_{jk}} = \frac{\partial g}{\partial p_j} \frac{\partial p_j}{\partial x_{jk}} = -\frac{1 + \ln p_j}{\ln 2} \cdot 2x_{jk} \quad \frac{\partial g}{\partial y_{jk}} = -\frac{1 + \ln p_j}{\ln 2} \cdot 2y_{jk}$$

Then the Lipschitz constant of g is

$$\begin{aligned} \eta^2 &= \sup_{\langle \varphi | \varphi \rangle \leq 1} \nabla g \cdot \nabla g \\ &= \sum_{j=1}^{d_A} \sum_{k=1}^{d_B} \left(\frac{\partial g}{\partial x_{jk}} \right)^2 + \left(\frac{\partial g}{\partial y_{jk}} \right)^2 \\ &= \sum_{j=1}^{d_A} \sum_{k=1}^{d_B} \frac{4(x_{jk}^2 + y_{jk}^2)}{(\ln 2)^2} [1 + \ln p_j(\varphi)]^2 \\ &= \sum_{j=1}^{d_A} \sum_{k=1}^{d_B} \frac{4|\varphi_{jk}|^2}{(\ln 2)^2} [1 + \ln p_j(\varphi)]^2 \end{aligned}$$

Note that $\sum_{k=1}^{d_B} |\varphi_{jk}|^2 = p_j(\varphi)$, $\nabla g \cdot \nabla g = \frac{4}{(\ln 2)^2} \sum_{j=1}^{d_A} p_j(\varphi) (1 + \ln p_j(\varphi))^2$.

Since $0 \leq p_j \leq 1$, we have $\ln p_j(\varphi) \leq 0$, hence $\sum_{j=1}^{d_A} p_j(\varphi) \ln p_j(\varphi) \leq 0$.

$$\begin{aligned} \sum_{j=1}^{d_A} p_j(\varphi) (1 + \ln p_j(\varphi))^2 &= \sum_{j=1}^{d_A} p_j(\varphi) (1 + 2 \ln p_j(\varphi) + (\ln p_j(\varphi))^2) \\ &= 1 + 2 \sum_{j=1}^{d_A} p_j(\varphi) \ln p_j(\varphi) + \sum_{j=1}^{d_A} p_j(\varphi) (\ln p_j(\varphi))^2 \\ &\leq 1 + \sum_{j=1}^{d_A} p_j(\varphi) (\ln p_j(\varphi))^2 \end{aligned}$$

Thus,

$$\begin{aligned}\nabla g \cdot \nabla g &\leq \frac{4}{(\ln 2)^2} \left[1 + \sum_{j=1}^{d_A} p_j(\varphi) (\ln p_j(\varphi))^2 \right] \\ &\leq \frac{4}{(\ln 2)^2} [1 + (\ln d_A)^2] \\ &\leq 8(\log_2 d_A)^2\end{aligned}$$

Proving $\sum_{j=1}^{d_A} p_j(\varphi) (\ln p_j(\varphi))^2 \leq (\ln d_A)^2$ for $d_A \geq 3$ takes some efforts and we will continue that later.

Consider any two unit vectors $|\varphi\rangle$ and $|\psi\rangle$, assume $S(\varphi_A) \leq S(\psi_A)$. If we choose the measurement M to be along the eigenbasis of φ_A , $H(M(\varphi_A)) = S(\varphi_A)$ and we have

$$S(\psi_A) - S(\varphi_A) \leq H(M(\psi_A)) - H(M(\varphi_A)) \leq \eta \| |\psi\rangle - |\varphi\rangle \|$$

Thus the Lipschitz constant of $S(\varphi_A)$ is upper bounded by $\sqrt{8} \log_2(d_A)$. □

From Levy's lemma, we have

If we define $\beta = \frac{1}{\ln(2)} \frac{d_A}{d_B}$, then we have

$$\Pr[H(\psi_A) < \log_2(d_A) - \alpha - \beta] \leq \exp \left(-\frac{1}{8\pi^2 \ln(2)} \frac{(d_A d_B - 1) \alpha^2}{(\log_2(d_A))^2} \right)$$

where $d_B \geq d_A \geq 3$ [HLW06].

Experimentally, we can have the following result:

As the dimension of the Hilbert space increases, the chance of getting an almost maximally entangled state increases (see Figure 1.6).

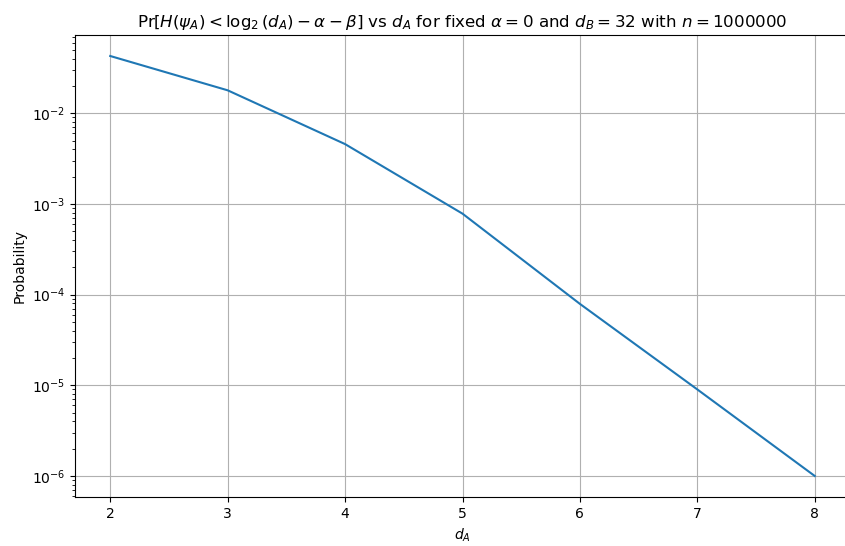


Figure 1.6: Entropy vs d_A

References for Chapter 1

- [BZ17] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2017.
- [Gro81] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Birkhäuser, 1981.
- [GMS15] V. P. Gupta, P. Mandayam, and V. S. Sunder. *The Functional Analysis of Quantum Information Theory*. 2015. arXiv: 1410.7188 [quant-ph]. URL: <https://arxiv.org/abs/1410.7188>.
- [Hay10] P. Hayden. “Concentration of measure effects in quantum information”. In: *Quantum Information Science and Its Contributions to Mathematics*. Vol. 68. Proceedings of Symposia in Applied Mathematics. American Mathematical Society, 2010, pp. 211–260. ISBN: 978-0-8218-4828-9. DOI: 10.1090/psapm/068.
- [HLW06] P. Hayden, D. W. Leung, and A. Winter. “Aspects of Generic Entanglement”. In: *Communications in Mathematical Physics* 265.1 (Mar. 2006), pp. 95–117. ISSN: 1432-0916. DOI: 10.1007/s00220-006-1535-6. URL: <http://dx.doi.org/10.1007/s00220-006-1535-6>.
- [KM] B. Kümmer and H. Maassen. “Elements of quantum probability”. In: *Quantum Probability Communications*, pp. 73–100. DOI: 10.1142/9789812816054_0003. URL: https://www.worldscientific.com/doi/abs/10.1142/9789812816054_0003.
- [NC10] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
- [Pag] D. N. Page. “Page’s conjecture”. In: *Physical Review Letters* ().
- [San95] J. Sanchez-Ruiz. “Page’s conjecture simple proof”. In: *Physical Review E* (1995).
- [Shi14] T. Shioya. *Metric measure geometry*. 2014. arXiv: 1410.0428 [math.MG]. URL: <https://arxiv.org/abs/1410.0428>.
- [Ver18] R. Vershynin. *High-dimensional probability: an introduction with applications in data science*. Cambridge University Press, 2018.