Chapter 1

Concentration of Measure And Quantum Entanglement

Non-commutative probability theory is a branch of generalized probability theory that studies the probability of events in non-commutative algebras (e.g. the algebra of observables in quantum mechanics). In the 20th century, non-commutative probability theory has been applied to the study of quantum mechanics as the classical probability theory is not enough to describe quantum mechanics [KM].

Recently, 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 $f : S^n \to \mathbb{R}$ 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. We assume the reader is an undergraduate student in mathematics and is familiar with the basic concepts of probability theory,

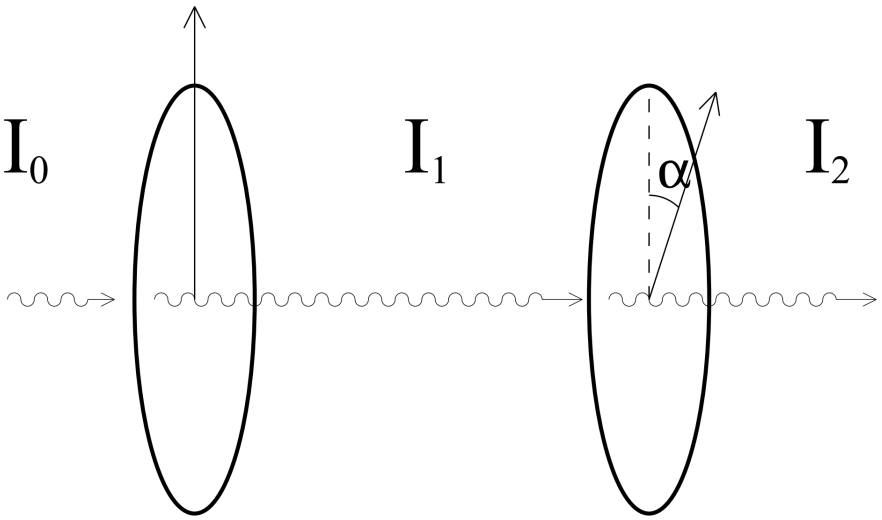

FIG. 1

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

measure theory, linear algebra, and some basic skills of mathematical analysis. To make the report more self-contained, we will add detailed annotated proofs that I understand and references for the original sources.

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.
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 (the event that the photon passes through the first filter but

not the third filter is the union of the event that the photon did not pass through the second filter and the event that the photon passed the second filter and did not pass through the third filter), we have

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

□

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

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

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 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.

This introduces a new model in mathematics explaining quantum mechanics: the non-commutative probability theory.

1.2 Non-commutative probability theory

The non-commutative probability theory is a branch of generalized probability theory that studies the probability of events in non-commutative algebras.

There are several main components of the generalized probability theory; let's see how we can formulate them, comparing with the classical probability theory.

First, we define the Hilbert space in case one did not make the step from the linear algebra courses like me.

Definition 2. *Hilbert space:*

A Hilbert space is a complete inner product space.

That is, a vector space equipped with an inner product that is complete (every Cauchy sequence converges to a limit).

To introduce an example of Hilbert space we use when studying quantum mechanics, we need to introduce a common inner product used in $\mathbb{C}^n$.

Definition 3. *Hermitian inner product:*

On $\mathbb{C}^n$, the Hermitian inner product is defined by

$$\langle u, v \rangle = \sum_{i=1}^{n} \overline{u_i} v_i$$

Proposition 4. *The Hermitian inner product on the complex vector space $\mathbb{C}^n$ makes it a Hilbert space.*

Proof. We first verify that the Hermitian inner product

$$\langle u, v \rangle = \sum_{i=1}^{n} \overline{u_i} v_i$$

on $\mathbb{C}^n$ satisfies the axioms of an inner product:

1. **Conjugate symmetry:** For all $u, v \in \mathbb{C}^n$,

$$\langle u, v \rangle = \sum_{i=1}^{n} \overline{u_i} v_i = \overline{\sum_{i=1}^{n} \overline{v_i} u_i} = \overline{\langle v, u \rangle}.$$

2. **Linearity:** For any $u, v, w \in \mathbb{C}^n$ and scalars $a, b \in \mathbb{C}$, we have

$$\langle u, av + bw\rangle = \sum_{i=1}^{n} \overline{u_i}(av_i + bw_i) = a\langle u, v\rangle + b\langle u, w\rangle.$$

3. **Positive definiteness:** For every $u = (u_1, u_2, \cdots, u_n) \in \mathbb{C}^n$, let $u_j = a_j + b_j i$, where $a_j, b_j \in \mathbb{R}$.

$$\langle u, u\rangle = \sum_{j=1}^{n} \overline{u_j} u_j = \sum_{i=1}^{n} (a_i^2 + b_i^2) \geq 0,$$

with equality if and only if $u = 0$.

Therefore, the Hermitian inner product is an inner product.

Next, we show that $\mathbb{C}^n$ is complete with respect to the norm induced by this inner product:

$$\|u\| = \sqrt{\langle u, u\rangle}.$$

Since $\mathbb{C}^n$ is finite-dimensional, every Cauchy sequence (with respect to any norm) converges in $\mathbb{C}^n$. This is a standard result in finite-dimensional normed spaces, which implies that $\mathbb{C}^n$ is indeed complete.

Therefore, since the Hermitian inner product fulfills the inner product axioms and $\mathbb{C}^n$ is complete, the complex vector space $\mathbb{C}^n$ with the Hermitian inner product is a Hilbert space. □

Another classical example of Hilbert space is $L^2(\Omega, \mathcal{F}, P)$, where $(\Omega, \mathcal{F}, P)$ is a measure space (Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω, and P is a measure on $\mathcal{F}$). The L^2 space is the space of all square integrable, complex-valued measurable functions on Ω.

The square integrable functions are the functions $f : \Omega \to \mathbb{C}$ such that

$$\int_\Omega |f(\omega)|^2 dP(\omega) < \infty$$

with inner product defined by

$$\langle f, g\rangle = \int_\Omega \overline{f(\omega)} g(\omega) dP(\omega)$$

Proposition 5. *$L^2(\Omega, \mathcal{F}, P)$ is a Hilbert space.*

Proof. We check the two conditions of the Hilbert space:

- Completeness: Let (f_n) be a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$. Then for any $\epsilon > 0$, there exists an N such that for all $m, n \geq N$, we have

$$\int_\Omega |f_m(\omega) - f_n(\omega)|^2 dP(\omega) < \epsilon^2$$

This means that (f_n) is a Cauchy sequence in the norm of $L^2(\Omega, \mathcal{F}, P)$.

- Inner product: The inner product is defined by

$$\langle f, g\rangle = \int_\Omega \overline{f(\omega)} g(\omega) dP(\omega)$$

This is a well-defined inner product on $L^2(\Omega, \mathcal{F}, P)$. We can check the properties of the inner product:

  - Linearity:
$$\langle af + bg, h\rangle = a\langle f, h\rangle + b\langle g, h\rangle$$
  - Conjugate symmetry:
$$\langle f, g\rangle = \overline{\langle g, f\rangle}$$
  - Positive definiteness:
$$\langle f, f\rangle \geq 0$$

□

Let $\mathcal{H}$ be a Hilbert space. $\mathcal{H}$ consists of complex-valued functions on a finite set $\Omega = \{1, 2, \cdots, n\}$, and the functions $(e_1, e_2, \cdots, e_n)$ form an orthonormal basis of $\mathcal{H}$. (We use Dirac notation $|k\rangle$ to denote the basis vector e_k [Par92].)

The detailed definition of the non-commutative probability space is given below:

As an analog to the classical probability space $(\Omega, \mathcal{F}, \mu)$, which consists of a sample space Ω and a probability measure μ on the state space $\mathcal{F}$, the non-commutative probability space $(\mathcal{H}, \mathcal{P}, \rho)$ consists of a Hilbert space $\mathcal{H}$ and a state ρ on the space of all orthogonal projections $\mathcal{P}$.
Definition 6. *Non-commutative probability space:*

A non-commutative probability space is a pair $(\mathcal{B}(\mathcal{H}), \mathcal{P})$, where $\mathcal{B}(\mathcal{H})$ is the set of all bounded linear operators on $\mathcal{H}$.

A linear operator on $\mathcal{H}$ is bounded if for all u such that $\|u\| \leq 1$, we have $\|Au\| \leq M$ for some $M > 0$.

$\mathcal{P}$ is the set of all orthogonal projections on $\mathcal{B}(\mathcal{H})$.

The set $\mathcal{P} = \{P \in \mathcal{B}(\mathcal{H}) : P^* = P = P^2\}$ is the set of all orthogonal projections on $\mathcal{B}(\mathcal{H})$.

As a counterpart for the initial probability distribution in the classical probability theory, we need to define the state in the non-commutative probability theory.
Definition 7. *Non-commutative probability state:*

A state on $(\mathcal{B}(\mathcal{H}), \mathcal{P})$ is a map $\rho : \mathcal{P} \to [0, 1]$ such that:

- *$\rho(O) = 0$, where O is the zero projection, and $\rho(I) = 1$, where I is the identity projection.*
- *If $P_1, P_2, \ldots, P_n$ are pairwise disjoint orthogonal projections, then $\rho(P_1 + P_2 + \cdots + P_n) = \sum_{i=1}^n \rho(P_i)$.*

An example of a density operator can be given as follows:

If $(|\psi_1\rangle, |\psi_2\rangle, \cdots, |\psi_n\rangle)$ is an orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of ρ, for the eigenvalues $p_1, p_2, \cdots, p_n$, then $p_j \geq 0$ and $\sum_{j=1}^{n} p_j = 1$.

We can write ρ as

$$\rho = \sum_{j=1}^{n} p_j |\psi_j\rangle\langle\psi_j|$$

(Under basis $|\psi_j\rangle$, it is a diagonal matrix with p_j on the diagonal.)

Then we need to introduce a theorem that ensures that every state on the space of all orthogonal projections on $\mathcal{H}$ can be represented by a density operator.

Theorem 8. *Gleason's theorem (Theorem 1.1.15 in [Par05])*

Let $\mathcal{H}$ be a Hilbert space over $\mathbb{C}$ or $\mathbb{R}$ of dimension $n \geq 3$. Let μ be a state on the space $\mathcal{P}$ of projections on $\mathcal{H}$. Then there exists a unique density operator ρ such that

$$\mu(P) = \mathrm{Tr}(\rho P)$$

for all $P \in \mathcal{P}$. $\mathcal{P}$ is the space of all orthogonal projections on $\mathcal{H}$.

This proof came from [Par05].

This theorem is a very important theorem in non-commutative probability theory; it states that any state on the space of all orthogonal projections on $\mathcal{H}$ can be represented by a density operator.

The counterpart of the random variable in the non-commutative probability theory is called an observable, which is a Hermitian operator on $\mathcal{H}$ (for all ψ, ϕ in the domain of A, we have $\langle A\psi, \phi\rangle = \langle \psi, A\phi\rangle$. This kind of operator ensures that our outcome interpreted as probability is a real number).

Definition 9. *Observable:*

Let $\mathcal{B}(\mathbb{R})$ be the set of all Borel sets on $\mathbb{R}$.

A random variable on the Hilbert space $\mathcal{H}$ is a projection-valued map (measure) $P : \mathcal{B}(\mathbb{R}) \to \mathcal{P}$.

With the following properties:

- *$P(\emptyset) = O$ (the zero projection)*
- *$P(\mathbb{R}) = I$ (the identity projection)*
- *For any sequence $A_1, A_2, \cdots, A_n \in \mathcal{B}(\mathbb{R})$, the following holds:*
  - *$P(\bigcup_{i=1}^{n} A_i) = \bigvee_{i=1}^{n} P(A_i)$*
  - *$P(\bigcap_{i=1}^{n} A_i) = \bigwedge_{i=1}^{n} P(A_i)$*
  - *$P(A^c) = I - P(A)$*
  - *If A_j are mutually disjoint (that is $P(A_i)P(A_j) = P(A_j)P(A_i) = O$ for $i \neq j$), then $P(\bigcup_{j=1}^{n} A_j) = \sum_{j=1}^{n} P(A_j)$*

Definition 10. *Probability of a random variable:*

For a system prepared in state ρ, the probability that the random variable given by the projection-valued measure P is in the Borel set A is $\mathrm{Tr}(\rho P(A))$.

When operators commute, we recover classical probability measures.

Definition 11. *Definition of measurement:*

A measurement (observation) of a system prepared in a given state produces an outcome x, x is a physical event that is a subset of the set of all possible outcomes. For each x, we associate a measurement operator M_x on $\mathcal{H}$.

Given the initial state (pure state, unit vector) u, the probability of measurement outcome x is given by:

$$p(x) = \|M_x u\|^2$$

Note that to make sense of this definition, the collection of measurement operators $\{M_x\}$ must satisfy the completeness requirement:

$$1 = \sum_{x \in X} p(x) = \sum_{x \in X} \|M_x u\|^2 = \sum_{x \in X} \langle M_x u, M_x u \rangle = \langle u, (\sum_{x \in X} M_x^* M_x) u \rangle$$

So $\sum_{x \in X} M_x^* M_x = I$.

Proposition 12. *Proposition of indistinguishability:*

Suppose that we have two systems $u_1, u_2 \in \mathcal{H}_1$, the two states are distinguishable if and only if they are orthogonal.

Proof. Ways to distinguish the two states:

1. Set $X = \{0, 1, 2\}$ and $M_i = |u_i\rangle\langle u_i|$, $M_0 = I - M_1 - M_2$
2. Then $\{M_0, M_1, M_2\}$ is a complete collection of measurement operators on $\mathcal{H}$.
3. Suppose the prepared state is u_1, then $p(1) = \|M_1 u_1\|^2 = \|u_1\|^2 = 1$, $p(2) = \|M_2 u_1\|^2 = 0$, $p(0) = \|M_0 u_1\|^2 = 0$.

If they are not orthogonal, then there is no choice of measurement operators to perfectly distinguish the two states.

□

Intuitively, if the two states are not orthogonal, then for any measurement (projection) there exists non-zero probability of getting the same outcome for both states.

Here is Table 1.1 summarizing the analog of classical probability theory and non-commutative (*quantum*) probability theory [Fer]:

1.3 Concentration of measure phenomenon

Definition 13. *η-Lipschitz function*

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

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

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

Table 1.1: Analog of classical probability theory and non-commutative (*quantum*) probability theory

Classical probability	**Non-commutative probability**
Sample space Ω, cardinality $\|\Omega\| = n$, example: $\Omega = \{0, 1\}$	Complex Hilbert space $\mathcal{H}$, dimension $\dim \mathcal{H} = n$, example: $\mathcal{H} = \mathbb{C}^2$
Common algebra of $\mathbb{C}$ valued functions	Algebra of bounded operators $\mathcal{B}(\mathcal{H})$
$f \mapsto \bar{f}$ complex conjugation	$P \mapsto P^*$ adjoint
Events: indicator functions of sets	Projections: space of orthogonal projections $\mathcal{P} \subseteq \mathcal{B}(\mathcal{H})$
functions f such that $f^2 = f = \overline{f}$	orthogonal projections P such that $P^* = P = P^2$
$\mathbb{R}$-valued functions $f = \overline{f}$	self-adjoint operators $A = A^*$
$\mathbb{I}_{f^{-1}(\{\lambda\})}$ is the indicator function of the set $f^{-1}(\{\lambda\})$	$P(\lambda)$ is the orthogonal projection to eigenspace
$f = \sum_{\lambda \in \text{Range}(f)} \lambda \mathbb{I}_{f^{-1}(\{\lambda\})}$	$A = \sum_{\lambda \in \text{sp}(A)} \lambda P(\lambda)$
Probability measure μ on Ω	Density operator ρ on $\mathcal{H}$
Delta measure δ_ω	Pure state $\rho = \|\psi\rangle\langle\psi\|$
μ is non-negative measure and $\sum_{i=1}^n \mu(\{i\}) = 1$	ρ is positive semi-definite and $\text{Tr}(\rho) = 1$
Expected value of random variable f is $\mathbb{E}_\mu(f) = \sum_{i=1}^n f(i)\mu(\{i\})$	Expected value of operator A is $\mathbb{E}_\rho(A) = \text{Tr}(\rho A)$
Variance of random variable f is $\text{Var}_\mu(f) = \sum_{i=1}^n (f(i) - \mathbb{E}_\mu(f))^2 \mu(\{i\})$	Variance of operator A is $\text{Var}_\rho(A) = \text{Tr}(\rho A^2) - \text{Tr}(\rho A)^2$
Covariance of random variables f and g is $\text{Cov}_\mu(f, g) = \sum_{i=1}^n (f(i) - \mathbb{E}_\mu(f))(g(i) - \mathbb{E}_\mu(g))\mu(\{i\})$	Covariance of operators A and B is $\text{Cov}_\rho(A, B) = \text{Tr}(\rho A \circ B) - \text{Tr}(\rho A)\,\text{Tr}(\rho B)$
Composite system is given by Cartesian product of the sample spaces $\Omega_1 \times \Omega_2$	Composite system is given by tensor product of the Hilbert spaces $\mathcal{H}_1 \otimes \mathcal{H}_2$
Product measure $\mu_1 \times \mu_2$ on $\Omega_1 \times \Omega_2$	Tensor product of space $\rho_1 \otimes \rho_2$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$
Marginal distribution $\pi_* v$	Partial trace $\text{Tr}_2(\rho)$

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.

Lemma 14. *Isoperimetric inequality on the sphere:*

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

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.

1.3.1 Levy's concentration theorem

Theorem 15. *Levy's concentration theorem:*

An arbitrary 1-Lipschitz function $f : S^n \to \mathbb{R}$ 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 the **Levy mean** of function f, that is, the level set $f^{-1} : \mathbb{R} \to S^n$ divides the sphere into equal halves, characterized by the following equality:*

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

We will prove the theorem via the Maxwell-Boltzmann distribution law. [Shi14]
Definition 16. *Gaussian measure:*

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

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

$x \in \mathbb{R}^k$, $\|x\|^2 = \sum_{i=1}^k x_i^2$ 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 Projective limit theorem. [Ver18]

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

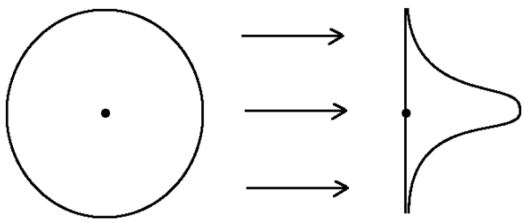

Figure 3.9 The projective central limit theorem: the projection of the uniform distribution on the sphere of radius $\sqrt{n}$ onto a line converges to the normal distribution $N(0,1)$ as $n \to \infty$.

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

Lemma 17. *Maxwell-Boltzmann distribution law:*

For any natural number k,

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

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

In other words,

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

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,

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

as $n \to \infty$.

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

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

So

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

□

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

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

Let x and x' be two given real numbers and $\gamma^1(-\infty, x] = \overline{\sigma}_\infty[-\infty, x']$, 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.

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

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

It is sufficient to show that,

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

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

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

By the lemma 14, 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 17, we have

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

Therefore,

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

□

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

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

In quantum communication, we can pass classical bits by sending quantum states. However, by the indistinguishability (Proposition 12) 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 18. *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.

1.4.1 Quantum communication

To surpass the Holevo bound, we need to use the entanglement of quantum states.
Definition 19. *Bell state:*

The Bell states are the following four states:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \quad |\Phi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$$

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \quad |\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

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

1.4.2 Superdense coding and entanglement

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.

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.

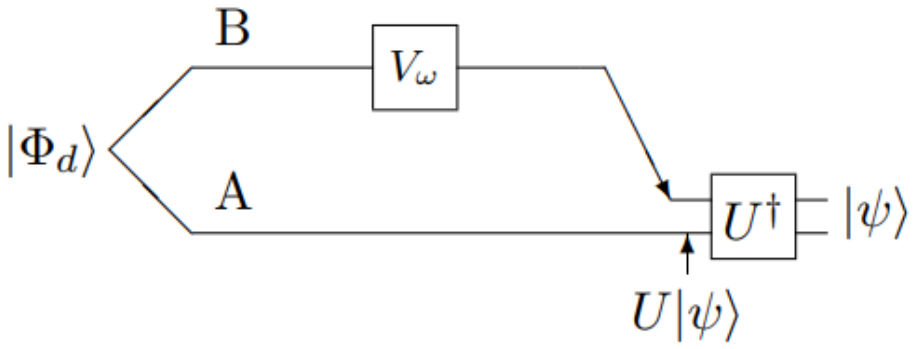

Figure 1.3: Superdense coding, image from [Hay10]

1.4.3 Hayden's concentration of measure phenomenon

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:

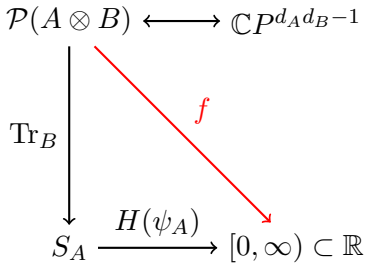

Figure 1.4: Mathematical structure for Hayden's concentration of measure phenomenon

- The red arrow is the concentration of measure effect. $f = H(\mathrm{Tr}_B(\psi))$.
- 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:

First, we need to define what is a random state in a bipartite system. In fact, for pure states, there is a unique uniform distribution under Haar measure that is unitarily invariant.

$U(n)$ is the group of all $n \times n$ **unitary matrices** over $\mathbb{C}$,

$$U(n) = \{A \in \mathbb{C}^{n \times n} : A^* A = A A^* = I_n\}$$

The uniqueness of such measurement came from the lemma below [Mec]

Lemma 20. *Let $(U(n), \|\cdot\|, \mu)$ be a metric measure space where $\|\cdot\|$ is the Hilbert-Schmidt norm and μ is the measure function.*

The Haar measure on $U(n)$ is the unique probability measure that is invariant under the action of $U(n)$ on itself.

That is, fixing $B \in U(n)$, $\forall A \in U(n)$, $\mu(A \cdot B) = \mu(B \cdot A) = \mu(B)$.

The Haar measure is the unique probability measure that is invariant under the action of $U(n)$ on itself.

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.

A random pure state φ is any random variable distributed according to the unitarily invariant probability measure on the pure states $\mathcal{P}(A)$ of the system A, denoted by $\varphi \in_R \mathcal{P}(A)$.

It is trivial that for the space of pure state, we can easily apply the Haar measure as the unitarily invariant probability measure since the space of pure state is S^n for some n. However, for the case of mixed states, that is a bit complicated and we need to use partial tracing to defined the rank-s random states.

Definition 21. *Rank-s random state.*

For a system A and an integer $s \geq 1$, consider the distribution onn the mixed states $\mathcal{S}(A)$ of A induced by the partial trace over the second factor form the uniform distribution on pure states of $A \otimes \mathbb{C}^s$. Any random variable ρ distributed as such will be called a rank-s random states; denoted as $\rho \in_R \mathcal{S}_s(A)$. And $\mathcal{P}(A) = \mathcal{S}_1(A)$.

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

Lemma 22. *Page's lemma for expected entropy of mixed states*

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

The expected value of the entropy of entanglement is known and satisfies a concentration inequality known as Page's formula [Pag; San95; BŻ17][15.72]. The detailed proof is not fully explored in this project and is intended to be done in the next semester.

$$\mathbb{E}[H(\psi_A)] \geq \log_2(d_A) - \frac{1}{2\ln(2)}\frac{d_A}{d_B}$$

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

Then we have bound for Lipschitz constant η of the map $H(\varphi_A)$

Lemma 23. *The Lipschitz constant η of $S(\varphi_A)$ is upper bounded by $\sqrt{8}\log_2(d_A)$ for $d_A \geq 3$.*

From Levy's lemma, we have

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

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

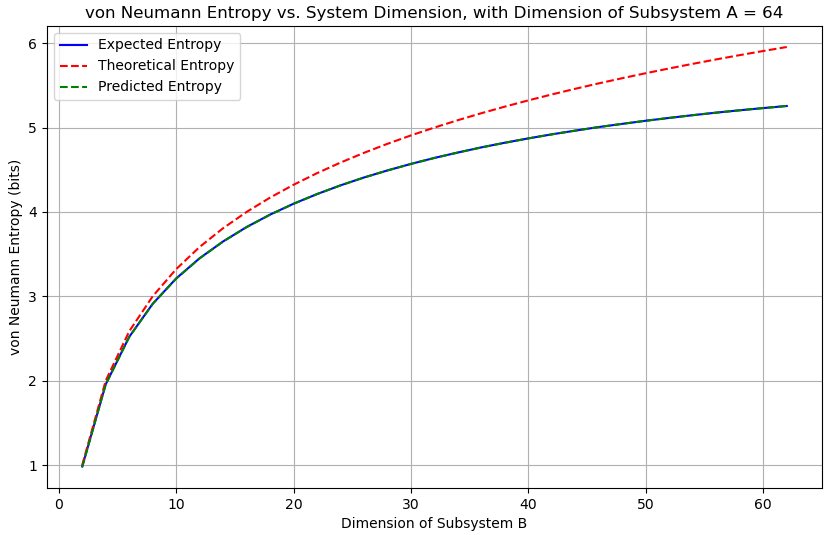

Figure 1.5: Entropy vs dimension

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).

In Hayden's work, the result is also extended to the multiparty case [Hay10], and the result is still under research and I will show the result in the final report if I have enough time.

1.4.4 Majorana stellar representation of the quantum state

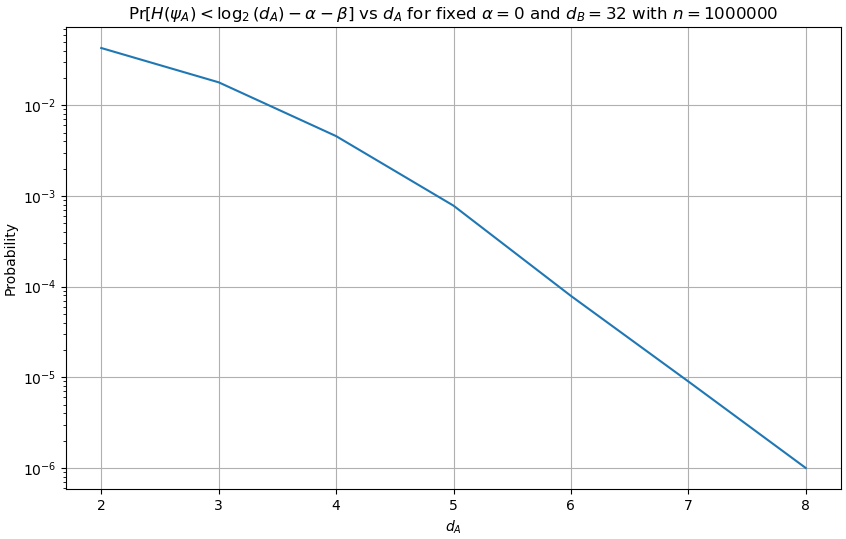

Figure 1.6: Entropy vs d_A

References for Chapter 1

[BŻ17] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2017.

[Fer] R. Feres. *Math 444 Lecture notes – the mathematics of quantum theory*. URL: https://www.math.wustl.edu/~feres/Math444Spring25/Math444Spring25Syllabus.html.

[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.

[Mec] E. Meckes. *The Random Matrix Theory of the Classical Compact Groups*.

[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* ().

[Par92] K. R. Parthasarathy. *An Introduction to Quantum Stochastic Calculus*. Vol. 85. Monographs in Mathematics. Birkhäuser Basel, 1992, pp. XI, 292. ISBN: 978-3-0348-9711-2. DOI: 10.1007/978-3-0348-8641-3.

[Par05] K. R. Parthasarathy. *Mathematical Foundation of Quantum Mechanics*. Vol. 85. Texts and Readings in Mathematics. Hindustan Book Agency, 2005, pp. XI, 292. ISBN: 978-93-86279-28-6. DOI: 10.1007/978-93-86279-28-6.

[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.