Concentration of Measure And Quantum Entanglement

Zheyuan Wu

March 29, 2026

Contents

Preface

This thesis studies how concentration of measure enters quantum information through the geometry of pure states. The basic question is probabilistic but the answer is geometric: once the state space is viewed as a high-dimensional metric-measure space, many physically relevant observables become sharply concentrated. In particular, the entropy of a random bipartite pure state is typically close to its maximal value, which explains why high-dimensional random states are generically highly entangled.

The exposition is organized in layers.

Chapter 0 collects the algebraic, probabilistic, geometric, and quantum-mechanical background needed later.

Chapter 1 develops the motivation from polarization experiments, recalls concentration on spheres, and then states the Hayden–Leung–Winter entanglement bound.

Chapter 2 reformulates the same phenomenon in the language of observable diameter and uses entropy-based simulations as a concrete probe on spheres and complex projective spaces.

Chapter 3 records an exploratory direction through Majorana stellar representation and related holomorphic models. It branches out from our main narrative but is included here for completeness and future reference as our journey into the geometry of quantum states continues.

How to use the dependency graph

The graph below is meant to be read from top to bottom. The first row lists undergraduate courses that supply the basic language. The middle rows list the sections and subsection-level themes that are used in the body of the thesis. The bottom node records the concentration-of-measure theorem that motivates the main narrative. Not every reader needs every path: a reader interested mainly in the entanglement theorem can follow the linear algebra, probability, quantum, and concentration branches first, while a reader interested in the geometric reformulation can continue through manifolds, Riemannian geometry, Hopf fibration, and observable diameter.
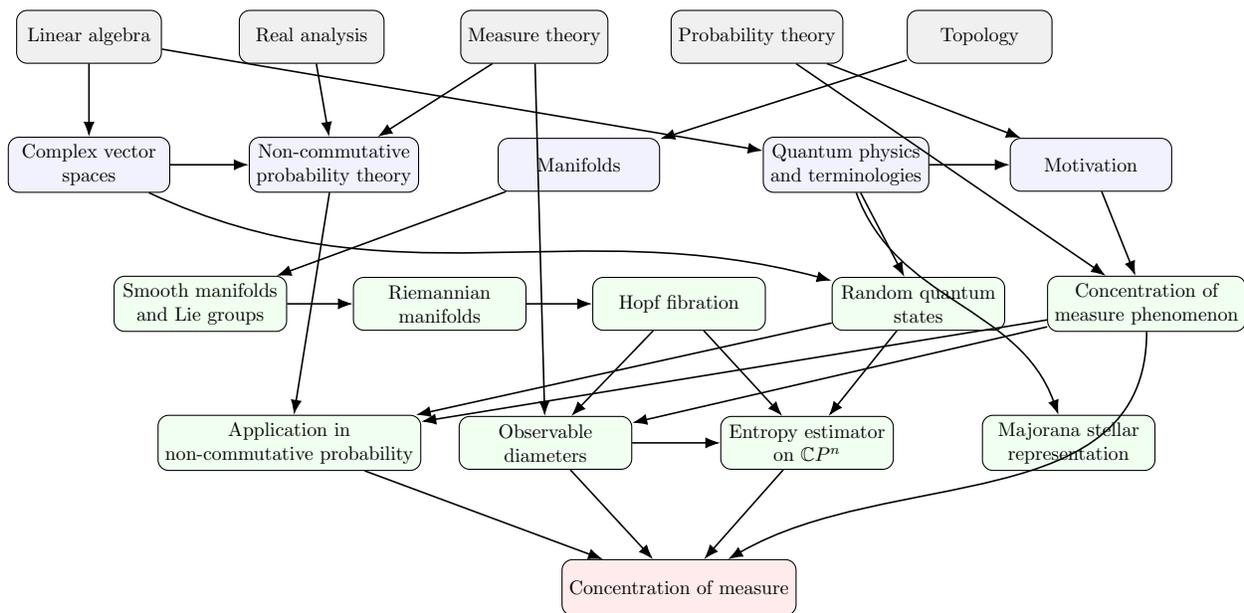
Figure 1: Dependency graph for the thesis. The central path runs from undergraduate background through geometry and quantum theory to concentration of measure, while the Majorana branch records a later exploratory direction.

Chapter 0: Brief definitions and basic concepts

As the future version of me might forgot everything we have over the summer, as I did for now, I will make a review again from the simple definition to recall the necessary information to tell you why we are here and how we are going to proceed.

This section serve as reference for definitions, notations, and theorems that we will use later. This section can be safely ignored if you are already familiar with the definitions and theorems.

But for the future self who might have no idea what I'm talking about, we will provided detailed definitions to you to understand the concepts.

0.1 Complex vector spaces

The main vector space we are interested in is $\mathbb{C}^n$; therefore, all the linear operators we defined are from $\mathbb{C}^n$ to $\mathbb{C}^n$.

Definition 1. *We denote a vector in vector space as $|\psi\rangle = (z_1, \ldots, z_n)$ (might also be infinite dimensional, and $z_i \in \mathbb{C}$).*

Here ψ is just a label for the vector, and you don't need to worry about it too much. This is also called the ket, where the counterpart $\langle\psi|$ is called the bra, used to denote the vector dual to ψ; such an element is a linear functional if you really want to know what that is.

Few additional notation will be introduced, in this document, we will follows the notation used in mathematics literature [Axl23]

- $\langle\psi|\varphi\rangle$ is the inner product between two vectors, and $\langle\psi| A |\varphi\rangle$ is the inner product between $A |\varphi\rangle$ and $\langle\psi|$, or equivalently $A^\dagger \langle\psi|$ and $|\varphi\rangle$.

- Given a complex matrix $A = \mathbb{C}^{n \times n}$,

  1. $\overline{A}$ is the complex conjugate of A.

> **Example**
>
> $$A = \begin{bmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \\ 7+i & 8+i & 9+i \end{bmatrix}, \overline{A} = \begin{bmatrix} 1-i & 2-i & 3-i \\ 4-i & 5-i & 6-i \\ 7-i & 8-i & 9-i \end{bmatrix}$$

2. $A^\top$ denotes the transpose of A.

Example

$$A = \begin{bmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \\ 7+i & 8+i & 9+i \end{bmatrix}, A^\top = \begin{bmatrix} 1+i & 4+i & 7+i \\ 2+i & 5+i & 8+i \\ 3+i & 6+i & 9+i \end{bmatrix}$$
---

3. $A^* = \overline{(A^\top)}$ denotes the complex conjugate transpose, referred to as the adjoint, or Hermitian conjugate of A.

Example

$$A = \begin{bmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \\ 7+i & 8+i & 9+i \end{bmatrix}, A^* = \begin{bmatrix} 1-i & 4-i & 7-i \\ 2-i & 5-i & 8-i \\ 3-i & 6-i & 9-i \end{bmatrix}$$
---

4. A is unitary if $A^*A = AA^* = I$.

5. A is self-adjoint (hermitian in physics literature) if $A^* = A$.

Motivation of Tensor product

Recall from the traditional notation of product space of two vector spaces V and W, that is, $V \times W$, is the set of all ordered pairs $(|v\rangle, |w\rangle)$ where $|v\rangle \in V$ and $|w\rangle \in W$.

The space has dimension $\dim V + \dim W$.

We want to define a vector space with the notation of multiplication of two vectors from different vector spaces.

That is

$$(|v_1\rangle + |v_2\rangle) \otimes |w\rangle = (|v_1\rangle \otimes |w\rangle) + (|v_2\rangle \otimes |w\rangle)$$

$$|v\rangle \otimes (|w_1\rangle + |w_2\rangle) = (|v\rangle \otimes |w_1\rangle) + (|v\rangle \otimes |w_2\rangle)$$

and enables scalar multiplication by

$$\lambda(|v\rangle \otimes |w\rangle) = (\lambda |v\rangle) \otimes |w\rangle = |v\rangle \otimes (\lambda |w\rangle)$$

And we wish to build a way to associate the basis of V and W with the basis of $V \otimes W$. That makes the tensor product a vector space with dimension $\dim V \times \dim W$.

Definition 2. *Definition of linear functional*

A linear functional is a linear map from V to $\mathbb{F}$.

Note the difference between a linear functional and a linear map.

A generalized linear map is a function $f : V \to W$ satisfying the condition.

4

- $f(|u\rangle + |v\rangle) = f(|u\rangle) + f(|v\rangle)$

- $f(\lambda |v\rangle) = \lambda f(|v\rangle)$

Definition 3. *A bilinear functional is a bilinear function $\beta : V \times W \to \mathbb{F}$ satisfying the condition that $|v\rangle \to \beta(|v\rangle, |w\rangle)$ is a linear functional for all $|w\rangle \in W$ and $|w\rangle \to \beta(|v\rangle, |w\rangle)$ is a linear functional for all $|v\rangle \in V$.*

The vector space of all bilinear functionals is denoted by $\mathcal{B}(V, W)$.

Definition 4. *Let V, W be two vector spaces.*

Let V' and W' be the dual spaces of V and W, respectively, that is $V' = \{\psi : V \to \mathbb{F}\}$ and $W' = \{\phi : W \to \mathbb{F}\}$, ψ, ϕ are linear functionals.

The tensor product of vectors $v \in V$ and $w \in W$ is the bilinear functional defined by $\forall(\psi, \phi) \in V' \times W'$ given by the notation

$$(v \otimes w)(\psi, \phi) = \psi(v)\phi(w)$$

The tensor product of two vector spaces V and W is the vector space $\mathcal{B}(V', W')$

Notice that the basis of such vector space is the linear combination of the basis of V' and W', that is, if $\{e_i\}$ is the basis of V' and $\{f_j\}$ is the basis of W', then $\{e_i \otimes f_j\}$ is the basis of $\mathcal{B}(V', W')$.

That is, every element of $\mathcal{B}(V', W')$ can be written as a linear combination of the basis.

Since $\{e_i\}$ and $\{f_j\}$ are bases of V' and W', respectively, then we can always find a set of linear functionals $\{\phi_i\}$ and $\{\psi_j\}$ such that $\phi_i(e_j) = \delta_{ij}$ and $\psi_j(f_i) = \delta_{ij}$.

Here $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$ is the Kronecker delta.

$$V \otimes W = \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \phi_i(v) \psi_j(w) : \phi_i \in V', \psi_j \in W' \right\}$$

Note that $\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \phi_i(v) \psi_j(w)$ is a bilinear functional that maps $V' \times W'$ to $\mathbb{F}$.

This enables basis-free construction of vector spaces with proper multiplication and scalar multiplication.

Examples of tensor product for vectors

Let $V = \mathbb{C}^2, W = \mathbb{C}^3$, choose bases $\{|0\rangle, |1\rangle\} \subset V, \{|0\rangle, |1\rangle, |2\rangle\} \subset W$.

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_1 |0\rangle + v_2 |1\rangle \in V, w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = w_1 |0\rangle + w_2 |1\rangle + w_3 |2\rangle \in W$$

.

Then the tensor product $v \otimes w$ is given by

$$v \otimes w = \begin{pmatrix} v_1 w_1 & v_1 w_2 & v_1 w_3 \\ v_2 w_1 & v_2 w_2 & v_2 w_3 \end{pmatrix} \in \mathbb{C}^6$$

Examples of tensor product for vector spaces

Let $V = \mathbb{C}^2, W = \mathbb{C}^3$, choose bases $\{|0\rangle, |1\rangle\} \subset V, \{|0\rangle, |1\rangle, |2\rangle\} \subset W$.
Then a basis of the tensor product is

$$\{|00\rangle, |01\rangle, |02\rangle, |10\rangle, |11\rangle, |12\rangle\},$$

where $|ij\rangle := |i\rangle \otimes |j\rangle$.
An example element of $V \otimes W$ is

$$|\psi\rangle = 2 |0\rangle \otimes |1\rangle + (1 + i) |1\rangle \otimes |0\rangle - i |1\rangle \otimes |2\rangle.$$

With respect to the ordered basis

$$(|00\rangle, |01\rangle, |02\rangle, |10\rangle, |11\rangle, |12\rangle),$$

this tensor corresponds to the coordinate vector

$$|\psi\rangle \longleftrightarrow \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1+i \\ 0 \\ -i \end{pmatrix} \in \mathbb{C}^6.$$

Using the canonical identification

$$\mathbb{C}^2 \otimes \mathbb{C}^3 \cong \mathbb{C}^{2 \times 3},$$

where

$$|i\rangle \otimes |j\rangle \longmapsto E_{ij},$$

the same tensor is represented by the matrix

$$|\psi\rangle \longleftrightarrow \begin{pmatrix} 0 & 2 & 0 \\ 1+i & 0 & -i \end{pmatrix}.$$

Definition 5. *The vector space defined by the tensor product is equipped with the unique inner product $\langle v \otimes w, u \otimes x \rangle_{V \otimes W} : V \otimes W \times V \otimes W \to \mathbb{F}$ defined by*

$$\langle v \otimes w, u \otimes x \rangle = \langle v, u \rangle_V \langle w, x \rangle_W$$

In practice, we ignore the subscript of the vector space and just write $\langle v \otimes w, u \otimes x \rangle = \langle v, u \rangle \langle w, x \rangle$.
Partial trace
Definition 6. *Let T be a linear operator on $\mathcal{H}$, $(e_1, e_2, \cdots, e_n)$ be a basis of $\mathcal{H}$ and $(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)$ be a basis of dual space $\mathcal{H}^*$. Then the trace of T is defined by*

$$\text{Tr}(T) = \sum_{i=1}^{n} \epsilon_i(T(e_i)) = \sum_{i=1}^{n} \langle e_i, T(e_i) \rangle$$

This is equivalent to the sum of the diagonal elements of T.

Definition 7. *Let T be a linear operator on $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are finite-dimensional Hilbert spaces.*

An operator T on $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ can be written as

$$T = \sum_{i=1}^{n} a_i A_i \otimes B_i$$

where A_i is a linear operator on $\mathcal{A}$ and B_i is a linear operator on $\mathcal{B}$.

The $\mathcal{B}$-partial trace of T ($\text{Tr}_{\mathcal{B}}(T) : \mathcal{L}(\mathcal{A} \otimes \mathcal{B}) \to \mathcal{L}(\mathcal{A})$) is the linear operator on $\mathcal{A}$ defined by

$$\text{Tr}_{\mathcal{B}}(T) = \sum_{i=1}^{n} a_i \, \text{Tr}(B_i) A_i$$

Or we can define the map $L_v : \mathcal{A} \to \mathcal{A} \otimes \mathcal{B}$ by

$$L_v(u) = u \otimes v$$

Note that $\langle u, L_v^*(u') \otimes v' \rangle = \langle u, u' \rangle \langle v, v' \rangle = \langle u \otimes v, u' \otimes v' \rangle = \langle L_v(u), u' \otimes v' \rangle$.

Therefore, $L_v^* \sum_j u_j \otimes v_j = \sum_j \langle v, v_j \rangle u_j$.

Then the partial trace of T can also be defined by

Let $\{v_j\}$ be a set of orthonormal basis of $\mathcal{B}$.

$$\text{Tr}_{\mathcal{B}}(T) = \sum_j L_{v_j}^*(T) L_{v_j}$$

Definition 8. *Let T be a linear operator on $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are finite-dimensional Hilbert spaces.*

Let ρ be a state on $\mathcal{B}$ consisting of orthonormal basis $\{v_j\}$ and eigenvalue $\{\lambda_j\}$.

The partial trace of T with respect to ρ is the linear operator on $\mathcal{A}$ defined by

$$\text{Tr}_{\mathcal{B},\rho}(T) = \sum_j \lambda_j L_{v_j}^*(T) L_{v_j}$$

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

0.2   Non-commutative probability theory

The constructions above explain why tensor products and traces appear before probability is mentioned again: they are the algebraic devices that let composite quantum systems behave like probabilistic systems with marginals and expectations. The next section packages these operations into the operator-theoretic language of states, observables, and expectation values, which is the setting used later for random quantum states and entropy.

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 9. *Hilbert space:*

A Hilbert space is a complete inner product space.

That is, a vector space equipped with an inner product, with the induced metric defined by the norm of the inner product, we have a metric space, which is complete. Reminds that complete mean that every Cauchy sequence, the sequence such that for any $\epsilon > 0$, there exists an N such that for all $m, n \geq N$, we have $|x_m - x_n| < \epsilon$, converges to a limit.

As a side note we will use later, we also defined the Borel measure on a space, here we use the following definition specialized for the space (manifolds) we are interested in.

Definition 10. *Borel measure:*

Let X be a topological space, then a Borel measure $\mu : \mathcal{B}(X) \to [0, \infty]$ on X is a measure on the Borel σ-algebra of X $\mathcal{B}(X)$ satisfying the following properties:

1. *$X \in \mathcal{B}(X)$.*

2. *Close under complement: If $A \in \mathcal{B}(X)$, then $A^c \in \mathcal{B}(X)$.*

3. *Close under countable unions: If $E_1, E_2, \cdots$ are disjoint Borel sets, then $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$.*

In later sections, we will use Lebesgue measure, and Haar measure for various circumstances, their detailed definition may be introduced in later sections.

Example

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

Proposition 11. *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. $\square$

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 function on Ω that is

1. **square integrable**: 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

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

2. **complex-valued**: functions are complex-valued measurable. $f = u + vi$ is complex-valued if u and v are real-valued measurable.

> **Example**
>
> **Proposition 12.** $L^2(\Omega, \mathcal{F}, P)$ *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:
> $$\langle f, f \rangle \geq 0$$
>
> $\square$

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

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 $\mathcal{P}$.

The detailed definition of the non-commutative probability space is given below:
Definition 13. *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 $\mathcal{B}(\mathcal{H})$.

Recall from classical probability theory, we call the initial probability distribution for possible outcomes in the classical probability theory as our *state*, similarly, we need to define the *state* in

the non-commutative probability theory.

Definition 14. *Non-commutative probability state:*

Given a non-commutative probability space $(\mathcal{B}(\mathcal{H}), \mathcal{P})$,

A state is a unit vector $|\psi\rangle$ in the Hilbert space $\mathcal{H}$, such that $\langle\psi|\,|\psi\rangle = 1$.

Every state uniquely defines a map $\rho : \mathcal{P} \to [0,1]$, $\rho(P) = \langle\psi|\,P\,|\psi\rangle$ (commonly named as density operator) 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 $\rho(P_1 + P_2 + \cdots + P_n) = \sum_{i=1}^{n} \rho(P_i)$.*

Note that the pure states are the density operators that can be represented by a unit vector $|\psi\rangle$ in the Hilbert space $\mathcal{H}$, whereas mixed states are the density operators that cannot be represented by a unit vector in the Hilbert space $\mathcal{H}$.

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.)

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 15. *Observable:*

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

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

Satisfies the following properties:

- *$P_A(\emptyset) = O$ (the zero projection)*

- *$P_A(\mathbb{R}) = I$ (the identity projection)*

- *For any sequence $A_1, A_2, \cdots, A_n \in \mathcal{B}(\mathbb{R})$, the following holds:*
  - *$P_A(\bigcup_{i=1}^{n} A_i) = \bigvee_{i=1}^{n} P_A(A_i)$*
  - *$P_A(\bigcap_{i=1}^{n} A_i) = \bigwedge_{i=1}^{n} P_A(A_i)$*
  - *$P_A(A^c) = I - P_A(A), \forall A \in \mathcal{B}(\mathbb{R})$*

If A is an observable determined by the map $P_A : \mathcal{B}(\mathbb{R}) \to \mathcal{P}(\mathcal{H})$, P_A is a spectral measure (a complete additive orthogonal projection valued measure on $\mathcal{B}(\mathbb{R})$). And every spectral measure can be represented by an observable. [Par05]

Proposition 16. *If A_j are mutually disjoint (that is $P_A(A_i)P_A(A_j) = P_A(A_j)P_A(A_i) = O$ for $i \neq j$), then $P_A(\bigcup_{j=1}^n A_j) = \sum_{j=1}^n P_A(A_j)$*

Definition 17. *Probability of a random variable:*

Let A be a real-valued observable on a Hilbert space $\mathcal{H}$. ρ be a state. The probability of observing the outcome $E \in \mathcal{B}(\mathbb{R})$ is given by:

$$\mu(E) = \mathrm{Tr}(\rho P_A(E))$$

Restriction of a quantum state to a commutative subalgebra defines an ordinary probability measure.

Example

Let
$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The eigenvalues of Z are $+1$ and -1, with corresponding normalized eigenvectors

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The spectral projections are

$$P_Z(\{1\}) = |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad P_Z(\{-1\}) = |1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The associated projection-valued measure P_Z satisfies

$$P_Z(\{1, -1\}) = I, \qquad P_Z(\emptyset) = 0.$$

Let
$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The normalized eigenvectors of X are

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \qquad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle),$$

with eigenvalues $+1$ and -1, respectively.
The corresponding spectral projections are

$$P_X(\{1\}) = |+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

$$P_X(\{-1\}) = |-\rangle\langle -| = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Compute
$$P_Z(\{1\})P_X(\{1\}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

On the other hand,

$$P_X(\{1\})P_Z(\{1\}) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Since

$$P_Z(\{1\})P_X(\{1\}) \neq P_X(\{1\})P_Z(\{1\}),$$

the projections do not commute.

Let ρ be a density operator on $\mathbb{C}^2$, i.e.

$$\rho \geq 0, \qquad \mathrm{Tr}(\rho) = 1.$$

For a pure state $|\psi\rangle$, one has

$$\rho = |\psi\rangle \langle\psi|.$$

The probability that a measurement associated with a PVM P yields an outcome in a Borel set $A \in \mathcal{B}$ is

$$\mathbb{P}(A) = \mathrm{Tr}(\rho\,P(A)).$$

For example, let

$$\rho = |0\rangle \langle0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then

$$\mathrm{Tr}\big(\rho\,P_Z(\{1\})\big) = 1, \qquad \mathrm{Tr}\big(\rho\,P_X(\{1\})\big) = \frac{1}{2}.$$

Definition 18. *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 $\sum_{x \in X} M_x^* M_x = I$.

Here is Table 1 summarizing the analog of classical probability theory and non-commutative (*quantum*) probability theory [Fer]:

Table 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
$f = \sum_{\lambda \in \mathrm{Range}(f)} \lambda \mathbb{I}_{f^{-1}(\{\lambda\})}$	$A = \sum_{\lambda \in \mathrm{sp}(A)} \lambda P(\lambda)$
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 $\mathrm{Tr}(\rho) = 1$
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 $\mathbb{E}_\rho(A) = \mathrm{Tr}(\rho A)$
Variance of random variable f is $\mathrm{Var}_\mu(f) = \sum_{i=1}^{n}(f(i) - \mathbb{E}_\mu(f))^2 \mu(\{i\})$	Variance of operator A is $\mathrm{Var}_\rho(A) = \mathrm{Tr}(\rho A^2) - \mathrm{Tr}(\rho A)^2$
Covariance of random variables f and g is $\mathrm{Cov}_\mu(f,g) = \sum_{i=1}^{n}(f(i) - \mathbb{E}_\mu(f))(g(i) - \mathbb{E}_\mu(g))\mu(\{i\})$	Covariance of operators A and B is $\mathrm{Cov}_\rho(A,B) = \mathrm{Tr}(\rho A \circ B) - \mathrm{Tr}(\rho A)\mathrm{Tr}(\rho B)$
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 $\mathrm{Tr}_2(\rho)$

0.3   Manifolds

Up to this point the emphasis has been algebraic and probabilistic. The concentration results used later, however, live naturally on curved spaces equipped with metrics and measures. For that reason the discussion now shifts from operator theory to manifold theory, starting with topological manifolds and then adding smooth and Riemannian structure until we can describe complex projective space as a genuine geometric state space.

In this section, we will introduce some basic definitions and theorems used in manifold theory that are relevant to our study. Assuming no prior knowledge of manifold theory but basic topology understanding. We will provide brief definitions and explanations for each term. From the most abstract Manifold definition to the Riemannian manifolds and related theorems.

0.3.1 Manifolds

Definition 19. *An m-manifold is a topological space X that is*

1. *Hausdorff: every distinct two points in X can be separated by two disjoint open sets.*

2. *Second countable: X has a countable basis.*

3. *Every point p has an open neighborhood $p \in U$ that is homeomorphic to an open subset of $\mathbb{R}^m$.*

Example

Let $X = \mathbb{R}$ and $\mathcal{B} = \{(a,b) | a, b \in \mathbb{R}, a < b\}$ (collection of all open intervals with rational endpoints).
Since the rational numbers are countable, so $\mathcal{B}$ is countable.
So $\mathbb{R}$ is second countable.
Likewise, $\mathbb{R}^n$ is also second countable.

Example

1-manifold is a curve and 2-manifold is a surface.

Theorem 20. *Whithney's Embedding Theorem:*

If X is a compact m-manifold, then X can be imbedded in $\mathbb{R}^n$ for some n.

This proof is from topology course, and use additional one lemma:
Lemma 21. *Let $\{U_i\}_{i=1}^n$ be a finite open cover of a normal space X (Every pair of closed sets in X can be separated by two open sets in X).*

Then there exists a partition of unity dominated by $\{U_i\}_{i=1}^n$.

Proof. Since X is a m compact manifold, $\forall x \in X$, there is an open neighborhood U_x of x such that U_x is homeomorphic to $\mathbb{R}^m$. That means there exists $\varphi_i : U_x \to \varphi(U_x) \subseteq \mathbb{R}^m$.

Where $\{U_x\}_{x \in X}$ is an open cover of X. Since X is compact, there is a finite subcover $\bigcup_{i=1}^k U_{x_i} = X$.

Apply the existence of a finite partition of unity, we can find a partition of unity dominated by $\{U_{x_i}\}_{i=1}^k$. With family of functions $\phi_i : \mathbb{R}^d \to [0,1]$.

Define $h_i : X \to \mathbb{R}^m$ by

$$h_i(x) = \begin{cases} \phi_i(x)\varphi_i(x) & \text{if } x = x_i \\ 0 & \text{otherwise} \end{cases}$$

We claim that h_i is continuous using pasting lemma.

On U_i, $h_i = \phi_i\varphi_i$ is product of two continuous functions therefore continuous.

On $X - \text{supp}(\phi_i)$, $h_i = 0$ is continuous.

By pasting lemma, h_i is continuous.

Define

$$F : X \to (\mathbb{R}^m \times \mathbb{R})^n$$

where $x \mapsto (h_1(x), \varphi_1(x), h_2(x), \varphi_2(x), \ldots, h_n(x), \varphi_n(x))$

We want to show that F is imbedding map.

1. F is continuous

   since it is a product of continuous functions.

2. F is injective

   that is, if $F(x_1) = F(x_2)$, then $x_1 = x_2$.

   By partition of unity, we have,

   $h_1(x_1) = h_1(x_2), h_2(x_1) = h_2(x_2), \ldots, h_n(x_1) = h_n(x_2)$.

   And $\varphi_1(x_1) = \varphi_1(x_2), \varphi_2(x_1) = \varphi_2(x_2), \ldots, \varphi_n(x_1) = \varphi_n(x_2)$.

   Because $\sum_{i=1}^n \varphi_i(x_1) = 1$, therefore there exists $\varphi_i(x_1) = \varphi_i(x_2) > 0$.

   Therefore $x1, x_2 \in \text{supp}(\phi_i) \subseteq U_i$.

   By definition of h, $h_i(x_1) = h_i(x_2)$, $\varphi_i(x_1)\phi_i(x_1) = \varphi_i(x_2)\phi_i(x_2)$.

   Using cancellation, $\phi_i(x_1) = \phi_i(x_2)$.

   Therefore $x_1 = x_2$ since $\phi_i(x_1) = \phi_i(x_2)$ is a homeomorphism.

   In this proof, φ ensures the imbedding is properly defined on the open sets

3. F is a homeomorphism.

   Note that if $f : X \to Y$ is continuous and X is compact, Y is Hausdorff, then f is a closed map.

   $F : X \to F(X)$ is a bijective map from a compact space to a Hausdorff space, therefore F is a closed map.

   Since F is continuous, then $F^{-1}(C)$ where C is a closed set in $F(X)$, $F^{-1}(C)$ is closed in X.

   Therefore F is a homeomorphism.

   $\square$

0.3.2 Smooth manifolds and Lie groups

This section is adopted from [LL12]

The topological definition of a manifold tells us what the space looks like locally, but not how to differentiate on it. The next step is therefore to add charts with smooth transition maps. Once this smooth structure is available, notions such as differentials, submersions, and group actions can be stated precisely, and these are exactly the tools needed later for the Hopf fibration.

Definition 22. *Let $U \subseteq \mathbb{R}^n$ and $f : U \to \mathbb{R}^n$ be a map.*

For any $a = (a_1, \cdots, a_n) \in U$, $j \in \{1, \cdots, n\}$, the j-th partial derivative of F at a is defined as

$$\frac{\partial f}{\partial x_j}(a) = \lim_{h \to 0} \frac{f(a_1, \cdots, a_j + h, \cdots, a_n) - f(a_1, \cdots, a_j, \cdots, a_n)}{h}$$
$$= \lim_{h \to 0} \frac{f(a + he_j) - f(a)}{h}$$

Definition 23. *Let $U \subseteq \mathbb{R}^n$ and $f : U \to \mathbb{R}^n$ be a map.*

If for any $j \in \{1, \cdots, n\}$, the j-th partial derivative of f is continuous at a, then f is continuously differentiable at a.

If $\forall a \in U$, $\frac{\partial f}{\partial x_j}$ exists and is continuous at a, then f is continuously differentiable on U, or a C^1 map. (Note that C^0 map is just a continuous map.)

Definition 24. *A function $f : U \to \mathbb{R}^n$ is smooth if it is of class C^k for every $k \geq 0$ on U. Such function is called a **diffeomorphism** if it is also a **bijection** and its **inverse is also smooth**.*

Definition 25. *Let M be a smooth manifold. A **chart** is a pair (U, φ) where $U \subseteq M$ is an open subset and $\varphi : U \to \hat{U} \subseteq \mathbb{R}^n$ is a homeomorphism (a continuous bijection map and its inverse is also continuous).*

If $p \in U$ and $\varphi(p) = 0$, then we say that p is the origin of the chart (U, φ).

*For $p \in U$, we note that the continuous function $\varphi(p) = (x_1(p), \cdots, x_n(p))$ gives a vector in $\mathbb{R}^n$. The $(x_1(p), \cdots, x_n(p))$ is called the **local coordinates** of p in the chart (U, φ).*

Definition 26. *Let M be a smooth manifold. An **atlas** is a collection of charts $\mathcal{A} = \{(U_\alpha, \phi_\alpha)\}_{\alpha \in I}$ such that $M = \bigcup_{\alpha \in I} U_\alpha$.*

*An atlas is said to be **smooth** if the transition maps $\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \to \phi_\alpha(U_\alpha \cap U_\beta)$ are smooth for all $\alpha, \beta \in I$.*

Definition 27. *A smooth manifold is a pair $(M, \mathcal{A})$ where M is a topological manifold and $\mathcal{A}$ is a smooth atlas.*

Definition 28. *Let M and N be smooth manifolds, and $f : M \to N$ be a smooth map. For each $p \in M$, the **differential** of f at p is the linear map*

$$df_p : T_p M \to T_{f(p)} N$$

Definition 29. *A smooth map $f : M \to N$ is a **smooth submersion** if for each $p \in M$, the differential $df_p : T_p M \to T_{f(p)} N$ is surjective.*

Or equivalently $\operatorname{rank}(df_p) = \dim N$ for each $p \in M$.

Here are some additional propositions that will be helpful for our study in later sections:

This one is from [LL12] Theorem 4.26

Theorem 30. *Let M and N be smooth manifolds and $\pi : M \to N$ is a smooth map. Then π is a smooth submersion if and only if every point of M is in the image of a smooth local section of π (a local section of π is a map $\sigma : U \to M$ defined on some open subset $U \subseteq N$ with $\pi \circ \sigma = Id_U$).*

0.3.3 Riemannian manifolds

Smooth manifolds still do not measure lengths, angles, or volumes. To connect the manifold side of the thesis with concentration of measure, we need a metric structure that turns local smooth data into global geometric data. Riemannian metrics provide exactly that extra layer, and Riemannian submersions will later transfer this geometry from spheres to complex projective spaces.

Definition 31. *Let M be a smooth manifold. A **Riemannian metric** on M is a smooth covariant tensor field $g \in \mathcal{T}^2(M)$ such that for each $p \in M$, g_p is an inner product on T_pM.*

$g_p(v,v) \geq 0$ for each $p \in M$ and each $v \in T_pM$. equality holds if and only if $v = 0$.

Definition 32. *Suppose $(\tilde{M}, \tilde{g})$ and (M, g) are smooth Riemannian manifolds, and $\pi : \tilde{M} \to M$ is a smooth submersion. Then π is said to be a **Riemannian submersion** if for each $x \in \tilde{M}$, the differential $d\pi_x : T_x\tilde{M} \to T_{\pi(x)}M$ restricts to a linear isometry from H_x onto $T_{\pi(x)}M$.*

In other words, $\tilde{g}_x(v, w) = g_{\pi(x)}(d\pi_x(v), d\pi_x(w))$ whenever $v, w \in H_x$.

Theorem 33. *Let $(\tilde{M}, \tilde{g})$ be a Riemannian manifold, let $\pi : \tilde{M} \to M$ be a surjective smooth submersion, and let G be a group acting on $\tilde{M}$. If the **action** is*

1. *isometric: the map $x \mapsto \varphi \cdot x$ is an isometry for each $\varphi \in G$.*

2. *vertical: every element $\varphi \in G$ takes each fiber to itself, that is $\pi(\varphi \cdot p) = \pi(p)$ for all $p \in \tilde{M}$.*

3. *transitive on fibers: for each $p, q \in \tilde{M}$ such that $\pi(p) = \pi(q)$, there exists $\varphi \in G$ such that $\varphi \cdot p = q$.*

Then there is a unique Riemannian metric on M such that π is a Riemannian submersion.

Proof. For each $p \in \tilde{M}$, let

$$V_p := \ker(d\pi_p) \subseteq T_p\tilde{M}$$

be the vertical space, and let

$$H_p := V_p^{\perp_{\tilde{g}}}$$

be its $\tilde{g}$-orthogonal complement. Since π is a surjective smooth submersion, each $d\pi_p : T_p\tilde{M} \to T_{\pi(p)}M$ is surjective, so

$$T_p\tilde{M} = V_p \oplus H_p,$$

and therefore the restriction

$$d\pi_p|_{H_p} : H_p \to T_{\pi(p)}M$$

is a linear isomorphism.

We first show that the group action preserves the horizontal distribution. Fix $\varphi \in G$. Since the action is vertical, we have

$$\pi(\varphi \cdot x) = \pi(x) \qquad \text{for all } x \in \tilde{M}.$$

Differentiating at p gives

$$d\pi_{\varphi \cdot p} \circ d\varphi_p = d\pi_p.$$

Hence if $v \in V_p = \ker(d\pi_p)$, then

$$d\pi_{\varphi \cdot p}(d\varphi_p v) = d\pi_p(v) = 0,$$

so $d\varphi_p(V_p) \subseteq V_{\varphi \cdot p}$. Since φ acts isometrically, $d\varphi_p$ is a linear isometry, and thus it preserves orthogonal complements. Therefore

$$d\varphi_p(H_p) = H_{\varphi \cdot p}.$$

We now define a metric on M. Let $m \in M$, and choose any $p \in \pi^{-1}(m)$. For $u, v \in T_m M$, let $\tilde{u}, \tilde{v} \in H_p$ be the unique horizontal lifts satisfying

$$d\pi_p(\tilde{u}) = u, \qquad d\pi_p(\tilde{v}) = v.$$

Define

$$g_m(u, v) := \tilde{g}_p(\tilde{u}, \tilde{v}).$$

This is a symmetric bilinear form on $T_m M$, and it is positive definite because $\tilde{g}_p$ is positive definite on H_p and $d\pi_p|_{H_p}$ is an isomorphism.

It remains to show that this definition is independent of the choice of p in the fiber. Suppose $p, q \in \pi^{-1}(m)$. By transitivity of the action on fibers, there exists $\varphi \in G$ such that $\varphi \cdot p = q$. Let $\tilde{u}_p, \tilde{v}_p \in H_p$ be the horizontal lifts of u, v at p, and define

$$\tilde{u}_q := d\varphi_p(\tilde{u}_p), \qquad \tilde{v}_q := d\varphi_p(\tilde{v}_p).$$

By the previous paragraph, $\tilde{u}_q, \tilde{v}_q \in H_q$. Moreover,

$$d\pi_q(\tilde{u}_q) = d\pi_q(d\varphi_p \tilde{u}_p) = d\pi_p(\tilde{u}_p) = u,$$

and similarly $d\pi_q(\tilde{v}_q) = v$. Thus $\tilde{u}_q, \tilde{v}_q$ are exactly the horizontal lifts of u, v at q. Since φ is an isometry,

$$\tilde{g}_q(\tilde{u}_q, \tilde{v}_q) = \tilde{g}_q(d\varphi_p \tilde{u}_p, d\varphi_p \tilde{v}_p) = \tilde{g}_p(\tilde{u}_p, \tilde{v}_p).$$

Therefore $g_m(u, v)$ is independent of the chosen point $p \in \pi^{-1}(m)$, so g is well defined on M.

Next we prove that g is smooth. Let $m_0 \in M$. Since π is a smooth submersion, there exists an open neighborhood $U \subseteq M$ of m_0 and a smooth local section

$$s : U \to \tilde{M} \qquad \text{such that} \qquad \pi \circ s = \mathrm{id}_U.$$

Over $s(U)$, the vertical bundle $V = \ker d\pi$ is a smooth subbundle of $T\tilde{M}$, and hence so is its orthogonal complement $H = V^\perp$. For each $x \in U$, the restriction

$$d\pi_{s(x)}|_{H_{s(x)}} : H_{s(x)} \to T_x M$$

is a linear isomorphism, and these isomorphisms depend smoothly on x. Thus they define a smooth vector bundle isomorphism

$$d\pi|_H : H|_{s(U)} \to TU,$$

whose inverse is also smooth.

If X, Y are smooth vector fields on U, define their horizontal lifts along s by

$$X_x^H := \left(d\pi_{s(x)}|_{H_{s(x)}} \right)^{-1} (X_x), \qquad Y_x^H := \left(d\pi_{s(x)}|_{H_{s(x)}} \right)^{-1} (Y_x).$$

19

Then X^H and Y^H are smooth vector fields along $s(U)$, and by construction,

$$g(X, Y)(x) = \tilde{g}_{s(x)}(X_x^H, Y_x^H).$$

Since the right-hand side depends smoothly on x, it follows that g is a smooth Riemannian metric on M.

By construction, for every $p \in \tilde{M}$ and every $\tilde{u}, \tilde{v} \in H_p$,

$$g_{\pi(p)}(d\pi_p \tilde{u}, d\pi_p \tilde{v}) = \tilde{g}_p(\tilde{u}, \tilde{v}).$$

Thus $d\pi_p : H_p \to T_{\pi(p)}M$ is an isometry for every p, so $\pi : (\tilde{M}, \tilde{g}) \to (M, g)$ is a Riemannian submersion.

Finally, uniqueness is immediate. Indeed, if g' is another Riemannian metric on M such that $\pi : (\tilde{M}, \tilde{g}) \to (M, g')$ is a Riemannian submersion, then for any $m \in M$, any $p \in \pi^{-1}(m)$, and any $u, v \in T_m M$, letting $\tilde{u}, \tilde{v} \in H_p$ denote the horizontal lifts of u, v, we must have

$$g'_m(u, v) = \tilde{g}_p(\tilde{u}, \tilde{v}) = g_m(u, v).$$

Hence $g' = g$.

Therefore there exists a unique Riemannian metric on M such that π is a Riemannian submersion.

$\square$

0.3.4   Hopf fibration

The previous subsection gives the abstract mechanism for pushing a metric through a quotient map. The Hopf fibration is the concrete instance needed in this thesis: it explains why the geometry of the sphere descends to complex projective space, and therefore why concentration on spheres is relevant to the geometry of pure quantum states.

There are some remaining steps for showing how the metric on Sphere induces the metric on complex projective space, now we will just drop the conclusion here so that we can continue our discussion:

- Every nonzero vector can be normalized, so each pure state has a representative on the unit sphere
$$S^{2n+1} \subset \mathbb{C}^{n+1}.$$

- Two unit vectors represent the same pure state exactly when they differ by a phase:
$$z \sim e^{i\theta} z.$$

- Therefore
$$\mathbb{C}P^n = S^{2n+1}/S^1.$$

The quotient map
$$p : S^{2n+1} \to \mathbb{C}P^n, \qquad p(z) = [z] = \{\lambda z : \lambda \in \mathbb{C}^\times\},$$

is the **Hopf fibration**.

The geometric picture is

$$S^{2n+1} \xrightarrow{\text{Hopf fibration}} \mathbb{C}P^n, \qquad \text{round metric} \rightsquigarrow \text{Fubini–Study metric}.$$

The sphere $S^{2n+1} \subset \mathbb{C}^{n+1}$ has the **round metric**

$$g_{\text{round}} = \sum_{j=0}^{n}(dx_j^2 + dy_j^2)\big|_{S^{2n+1}},$$

induced from the Euclidean metric on $\mathbb{R}^{2n+2}$.

In homogeneous coordinates $[z] \in \mathbb{C}P^n$, the **Fubini–Study metric** is

$$g_{FS} = \frac{\langle dz, dz \rangle \langle z, z \rangle - |\langle z, dz \rangle|^2}{\langle z, z \rangle^2}.$$

0.4 Quantum physics and terminologies

The geometric discussion above identifies the right state space, but the thesis ultimately studies physical observables on that space. We now return to quantum terminology and translate the geometric objects into the language of states, measurements, Haar sampling, and reduced density matrices. This is the point where the manifold picture and the operator picture meet.

In this section, we will introduce some terminologies and theorems used in quantum physics that are relevant to our study. Assuming no prior knowledge of quantum physics, we will provide brief definitions and explanations for each term.

One might ask, what is the fundamental difference between a quantum system and a classical system, and why can we not directly apply those theorems in classical computers to a quantum computer? It turns out that quantum error-correcting codes are hard due to the following definitions and features for quantum computing.

Definition 34. *All quantum operations can be constructed by composing four kinds of transformations: (adapted from Chapter 10 of [BZ17])*

1. *Unitary operations. $U(\cdot)$ for any quantum state. It is possible to apply a non-unitary operation for an open quantum system, but that is usually not the focus for quantum computing and usually leads to non-recoverable loss of information that we wish to obtain.*

2. *Extend the system. Given a quantum state $\rho \in \mathcal{H}^N$, we can extend it to a larger quantum system by "entangle" (For this report, you don't need to worry for how quantum entanglement works) it with some new states $\sigma \in \mathcal{H}^K$ (The space where the new state dwells is usually called ancilla system) and get $\rho' = \rho \otimes \sigma \in \mathcal{H}^N \otimes \mathcal{K}$.*

3. *Partial trace. Given a quantum state $\rho \in \mathcal{H}^N$ and some reference state $\sigma \in \mathcal{H}^K$, we can trace out some subsystems and get a reduced state on the remaining subsystem.*

4. Selective measurement. Given a quantum state, we measure it and get a classical bit; unlike the classical case, the measurement is a probabilistic operation. (More specifically, this is some projection to a reference state corresponding to a classical bit output. For this report, you don't need to worry about how such a result is obtained and how the reference state is constructed.)

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

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

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

Lemma 35. *Let $(U(n), \|\cdot\|, \mu)$ be a metric measure space where $\|\cdot\|$ is the Hilbert-Schmidt norm and μ is a Borel probability measure.*

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

That is, for every Borel set $E \subseteq U(n)$ and every $A \in U(n)$, $\mu(AE) = \mu(EA) = \mu(E)$.

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

A finite-dimensional quantum system is modeled by a complex Hilbert space (a complete inner product space)

$$\mathcal{H} \cong \mathbb{C}^{n+1}.$$

A **pure state** is represented by a unit vector

$$\psi \in \mathcal{H}, \qquad \|\psi\| = 1.$$

A **mixed state** is represented by a density matrix

$$\rho = \sum_{j=1}^{n} p_j |\psi_j\rangle\langle\psi_j|, \qquad \sum_{j=1}^{n} p_j = 1, \qquad p_j \geq 0.$$

Some key comparisons between pure states and mixed states:

Pure states describe maximal information; mixed states describe probabilistic mixtures or partial information.

Pure states form a curved geometric space; mixed states form a convex set inside the space of matrices.

Pure states live in the complex projective space.

- Two nonzero vectors that differ by a nonzero complex scalar represent the same physical state:

$$\psi \sim \lambda\psi, \qquad \lambda \in \mathbb{C}^\times.$$

- In particular, multiplying by a phase $e^{i\theta}$ does not change any physical predictions.

- Therefore the physical pure state is not a single vector, but the *complex line* spanned by that vector.

Hence the space of pure states (denoted by $\mathcal{P}(\mathcal{H})$) is

$$\mathcal{P}(\mathcal{H}) = (\mathcal{H} \setminus \{0\})/\mathbb{C}^{\times}.$$

After choosing a basis $\mathcal{H} \cong \mathbb{C}^{n+1}$, this becomes

$$\mathcal{P}(\mathcal{H}) \cong \mathbb{C}P^n.$$

Proposition 36. *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.

$\square$

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.

0.4.1 Random quantum states

The preceding material identifies the spaces and symmetries of quantum states. The next step is probabilistic: once Haar invariance is available, we can speak about random pure states and random mixed states in a way that matches the geometric viewpoint developed earlier. These definitions are the starting point for the concentration statements proved in the next chapter.

First, we need to define what is a random state in a bipartite system.
Definition 37. *Pure state:*

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 by sampling unit vectors on 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 define the rank-s random states.

Definition 38. *Rank-s random state.*

For a system A and an integer $s \geq 1$, consider the distribution on the mixed states $\mathcal{S}(A)$ of A induced by the partial trace over the second factor from 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 state, denoted by $\rho \in_R \mathcal{S}_s(A)$. And $\mathcal{P}(A) = \mathcal{S}_1(A)$.

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 $f : S^n \to \mathbb{R}$ 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.

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.
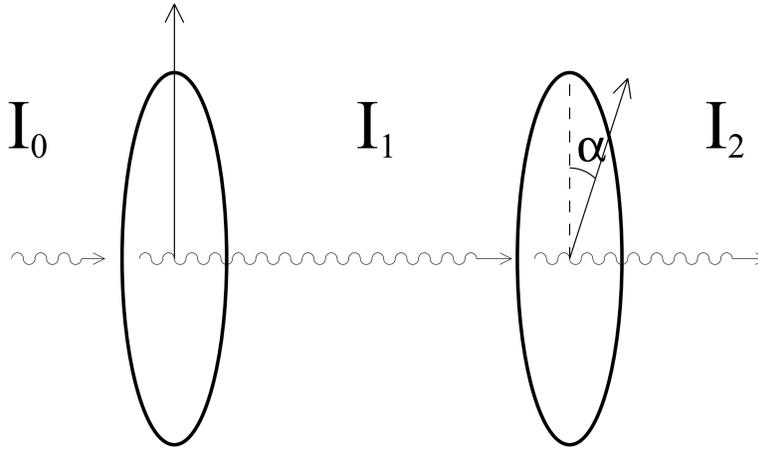

FIG. 1

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

Theorem 39. *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

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

$\square$

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

complement rule, we have

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

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

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

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 $\mathrm{Prob}(P_1 = 1, P_3 = 0)$ 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:

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

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

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

The probability that the photon then passes the second filter is

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

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

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

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 40. *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

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

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 41. *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.

1.2.1 Levy's concentration theorem

Theorem 42. *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 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 43. *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 Poincaré limit theorem. [Ver18]

If $X \sim \text{Unif}(S^n(\sqrt{n}))$, then for any fixed unit vector x we have $\langle X, x \rangle \to N(0, 1)$ in distribution as $n \to \infty$.
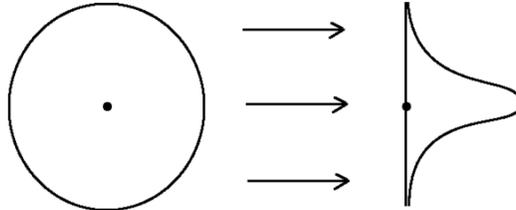
Figure 3.9 The projective central limit theorem: the projection of the uniform distribution on the sphere of radius $\sqrt{n}$ onto a line converges to the normal distribution $N(0, 1)$ as $n \to \infty$.

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

Lemma 44. *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,

$$\frac{d(\pi_{n,k})_*\sigma^n(x)}{dx} = \frac{\text{vol}_{n-k}(\pi_{n,k}^{-1}(x))}{\text{vol}_k(S^n(\sqrt{n}))}$$

$$= \frac{(n - \|x\|^2)^{\frac{n-k}{2}}}{\int_{\|x\|\leq\sqrt{n}}(n - \|x\|^2)^{\frac{n-k}{2}}dx}$$

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

$$\frac{(n - \|x\|^2)^{\frac{n-k}{2}}}{\int_{\|x\|\leq\sqrt{n}}(n - \|x\|^2)^{\frac{n-k}{2}}dx} = \frac{e^{-\frac{\|x\|^2}{2}}}{\int_{x\in\mathbb{R}^k} e^{-\frac{\|x\|^2}{2}}dx}$$

$$= \frac{1}{(2\pi)^{\frac{k}{2}}}e^{-\frac{\|x\|^2}{2}}$$

$$= \frac{d\gamma^k(x)}{dx}$$

$\square$

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.

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

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,

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

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 41, 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 44, 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,

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

$\square$

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.

1.3 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 36) 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 45. *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 46. *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.

1.3.2 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.
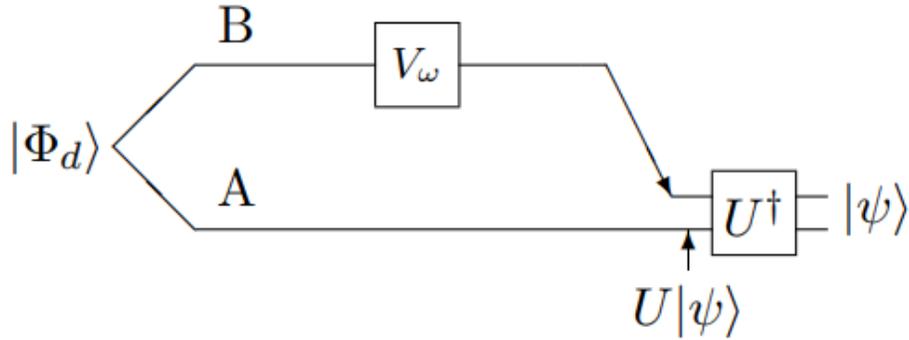
Figure 1.3: Superdense coding, image from [Hay10]

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:

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

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

- $\mathcal{S}(A)$ 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 47. *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].

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

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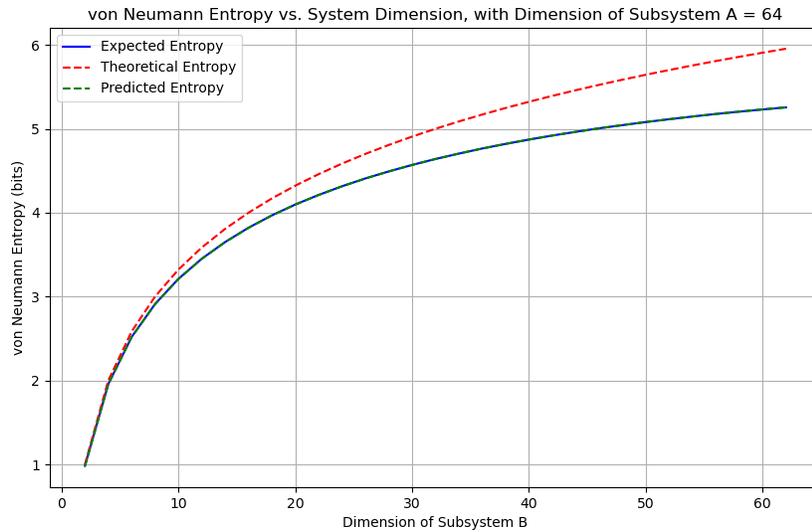

Figure 1.5: Entropy vs dimension

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

Lemma 48. *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 $g : A \otimes B \to \mathbb{R}$ defined as $g(\varphi) = H(M(\varphi_A))$, where $M : A \otimes B \to \mathcal{P}(A)$ is any fixed complete von Neumann measurement and $H : \mathcal{P}(A) \otimes \mathcal{P}(B) \to \mathbb{R}$ 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 $|\varphi\rangle = \sum_{j=1}^{d_A} \sum_{k=1}^{d_B} \varphi_{jk} |e_j\rangle_A |f_k\rangle_B$.

35

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

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

Then the Lipschitz constant of g is

$$\eta^2 = \sup_{\langle\varphi|\varphi\rangle\leq 1} \nabla g \cdot \nabla g$$

$$= \sum_{j=1}^{d_A}\sum_{k=1}^{d_B} \left(\frac{\partial g}{\partial x_{jk}}\right)^2 + \left(\frac{\partial g}{\partial y_{jk}}\right)^2$$

$$= \sum_{j=1}^{d_A}\sum_{k=1}^{d_B} \frac{4(x_{jk}^2 + y_{jk}^2)}{(\ln 2)^2}[1 + \ln p_j(\varphi)]^2$$

$$= \sum_{j=1}^{d_A}\sum_{k=1}^{d_B} \frac{4|\varphi_{jk}|^2}{(\ln 2)^2}[1 + \ln p_j(\varphi)]^2$$

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 $\sum_{j=1}^{d_A} p_j(\varphi) \ln p_j(\varphi) \leq 0$.

$$\sum_{j=1}^{d_A} p_j(\varphi)(1 + \ln p_j(\varphi))^2 = \sum_{j=1}^{d_A} p_j(\varphi)(1 + 2\ln p_j(\varphi) + (\ln p_j(\varphi))^2)$$

$$= 1 + 2\sum_{j=1}^{d_A} p_j(\varphi) \ln p_j(\varphi) + \sum_{j=1}^{d_A} p_j(\varphi)(\ln p_j(\varphi))^2$$

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

36

Thus,

$$\nabla g \cdot \nabla g \leq \frac{4}{(\ln 2)^2}[1 + \sum_{j=1}^{d_A} p_j(\varphi)(\ln p_j(\varphi))^2]$$

$$\leq \frac{4}{(\ln 2)^2}[1 + (\ln d_A)^2]$$

$$\leq 8(\log_2 d_A)^2$$

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 $\sqrt{8}\log_2(d_A)$. $\qquad\square$

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).
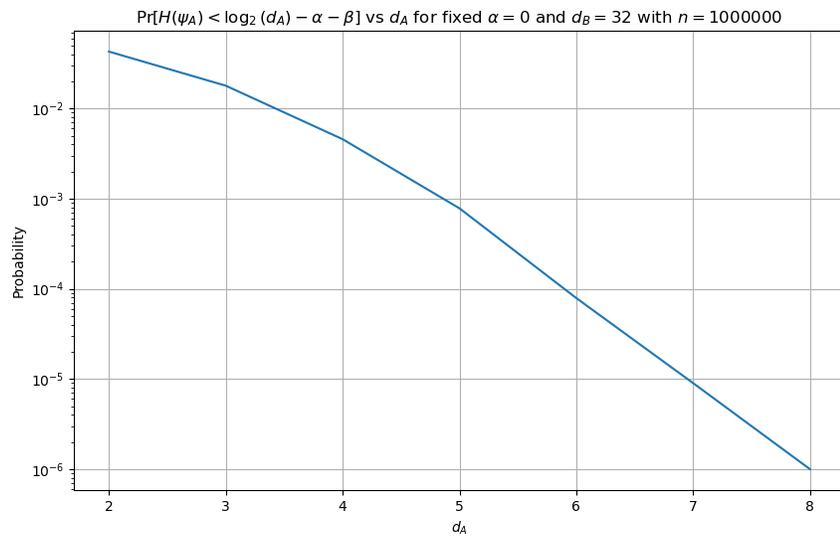
Figure 1.6: Entropy vs d_A

Chapter 2

Levy's family and observable diameters

Abstract

The concentration of measure phenomenon comes with an important notion called the observable diameter, which measures how tightly every 1-Lipschitz function $f : X \to \mathbb{R}$ concentrates on a metric-measure space X outside a set of measure at most κ.

From Hayden's work, we know that for a random pure state in a bipartite system the entropy is nearly maximal and the entropy function has Lipschitz constant upper bounded by $\sqrt{8} \log_2(d_A)$ for $d_A \geq 3$. We can use the entropy function as a proxy for the observable diameter.

Altogether, we study how entropy concentration reflects the geometry of high-dimensional spheres, generic bipartite projective state spaces, and symmetric quantum state manifolds.

In this section, we will explore how the results from Hayden's concentration of measure theorem can be understood in terms of observable diameters from Gromov's perspective and what properties they reveal for entropy functions.

We will try to use the results from previous sections to estimate the observable diameter for complex projective spaces.

2.1   Observable diameters

This chapter starts from the qualitative concentration statements in Chapter 1 and asks for a geometric quantity that records the same phenomenon directly at the level of metric-measure spaces. Observable diameter is that quantity. It translates concentration of real observables into an invariant of the underlying space, which makes it possible to compare spheres and complex projective spaces on the same footing.

Recall from previous sections, 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.

Definition 49. *Let X be a topological space with the following:*

1. *X is complete.*

2. *X is a metric space with metric d_X.*

3. *X has a Borel probability measure μ_X.*

*Then (X, d_X, μ_X) is a **metric measure space**.*

Definition 50. *Let (X, d_X) be a metric space. The **diameter** of a set $A \subset X$ is defined as*

$$\mathrm{diam}(A) = \sup_{x,y \in A} d_X(x, y).$$

Definition 51. *Let (X, d_X, μ_X) be a metric measure space. For any real number $\alpha \leq 1$, the **partial diameter** of X is defined as*

$$\mathrm{diam}(X; \alpha) = \inf_{A \subseteq X, \ \mu_X(A) \geq \alpha} \mathrm{diam}(A).$$

This definition generalizes the relation between the measure and metric in a metric-measure space. Intuitively, a space with smaller partial diameter can carry more mass inside the same diameter constraint.

However, in higher dimensions, the volume may tend to concentrate more around a small neighborhood of a set, as we saw earlier for high-dimensional spheres. We can safely cut away $\kappa > 0$ mass to reduce the diameter significantly, and this yields a better measure of concentration.

Definition 52. *Let X be a metric-measure space, Y be a metric space, and $f : X \to Y$ be a 1-Lipschitz function. Then $f_* \mu_X = \mu_Y$ is a push-forward measure on Y.*

*For any real number $\kappa > 0$, the κ-**observable diameter with screen** Y is defined as*

$$\mathrm{ObserDiam}_Y(X; -\kappa) = \sup\{\mathrm{diam}(f_* \mu_X; 1 - \kappa) \mid f : X \to Y \text{ is 1-Lipschitz}\}.$$

*And the **observable diameter with screen** Y is defined as*

$$\mathrm{ObserDiam}_Y(X) = \inf_{\kappa > 0} \max\{\mathrm{ObserDiam}_Y(X; -\kappa), \kappa\}.$$

*If $Y = \mathbb{R}$, we call it the **observable diameter**.*

If we collapse it naively via

$$\inf_{\kappa > 0} \mathrm{ObserDiam}_Y(X; -\kappa),$$

we typically get something degenerate: as $\kappa \to 1$, the condition "mass $\geq 1 - \kappa$" becomes almost empty, so $\mathrm{diam}(\nu; 1 - \kappa)$ can be forced to be 0 by taking a tiny set of positive mass. Hence the infimum tends to 0 for essentially any non-atomic space.

This is why one either:

1. keeps $\mathrm{ObserDiam}_Y(X; -\kappa)$ as a *function of κ* (picking κ to be small but not 0), or

2. if one insists on a single number, balances "spread" against "exceptional mass" by defining $\text{ObserDiam}_Y(X) = \inf_{\kappa > 0} \max\{\text{ObserDiam}_Y(X; -\kappa), \kappa\}$ as above.

The point of the $\max\{\cdot, \kappa\}$ is that it prevents cheating by taking κ close to 1: if κ is large then the maximum is large regardless of how small $\text{ObserDiam}_Y(X; -\kappa)$ is, so the infimum is forced to occur where the exceptional mass and the observable spread are small.

Few additional propositions in [Shi14] will help us estimate the observable diameter for complex projective spaces.

Proposition 53. *Let X and Y be two metric-measure spaces and $\kappa > 0$, and let $f : Y \to X$ be a 1-Lipschitz function (Y dominates X, denoted as $X \prec Y$). Then:*

1. $\text{diam}(X; 1 - \kappa) \le \text{diam}(Y; 1 - \kappa)$

2. $\text{ObserDiam}(X; -\kappa) \le \text{diam}(X; 1 - \kappa)$, *and* $\text{ObserDiam}(X)$ *is finite.*

3. $\text{ObserDiam}(X; -\kappa) \le \text{ObserDiam}(Y; -\kappa)$

Proof. Since f is 1-Lipschitz, we have $f_*\mu_Y = \mu_X$. Let A be any Borel set of Y with $\mu_Y(A) \ge 1 - \kappa$ and $\overline{f(A)}$ be the closure of $f(A)$ in X. We have $\mu_X(\overline{f(A)}) = \mu_Y(f^{-1}(\overline{f(A)})) \ge \mu_Y(A) \ge 1 - \kappa$ and by the 1-Lipschitz property, $\text{diam}(\overline{f(A)}) \le \text{diam}(A)$, so $\text{diam}(X; 1-\kappa) \le \text{diam}(A) \le \text{diam}(Y; 1-\kappa)$.

Let $g : X \to \mathbb{R}$ be any 1-Lipschitz function. Since $(\mathbb{R}, |\cdot|, g_*\mu_X)$ is dominated by X, $\text{diam}(\mathbb{R}; 1-\kappa) \le \text{diam}(X; 1 - \kappa)$. Therefore, $\text{ObserDiam}(X; -\kappa) \le \text{diam}(X; 1 - \kappa)$.

And

$$\text{diam}(g_*\mu_X; 1 - \kappa) \le \text{diam}((g \circ f)_*\mu_Y; 1 - \kappa) \le \text{ObserDiam}(Y; -\kappa).$$

$\square$

Proposition 54. *Let X be a metric-measure space. Then for any real number $t > 0$, we have*

$$\text{ObserDiam}(tX; -\kappa) = t\,\text{ObserDiam}(X; -\kappa)$$

where $tX = (X, td_X, \mu_X)$.

Proof.

$$
\begin{aligned}
\text{ObserDiam}(tX; -\kappa) &= \sup\{\text{diam}(f_*\mu_X; 1 - \kappa) \mid f : tX \to \mathbb{R} \text{ is 1-Lipschitz}\} \\
&= \sup\{\text{diam}(f_*\mu_X; 1 - \kappa) \mid t^{-1}f : X \to \mathbb{R} \text{ is 1-Lipschitz}\} \\
&= \sup\{\text{diam}((tg)_*\mu_X; 1 - \kappa) \mid g : X \to \mathbb{R} \text{ is 1-Lipschitz}\} \\
&= t\sup\{\text{diam}(g_*\mu_X; 1 - \kappa) \mid g : X \to \mathbb{R} \text{ is 1-Lipschitz}\} \\
&= t\,\text{ObserDiam}(X; -\kappa).
\end{aligned}
$$

$\square$

2.1.1 Observable diameter for class of spheres

With the basic formalism in place, the next step is to test it on the standard family where concentration is best understood: high-dimensional spheres. This is the model case in which Gaussian limits and scaling arguments can be made explicit, and it provides the comparison space that will later control complex projective space through the Hopf fibration.

In this section, we will try to use the results from previous sections to estimate the observable diameter for the class of spheres.

Theorem 55. *For any real number κ with $0 < \kappa < 1$, we have*

$$\mathrm{ObserDiam}(S^n(1); -\kappa) = O\left(\frac{1}{\sqrt{n}}\right).$$

Proof. First, recall that by the Maxwell-Boltzmann distribution law, for any $n > 0$, if $I(r)$ denotes the measure of the standard Gaussian measure on the interval $[0, r]$, then we have

$$
\begin{aligned}
\lim_{n\to\infty} \mathrm{ObserDiam}(S^n(\sqrt{n}); -\kappa) &= \lim_{n\to\infty} \sup\{\mathrm{diam}((\pi_{n,k})_* \sigma^n; 1 - \kappa) \mid \pi_{n,k} \text{ is 1-Lipschitz}\} \\
&= \lim_{n\to\infty} \sup\{\mathrm{diam}(\gamma^1; 1 - \kappa) \mid \gamma^1 \text{ is the standard Gaussian measure}\} \\
&= \mathrm{diam}(\gamma^1; 1 - \kappa) \\
&= 2I^{-1}\left(\frac{1 - \kappa}{2}\right).
\end{aligned}
$$

By Proposition 54, we have

$$\mathrm{ObserDiam}(S^n(\sqrt{n}); -\kappa) = \sqrt{n}\, \mathrm{ObserDiam}(S^n(1); -\kappa).$$

So

$$\mathrm{ObserDiam}(S^n(1); -\kappa) = \frac{1}{\sqrt{n}} 2I^{-1}\left(\frac{1 - \kappa}{2}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

$\square$

From the previous discussion, we see that the only remaining step in finding the observable diameter of $\mathbb{C}P^n$ is to identify a 1-Lipschitz map with the correct push-forward measure.

To find such a metric, we need some additional results.

Definition 56. *Let M be a smooth manifold. A **Riemannian metric** on M is a smooth covariant tensor field $g \in \mathcal{T}^2(M)$ such that for each $p \in M$, g_p is an inner product on T_pM.*

$g_p(v, v) \geq 0$ *for each $p \in M$ and each $v \in T_pM$. Equality holds if and only if $v = 0$.*

Theorem 57. *Let $(\tilde{M}, \tilde{g})$ be a Riemannian manifold, let $\pi : \tilde{M} \to M$ be a surjective smooth submersion, and let G be a group acting on $\tilde{M}$. If the **action** is*

1. *isometric: the map $x \mapsto \varphi \cdot x$ is an isometry for each $\varphi \in G$.*

2. *vertical: every element $\varphi \in G$ takes each fiber to itself, that is $\pi(\varphi \cdot p) = \pi(p)$ for all $p \in \tilde{M}$.*

3. transitive on fibers: for each $p, q \in \tilde{M}$ such that $\pi(p) = \pi(q)$, there exists $\varphi \in G$ such that $\varphi \cdot p = q$.

Then there is a unique Riemannian metric on M such that π is a Riemannian submersion.

A natural measure for $\mathbb{C}P^n$ is the normalized volume measure on $\mathbb{C}P^n$ induced from the Fubini-Study metric [Lee18, Example 2.30].

Definition 58. *Let n be a positive integer, and consider the complex projective space $\mathbb{C}P^n$ defined as the quotient space of $\mathbb{C}^{n+1}$ by the equivalence relation $z \sim z'$ if there exists $\lambda \in \mathbb{C}$ such that $z = \lambda z'$. The map $\pi : \mathbb{C}^{n+1} \setminus \{0\} \to \mathbb{C}P^n$ sending each point in $\mathbb{C}^{n+1} \setminus \{0\}$ to its span is a surjective smooth submersion.*

Identifying $\mathbb{C}^{n+1}$ with $\mathbb{R}^{2n+2}$ with its Euclidean metric, we can view the unit sphere S^{2n+1} with its round metric $\mathring{g}$ as an embedded Riemannian submanifold of $\mathbb{C}^{n+1} \setminus \{0\}$. Let $p : S^{2n+1} \to \mathbb{C}P^n$ denote the restriction of the map π. Then p is smooth, and it is surjective, because every 1-dimensional complex subspace contains elements of unit norm.

There are many additional properties for such a construction, and here we check only the point needed later.

We need to show that it is a submersion.

Proof. Let $z_0 \in S^{2n+1}$ and set $\zeta_0 = p(z_0) \in \mathbb{C}P^n$. Since π is a smooth submersion, it has a smooth local section $\sigma : U \to \mathbb{C}^{n+1}$ defined on a neighborhood U of ζ_0 and satisfying $\sigma(\zeta_0) = z_0$ by the local section theorem (Theorem 30). Let $v : \mathbb{C}^{n+1} \setminus \{0\} \to S^{2n+1}$ be the radial projection onto the sphere:

$$v(z) = \frac{z}{|z|}.$$

Since dividing an element of $\mathbb{C}^{n+1}$ by a nonzero scalar does not change its span, it follows that $p \circ v = \pi$. Therefore, if we set $\tilde{\sigma} = v \circ \sigma$, then $\tilde{\sigma}$ is a smooth local section of p. Applying the local section theorem again shows that p is a submersion.

Define an action of S^1 on S^{2n+1} by complex multiplication:

$$\lambda(z^1, z^2, \ldots, z^{n+1}) = (\lambda z^1, \lambda z^2, \ldots, \lambda z^{n+1})$$

for $\lambda \in S^1$ (viewed as a complex number of norm 1) and $z = (z^1, z^2, \ldots, z^{n+1}) \in S^{2n+1}$. This action is isometric, vertical, and transitive on fibers of p.

By Theorem 57, there is a unique metric on $\mathbb{C}P^n$ such that the map $p : S^{2n+1} \to \mathbb{C}P^n$ is a Riemannian submersion. This metric is called the Fubini–Study metric. $\square$

2.1.2 Observable diameter for complex projective spaces

The sphere estimate by itself is not yet a statement about quantum state space. The missing geometric bridge is the Hopf fibration, which realizes complex projective space as a quotient of the

sphere by phase. Because this quotient map is a Riemannian submersion, the concentration scale on the sphere can be transferred to $\mathbb{C}P^n$.

Using the projection map and Hopf's fibration, we can estimate the observable diameter for complex projective spaces from the observable diameter of spheres.

Theorem 59. *For any real number κ with $0 < \kappa < 1$, we have*

$$\mathrm{ObserDiam}(\mathbb{C}P^n(1); -\kappa) \leq O\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Recall from Example 2.30 in [Lee18] that the Hopf fibration $f_n : S^{2n+1}(1) \to \mathbb{C}P^n$ is 1-Lipschitz with respect to the Fubini-Study metric on $\mathbb{C}P^n$. The push-forward $(f_n)_* \sigma^{2n+1}$ coincides with the normalized volume measure on $\mathbb{C}P^n$ induced from the Fubini-Study metric.

By Proposition 53, we have

$$\mathrm{ObserDiam}(\mathbb{C}P^n(1); -\kappa) \leq \mathrm{ObserDiam}(S^{2n+1}(1); -\kappa) \leq O\left(\frac{1}{\sqrt{n}}\right).$$

$\square$

2.2 Use entropy function as estimator of observable diameter for complex projective spaces

The previous subsection gives an abstract upper bound. The next question is how to test that scale numerically on observables that arise naturally in quantum information. Entropy is the obvious candidate: it is physically meaningful, it is already controlled by Hayden's theorem, and after normalization by a Lipschitz constant it supplies a computable lower-bound proxy for observable diameter.

In this section we describe a Monte Carlo pipeline for comparing concentration phenomena across three metric-measure spaces using real-valued entropy observables. The goal is not to compute the exact observable diameter

$$\mathrm{ObsDiam}_{\mathbb{R}}(X; -\kappa) = \sup_{f \in \mathrm{Lip}_1(X, \mathbb{R})} \mathrm{diam}(f_* \mu_X; 1 - \kappa),$$

but to estimate it by choosing a specific observable $f : X \to \mathbb{R}$ and then measuring the partial diameter of its push-forward distribution. Thus all numerical quantities below should be interpreted as *entropy-based observable-diameter proxies*, not exact observable diameters in Gromov's sense [Gro81; Shi14].

The screen is $\mathbb{R}$ equipped with the Euclidean metric, and for a fixed $\kappa \in (0, 1)$ we set

$$\alpha = 1 - \kappa.$$

Given sampled values $y_1, \ldots, y_N \in \mathbb{R}$ of the observable, the code sorts them and computes the shortest interval $[a, b]$ containing at least $\lceil \alpha N \rceil$ samples. Its width

$$b - a$$

44

is the empirical partial diameter of the push-forward measure on $\mathbb{R}$.

To compare this width with the true observable diameter, the code also estimates an empirical Lipschitz constant of the chosen observable. If $x_i, x_j \in X$ are sampled states and $f(x_i), f(x_j)$ are the corresponding observable values, then the sampled slopes are

$$\frac{|f(x_i) - f(x_j)|}{d_X(x_i, x_j)},$$

where d_X is the metric of the ambient space. The code records both the maximum sampled slope and the 0.99-quantile of these slopes. Dividing the empirical partial diameter by these sampled Lipschitz constants gives two normalized proxies:

$$\frac{\text{diam}(f_* \mu_X; 1 - \kappa)}{L_{\max}} \quad \text{and} \quad \frac{\text{diam}(f_* \mu_X; 1 - \kappa)}{L_{0.99}}.$$

If the chosen observable were exactly 1-Lipschitz, these normalized quantities would coincide with the raw width. In practice they should be viewed only as heuristic lower-scale corrections.

2.2.1 Random sampling using standard uniform measure on the unit sphere

The first family of spaces is the real unit sphere

$$S^{m-1} = \left\{ x = (x_1, \ldots, x_m) \in \mathbb{R}^m : \|x\|_2 = 1 \right\},$$

equipped with the geodesic distance

$$d_S(x, y) = \arccos\langle x, y \rangle$$

and the normalized Riemannian volume measure. This is the standard metric-measure structure used in concentration of measure on spheres [Lee18; Ver18; Shi14].

Sampling is performed by drawing a standard Gaussian vector $g \in \mathbb{R}^m$ and normalizing:

$$x = \frac{g}{\|g\|_2}.$$

This produces the uniform distribution on S^{m-1}.

The observable is a Shannon entropy built from the squared coordinates:

$$f_{\text{sphere}}(x) = -\sum_{i=1}^{m} x_i^2 \log_2(x_i^2).$$

Since $(x_1^2, \ldots, x_m^2)$ is a probability vector, f_{sphere} takes values in $[0, \log_2 m]$, and the code records $\log_2 m$ as the natural upper bound of the observable.

For each chosen dimension m, the experiment generates N independent samples $x^{(1)}, \ldots, x^{(N)}$, computes the values
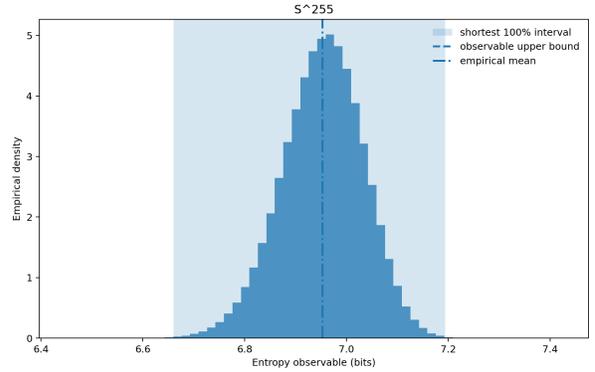
$$f_{\text{sphere}}(x^{(1)}), \ldots, f_{\text{sphere}}(x^{(N)}),$$

and then evaluates the shortest interval containing mass at least $1 - \kappa$. This gives an empirical observable-diameter proxy for the sphere family. The code also computes the empirical mean, median, standard deviation, and the normalized proxies obtained from sampled Lipschitz ratios.

The experiment produces histograms of the observable values, upper-tail deficit plots for $\log_2 m - f_{\text{sphere}}(x)$, and family-wise comparisons of partial diameter, standard deviation, and mean deficit. When available, these plots are overlaid with theoretical concentration scales derived from Lévy's lemma and related results [Lee18].
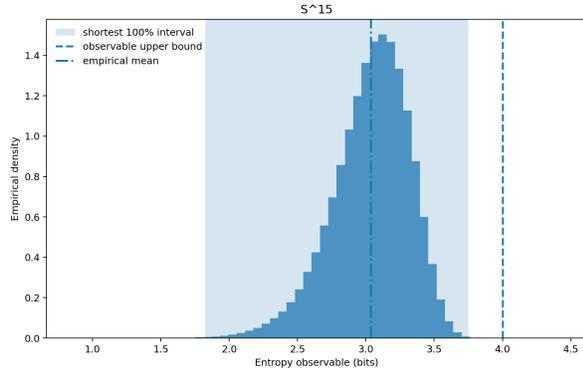
Entropy distribution for S^{15} Entropy distribution for S^{255}

2.2.2 Visualized the concentration of measure phenomenon on complex projective space

The second family is complex projective space

$$\mathbb{C}P^{d_A d_B - 1},$$

viewed as the space of pure states in $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ modulo global phase. Geometrically, this space is equipped with the Fubini–Study metric and its associated normalized volume measure [Lee18; BZ17]. Numerically, a projective point is represented by a unit vector

$$\psi \in \mathbb{C}^{d_A d_B}, \qquad \|\psi\| = 1,$$

and distances are computed by the Fubini–Study formula

$$d_{FS}([\psi], [\phi]) = \arccos |\langle \psi, \phi \rangle|.$$

Sampling is implemented by drawing a complex Gaussian matrix

$$G \in \mathbb{C}^{d_A \times d_B},$$

with independent standard complex normal entries, and then normalizing it so that

$$\psi = \frac{\text{vec}(G)}{\|\text{vec}(G)\|}.$$

This is equivalent to Haar sampling on the unit sphere in $\mathbb{C}^{d_A d_B}$ and hence induces the standard unitarily invariant measure on $\mathbb{C}P^{d_A d_B - 1}$ [BZ17; NC10].

The real-valued observable is the bipartite entanglement entropy. Writing

$$\rho_A = \mathrm{Tr}_B \, |\psi\rangle\langle\psi|,$$

the code defines

$$f_{\mathrm{CP}}([\psi]) = S(\rho_A) = -\mathrm{Tr}(\rho_A \log_2 \rho_A).$$

Equivalently, if $\lambda_1, \ldots, \lambda_{d_A}$ are the eigenvalues of ρ_A, then

$$f_{\mathrm{CP}}([\psi]) = -\sum_{i=1}^{d_A} \lambda_i \log_2 \lambda_i.$$

This observable takes values in $[0, \log_2 d_A]$.

For each dimension pair (d_A, d_B), the experiment samples N independent Haar-random pure states, computes the entropy values, and then forms the empirical push-forward distribution on $\mathbb{R}$. The shortest interval containing mass at least $1 - \kappa$ is reported as the entropy-based observable-diameter proxy. In addition, the code plots histograms, upper-tail deficit plots for

$$\log_2 d_A - S(\rho_A),$$

and family-wise comparisons of partial diameter, standard deviation, and mean deficit. When available, these plots are overlaid with the Page average entropy and with Hayden-style concentration scales, which serve as theoretical guides rather than direct outputs of the simulation [Hay10; HLW06; San95].
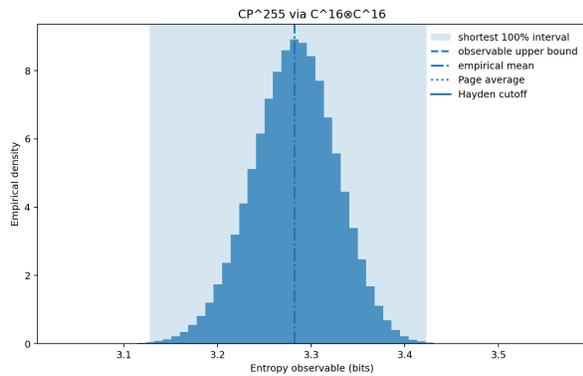
Entropy distribution for $(d_A, d_B) = (16, 16)$
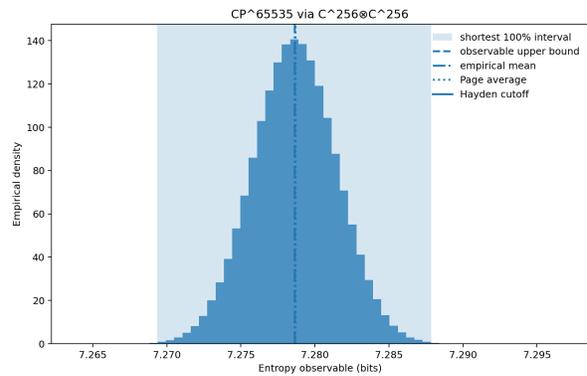
Entropy distribution for $(d_A, d_B) = (256, 256)$

2.3 A conjecture on observable diameter for complex projective spaces

The numerical section does not compute $\mathrm{ObserDiam}(\mathbb{C}P^n(1); -\kappa)$ exactly, but it does produce a natural lower-bound proxy. If $f : \mathbb{C}P^n \to \mathbb{R}$ has Lipschitz constant $L_f > 0$, then $L_f^{-1} f$ is 1-Lipschitz, so

$$\frac{\mathrm{diam}(f_* \mu; 1 - \kappa)}{L_f} \leq \mathrm{ObserDiam}(\mathbb{C}P^n(1); -\kappa).$$

Taking f to be the entropy observable means that the normalized entropy widths from the simulations cannot determine the full observable diameter, but they do give a computable lower bound on its scale. In that sense the entropy function is a probe of projective-space concentration, and the conjecture below asks that the upper bound coming from the Hopf-fibration argument has the same order as the concentration already suggested by entropy.

Theorem 60 (Wu's conjecture). *For $0 < \kappa < 1$,*

$$\mathrm{ObserDiam}(\mathbb{C}P^n(1); -\kappa) = O\left(\frac{1}{\sqrt{n}}\right).$$

Sketch of the proof. The expected upper bound should come from the same geometric mechanism as on the sphere. The Hopf fibration $S^{2n+1}(1) \rightarrow \mathbb{C}P^n(1)$ is 1-Lipschitz and sends the normalized spherical measure to the normalized Fubini–Study volume measure, so domination already gives

$$\mathrm{ObserDiam}(\mathbb{C}P^n(1); -\kappa) \leq \mathrm{ObserDiam}(S^{2n+1}(1); -\kappa).$$

Thus the sphere estimate suggests the order $n^{-1/2}$ on projective space as well. To compare this upper bound with entropy data, one studies the entropy observable $f([\psi]) = S(\mathrm{Tr}_B |\psi\rangle\langle\psi|)$ on the relevant projective state space, normalizes by a Lipschitz bound for f, and uses the resulting widths as lower-bound proxies for observable diameter. The remaining work is to relate these entropy-based lower bounds to intrinsic geometric concentration, possibly through curvature methods such as Fubini–Study geometry and Lévy–Gromov type inequalities.

References

[Axl23]   S. Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer International Publishing, 2023. ISBN: 9783031410260. URL: `https://books.google.com/books?id=OdnfEAAAQBAJ`.

[BZ17]   I. Bengtsson and K. Zyczkowski. *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`.

[Lee18]   J. M. Lee. *Introduction to Riemannian Manifolds*. 2nd. Graduate Texts in Mathematics. Cham, Switzerland: Springer, 2018. ISBN: 978-3-319-91755-9.

[LL12]   J. M. Lee and J. M. Lee. *Introduction to Smooth Manifolds*. eng. New York: Springer, 2012. ISBN: 978-1-4899-9475-2 978-1-4419-9982-5.

[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.