Measure concentration in complex projective space and quantum entanglement

Zheyuan Wu

Washington University in St. Louis

March 25, 2026

Table of Contents

Why I'm here?

"I think I can safely say that nobody understands quantum mechanics."

— Richard Feynman

Non-commutative probability space

We begin our discussion on a general type of probability space.

Non-commutative probability space

A non-commutative probability space is a pair $(\mathscr{B}(\mathscr{H}), \mathscr{P})$, where $\mathscr{B}(\mathscr{H})$ is the set of all **bounded** linear operators on $\mathscr{H}$.
$\mathscr{P}$ is the set of all orthogonal projections on $\mathscr{B}(\mathscr{H})$.
The set $\mathscr{P} = \{P \in \mathscr{B}(\mathscr{H}) : P^* = P = P^2\}$ is the set of all orthogonal projections on $\mathscr{B}(\mathscr{H})$.

Classical probability	**Non-commutative probability**
Sample space Ω, cardinality $\lvert\Omega\rvert = n$, example: $\Omega = \{0, 1\}$	Complex Hilbert space $\mathscr{H}$, dimension $\dim \mathscr{H} = n$, example: $\mathscr{H} = \mathbb{C}^2$
Common algebra of $\mathbb{C}$ valued functions	Algebra of bounded operators $\mathcal{B}(\mathscr{H})$
Events: indicator functions of sets	Projections: space of orthogonal projections $\mathscr{P} \subseteq \mathscr{B}(\mathscr{H})$
functions f such that $f^2 = f = \overline{f}$	orthogonal projections P such that $P^* = P = P^2$
$\mathbb{I}_{f^{-1}(\{\lambda\})}$ is the indicator function of the set $f^{-1}(\{\lambda\})$	$P(\lambda)$ is the orthogonal projection to eigenspace
$f = \sum_{\lambda \in \text{Range}(f)} \lambda \mathbb{I}_{f^{-1}(\{\lambda\})}$	$A = \sum_{\lambda \in \text{sp}(A)} \lambda P(\lambda)$
Probability measure μ on Ω	Density operator ρ on $\mathscr{H}$

Quantum states

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

Definition of (Quantum) State

A state is a unit vector $|\psi\rangle$ in the Hilbert space $\mathscr{H}$, such that $\langle\psi| |\psi\rangle = 1$.
Every state uniquely defines a map $\rho : \mathscr{P} \to [0, 1]$, $\rho(P) = \langle\psi| P |\psi\rangle$ (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 $\rho(P_1 + P_2 + \cdots + P_n) = \sum_{i=1}^{n} \rho(P_i)$.

Here ψ is just a label for the vector. $|\cdot\rangle$ is called the ket (column vector), where the counterpart $\langle\psi|$ is called the bra, used to denote the vector dual to ψ (row vector/linear functional of $|\psi\rangle$).

Quantum measurements

Definition of Quantum 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 $\mathscr{H}$.

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 $\sum_{x\in X} M_x^* M_x = I$ (Law of total probability).

In probability theory, an important measurement of uncertainty is entropy.
It characterizes the information content of a random variable.

Shannon entropy

Given a classical probability vector $p = (p_1, \ldots, p_n)$ with $\sum_i p_i = 1$,

$$H(p) = -\sum_{i=1}^{n} p_i \log_2 p_i.$$

This measures uncertainty of a *chosen measurement outcome*.

Information theory in quantum systems

von Neumann entropy

For a density matrix ρ,

$$S(\rho) = -\operatorname{Tr}(\rho \log_2 \rho).$$

This measures the intrinsic uncertainty of the quantum state and is basis-independent.

Entanglement entropy

For a bipartite pure state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, define the reduced state $\rho_A = \operatorname{Tr}_B\big(|\Psi\rangle\langle\Psi|\big)$. Its entanglement entropy is

$$E(|\Psi\rangle) = H(\rho_A).$$

Thus entanglement entropy is the von Neumann entropy of a subsystem, and it measures how entangled the bipartite pure state is.

Conclusion of Non-commutative probability space

Classical probability	**Non-commutative probability**
Sample space Ω, cardinality $\lvert\Omega\rvert = n$, example: $\Omega = \{0,1\}$	Complex Hilbert space $\mathscr{H}$, dimension $\dim \mathscr{H} = n$, example: $\mathscr{H} = \mathbb{C}^2$
Common algebra of $\mathbb{C}$ valued functions	Algebra of bounded operators $\mathcal{B}(\mathscr{H})$
$f \mapsto \bar{f}$ complex conjugation	$P \mapsto P^*$ adjoint
Events: indicator functions of sets	Projections: space of orthogonal projections $\mathscr{P} \subseteq \mathscr{B}(\mathscr{H})$
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 $\mathscr{H}$
Delta measure δ_ω	Pure state $\rho = \lvert\psi\rangle\langle\psi\rvert$
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 $\mathscr{H}_1 \otimes \mathscr{H}_2$
Product measure $\mu_1 \times \mu_2$ on $\Omega_1 \times \Omega_2$	Tensor product of space $\rho_1 \otimes \rho_2$ on $\mathscr{H}_1 \otimes \mathscr{H}_2$
Marginal distribution $\pi_* v$	Partial trace $\mathrm{Tr}_2(\rho)$

So what?

Lemma: That's all we need.

All quantum operations can be constructed by composing four kinds of transformations:

1. Unitary operations. $U(\cdot)$ for any quantum state. $A^*A = AA^* = I$, A is the matrix of U. (It is possible to apply a non-unitary operation for an open quantum system, but usually leads to non-recoverable loss of information)
2. Extend the system. Given a quantum state $\rho \in \mathcal{H}^N$, we can extend it to a larger quantum system by "entangle" it with some new states $\sigma \in \mathcal{H}^K$ and get $\rho' = \rho \otimes \sigma \in \mathcal{H}^N \otimes \mathcal{H}^K$.
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 new state $\rho' \in \mathcal{H}^{N-K}$.
4. Selective measurement. Given a quantum state, we measure it and get a classical bit.

Quantum states: pure vs. mixed

- 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

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

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

Key distinction

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

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

Relation with the sphere

- 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 induced riemmanian metric: Fubini–Study metric

Definition of Riemannian metric

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 (Vector space formed by the tangent vectors relative to the manifold M at p).
$g_p(v, v) \geq 0$ for each $p \in M$ and each $v \in T_pM$. equality holds if and only if $v = 0$.

- 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**

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

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

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

So what?

With everything we have here, we are ready to answer the question:

How a random bipartite pure state $\mathcal{P}(A \otimes B)$ is distributed on the complex projective space? And how entangled $H(\psi_A)$ it is?

Maxwell-Boltzmann Distribution Law

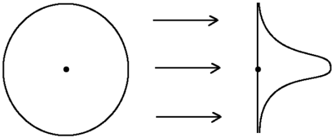


Figure 3.9 The projective central limit theorem: the projection of the uniform distribution on the sphere of radius $\sqrt{n}$ onto a line converges to the normal distribution $N(0,1)$ as $n \to \infty$.

Consider the orthogonal projection $0 \leq k < n$

$$\pi_{n,k} : S^n(\sqrt{n}) \to \mathbb{R}^k.$$

Its push-forward measure converges to the standard Gaussian as dimensions increase $n \to \infty$.

$$(\pi_{n,k})_* \sigma^n \to \gamma^k.$$

Another familiar name when $k = 1$ is the central limit theorem.

Levy Concentration

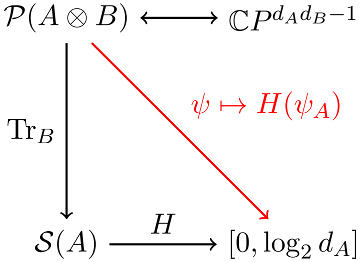
Levy's lemma

If $f : S^n \to \mathbb{R}$ is 1-Lipschitz, then there exists a a_0 such that for $\epsilon > 0$,

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

- In high dimension, most Lipschitz observables are almost constant.
- Here a_0 resembles the "median" of the set $f(S^n)$, that is half of the measure of the observations is bounded below/above by a_0.

How the Entropy Observable Fits In



- Recall that $\mathcal{P}(A \otimes B)$ is the set of pure states on $A \otimes B$. Tr_B is the partial trace over B. $\mathcal{S}(A)$ is the set of mixed states on A. H is the shannon entropy function, $H(\psi_A)$ is the entanglement entropy function.
- The red arrow is the observable to which concentration is applied.

Ingredients Behind the Tail Bound

Page-type lower bound

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

Lipschitz estimate for $H(\psi_A)$

The Lipschitz constant for the function $H(\psi_A)$ should be upper bounded by $\sqrt{8}\log_2(d_A)$, for $d_A \geq 3$.

Levy concentration plus these two estimates produces the exponential entropy tail bound.

Generic Entanglement Theorem

Hayden–Leung–Winter

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

$$\beta = \frac{1}{\ln(2)} \frac{d_A}{d_B}.$$

For $d_B \geq d_A \geq 3$, with $\alpha \geq 0$ by our choice,

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

As $d_B \to \infty$, with overwhelming probability $1 - \exp\left(-\frac{1}{8\pi^2 \ln(2)} \frac{(d_A d_B - 1)\alpha^2}{(\log_2 d_A)^2}\right) = 1 - \Theta(e^{-c d_B})$, a random pure state is almost maximally entangled $\log_2(d_A) - \frac{1}{2\ln(2)} \frac{d_A}{d_B} = \log_2(d_A) - \Theta(\frac{1}{d_B})$.

A natural question from the observables

What does the hayden–leung–winter theorem generalize the behavior of the lipschitz function $S^n \to \mathbb{R}$ and the lipschitz function $\mathbb{C}P^n \to \mathbb{R}$ as $n \to \infty$?

Observable diameter: the inner definition

Partial diameter on $\mathbb{R}$

Let μ be a Borel probability measure on $\mathbb{R}$ and let $\alpha \in (0,1]$. The **partial diameter** of μ at mass level α is

$$\operatorname{diam}(\mu;\alpha) := \inf_{A\subset\mathbb{R}} \{\operatorname{diam}(A) : \mu(A) \geq \alpha\}.$$

where

$$\operatorname{diam}(A) := \sup_{x,y\in A} |x-y|.$$

- The partial diameter asks for: what is the shortest interval I need to capture at least α of the mass (measure)?
- If 1 is the total measure of the space, $\operatorname{diam}(\mu; 1-\kappa)$, measures how tightly we can capture *most* of the distribution, allowing us to discard a set of mass at most κ.

Observable diameter of a metric-measure space

Definition

Let $X = (X, d_X, \mu_X)$ be a metric-measure space and let $\kappa > 0$. The **observable diameter** of X is

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

where $\mathrm{Lip}_1(X, \mathbb{R})$ is the set of all 1-Lipschitz functions $f : X \to \mathbb{R}$, and $f_* \mu_X$ is the pushforward measure on $\mathbb{R}$.

- Each 1-Lipschitz function f is viewed as an **observable** on X.
- The pushforward measure $f_* \mu_X$ is the distribution of the values of that observable.
- If $\mathrm{ObsDiam}_{\mathbb{R}}(X; -\kappa)$ is small, then *every* 1-Lipschitz observable is strongly concentrated.

A Geometric Consequence

Projective-space estimate from Gromov

For $0 < \kappa < 1$,

$$\mathrm{ObsDiam}(\mathbb{C}P^n(1); -\kappa) \le O(\sqrt{n}).$$

- First estimate observable diameter on spheres via Gaussian limits.
- Then use the Hopf map $S^{2n+1}(1) \to \mathbb{C}P^n$.
- This gives a geometric explanation for why many projective-space observables concentrate.

Hayden's work suggests that if the entropy function is a "good" proxy for the observable diameter, the difference of **the order of the growth** between $\mathrm{ObsDiam}(\mathbb{C}P^n(1); -\kappa)$ and $\mathrm{ObsDiam}(S^{2n+1}(1); -\kappa)$ should be smaller than $O(\sqrt{n})$.

A conjecture

Wu's conjecture

For $0 < \kappa < 1$,

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

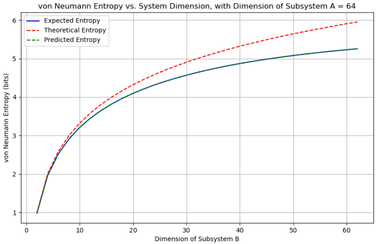
Additional works need to be done to verify this conjecture.

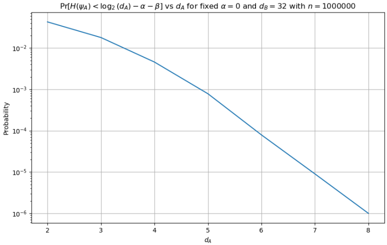
- Entropy function is not globally lipschitz, so we need to bound the deficit of the entropy function.
- Normalize by the Lipschitz constant of f to obtain a weak lower bound for the observable diameter with the algebraic varieties on $\mathbb{C}P^n(1)$.
- Continue to study and interpret the overall concentration mechanism geometrically through the positive Ricci curvature of the Fubini–Study metric and Lévy–Gromov type inequalities.

Entropy-Based Simulations

- Sample Haar-random pure states in $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$.
- Compute reduced density matrices and entanglement entropy.
- Measure shortest intervals containing mass $1 - \kappa$ in the entropy distribution.
- Compare concentration across:
  - real spheres,
  - complex projective spaces

What the Data Suggests



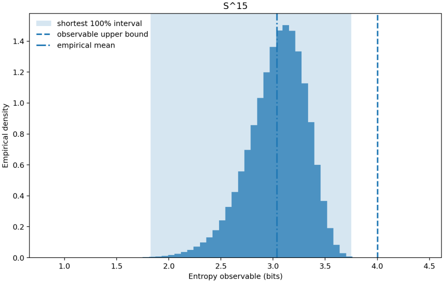


Entropy vs. ambient dimension

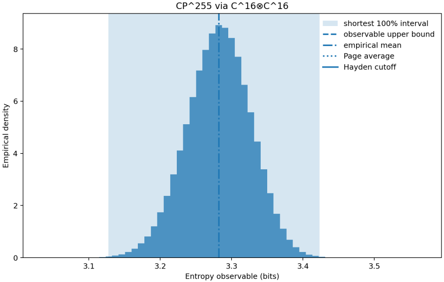
Entropy vs. subsystem dimension

As dimension increases, the entropy distribution concentrates near the maximal value.

Results for concentration of random states in lower dimensional spaces

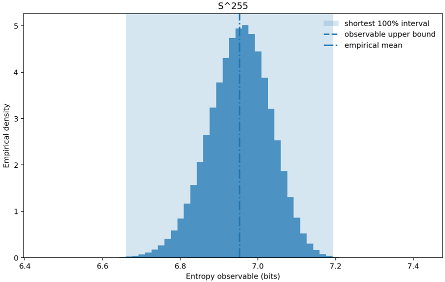


Entropy distribution for S^{15}

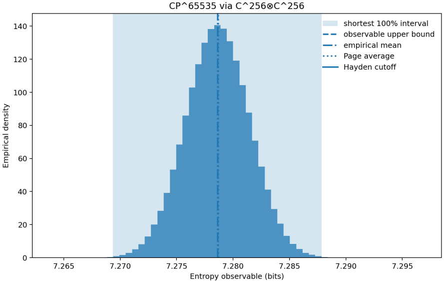


Entropy distribution for $\mathbb{C}P^{16} \otimes \mathbb{C}P^{16}$

Results for concentration of random states in higher dimensional spaces



Entropy distribution for S^{255}



Entropy distribution for $\mathbb{C}P^{256} \otimes \mathbb{C}P^{256}$

Conclusion and Outlook

- Complex projective space provides the natural geometric setting for pure quantum states.
- Concentration of measure explains generic high entanglement in large bipartite systems.
- Observable diameter gives a way to phrase concentration geometrically.
- Ongoing directions:
  - sharper estimates for $\mathbb{C}P^n$
  - deeper use of Fubini–Study geometry
  - recursive learning on new theorems and mathematical tools

Q&A