Overview of quantum learning theory

Junseo Lee

QISCA Summer School 2025
Quantum Learning and Complexity Theory – **Lecture 1**

July 12, 2025

Course Details

- Introductory-level overview of **quantum learning theory** and **quantum complexity theory**
- Covers fundamental tools for analyzing the **sample complexity** of quantum learning tasks
  - Includes techniques involving the *Haar measure*

BQP QMA

- Core algorithms for learning:
  - Quantum states, unitaries, processes, and circuits
- Introduction to key **quantum complexity classes**

P. NP

- Theoretical CS background for foundational papers in quantum learning
- Includes a glimpse of **recent research trends**
- Topics may be adjusted based on participant interest

Prerequisites

- Participants should already be familiar with:
  - **Linear algebra**, **probability theory**, and **basic quantum information theory**
  - Density matrix formalism, POVMs, Pauli matrices
  - Bloch sphere representation
  - Basic quantum channels (e.g., depolarizing, dephasing)
  - Unitary and Hermitian matrices, spectral decomposition
  - Tensor product notation and partial trace
  - Entanglement measures (e.g., von Neumann entropy, mutual information)
- These topics will not be reviewed during the course. (Only a brief review below.)
- Students lacking this background may find the material difficult to follow.

Density Matrix Formalism

Review: **Quantum Computation and Quantum Information**, *Chapters 1–2*

By tracing out subsystem A of a pure state $|\psi\rangle$, we obtain a mixed state ρ:

$$|\psi\rangle = \sum_i \sqrt{p_i}\, |\psi_i\rangle_A\, |i\rangle_B \quad \Rightarrow \quad \rho = \mathrm{Tr}_A(|\psi\rangle\langle\psi|) = \sum_i p_i\, |\psi_i\rangle\langle\psi_i| \tag{1}$$

Bloch Sphere Representation

Review: **Quantum Computation and Quantum Information**, *Chapters 1–2*

Any 1-qubit pure state $|\psi\rangle$ can be written as:

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle , \quad |\alpha|^2 + |\beta|^2 = 1 \tag{2}$$

It can equivalently be represented on the Bloch sphere as:

$$|\psi\rangle = e^{i\gamma} \left(\cos\frac{\theta}{2} |0\rangle + e^{i\phi} \sin\frac{\theta}{2} |1\rangle \right) \tag{3}$$

POVM

Review: **Quantum Computation and Quantum Information**, *Chapters 1–2*

A measurement is described by a set of operators $\{M_m\}$. Given a state ρ, the probability of outcome m is:

$$p_m = \mathrm{Tr}(M_m^\dagger M_m \rho) \tag{4}$$

The post-measurement state becomes:

$$\rho \to \frac{M_m \rho M_m^\dagger}{\mathrm{Tr}(M_m^\dagger M_m \rho)} \tag{5}$$

(Pauli) Observables

Review: **Quantum Computation and Quantum Information**, *Chapters 1–2*

Define $P_n = \{I, X, Y, Z\}^{\otimes n}$ to be the set of n-qubit Pauli strings.

Then any observable O on n qubits can be expressed as a linear combination of Pauli operators:

$$O = \sum_{\sigma_x \in P_n} \boxed{\alpha_x} \sigma_x \tag{6}$$

Unitary Transformation

Review: **Quantum Computation and Quantum Information**, *Chapters 1–2*

The evolution of a closed quantum system is described by a unitary transformation. That is, the state ρ evolves under a unitary U as:

$$\rho' = U\rho U^\dagger \tag{7}$$

Review: **Quantum Computation and Quantum Information**

The inner product $|\langle\psi|\phi\rangle|^2$ between two pure states $|\psi\rangle$, $|\phi\rangle$ can be estimated via the **swap test** with ϵ-additive error using $O(1/\epsilon^2)$ repetitions.

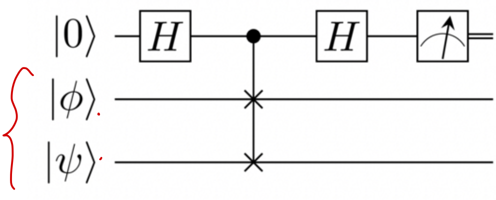


Figure: Swap test circuit

→ mixed.

Survey Overview

Note: This course approaches the subject from a mathematical and theoretical computer science perspective.

Student Type

- Undergraduate students (80%)
- Master's students (4%)
- Integrated MS–PhD program students (12%)
- Others (4%)

Major

- Physics (40%)
- EECS (40%)
- Other engineering (12%)
- Chemistry (8%)

What This Course Does *Not* Cover

- This course focuses on **provable learning models**, especially in the context of quantum state learning.
- Core topics include:
  - Sample complexity analysis
  - Related mathematical foundations
- We do **not** cover heuristic or variational quantum machine learning models such as:
  - Quantum Neural Networks (QNN), Quantum Convolutional Neural Networks (QCNN), Quantum GANs (QGAN)
  - Tasks like training classifiers on datasets such as MNIST
- However, we will briefly discuss the use of **quantum neural estimation** in the context of **quantum property testing**. (Lec 4)

Course Policies

- **No lecture recordings** will be provided.
- Lectures will primarily be held online via Zoom.
  - Access details will be shared later.
  - Lecture 4 or 8 may optionally be held **in person**, depending on circumstances, and may include a Q&A session.
- All lectures are scheduled for **Saturdays at 3 PM**, **except Lecture 2**, which will be held on **Sunday**.
- Problem sets may be assigned during the course.
- Any changes to the lecture schedule will be announced in advance.

Tentative Schedule and Topics

- **Jul 12** – Overview of quantum learning theory
- **Jul 20** – Haar measures and classical shadows (1)
- **Jul 26** – Haar measures and classical shadows (2)
- **Aug 02** – Quantum property neural estimation
- **Aug 09** – Complexity of learning quantum states
- **Aug 16** – Quantum complexity and homology problems
- **Aug 23** – Learning quantum circuits and unitaries
- **Aug 30** – Open problems and discussion session

Note: Lecture topics are subject to minor changes.

Instructors

Junseo Lee
Lectures: Weeks 1, 5, 6, 8
Research interests:

- Quantum complexity theory
- Quantum learning theory
- Quantum algorithms

Myeongjin Shin
Lectures: Weeks 2, 3, 4, 7
Research interests:

- Quantum property estimation
- Quantum learning theory
- Quantum algorithms

What is quantum learning?

state, . .

"How can we learn about quantum objects and phenomena?"

- Arguably one of the most basic scientific questions:
  - What is this object?
  - Is it equal to some other object?
  - What are its statistics?

$\rho \overset{?}{\approx} \sigma$

$\mathrm{tr}(\rho^2)$.

Professor A sets up an experiment in his lab that produces one copy of a quantum state ρ at a time.

Question:
How many copies of ρ are required to learn useful information about it?

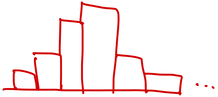
poly(n)

$O(2^n)$.

Why Do We Care About Quantum Learning Problems?

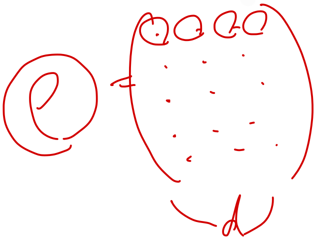
- Physical phenomena are fundamentally quantum.
- It is useful for verifying the outcome of quantum computations.
- It is the natural non-commutative analogue of distribution learning and testing.
- Quantum advantage?

- **Classical distribution:** A probability distribution over d elements is specified by a vector $p \in \mathbb{R}^d$ on the **simplex**, i.e.,
  - $p_x \geq 0$ for all $x = 1, 2, \ldots, d$
  - $\sum_{x=1}^{d} p_x = 1$

$p =$ 

- **Quantum distribution:** A mixed quantum state over d dimensions is specified by a matrix $\rho \in \mathbb{C}^{d \times d}$ on the **spectrahedron**, i.e.,
  - ρ is Hermitian: $\rho = \rho^\dagger$
  - $\rho \succeq 0$ (positive semidefinite)
  - $\mathrm{Tr}(\rho) = 1$

$O(d^2)$

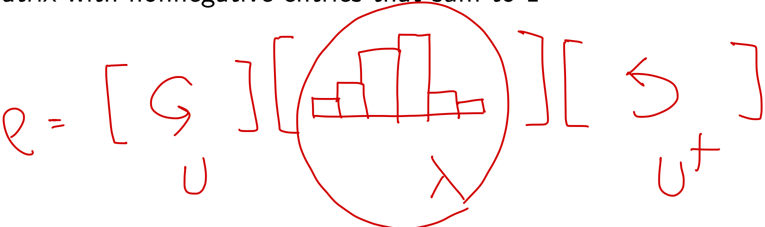


By the spectral theorem, any quantum state ρ can be written as

$$\rho = U \Lambda U^\dagger$$

where:

- U is a unitary matrix (represents a change of basis or rotation)
- Λ is a diagonal matrix with nonnegative entries that sum to 1



Key Difference 1: Measurement

One does not simply sample from a quantum state! To interact with a quantum state, one must specify a **measurement** $\mathcal{M}$.

A measurement is a collection of $d \times d$ matrices $\mathcal{M} = \{M_1, M_2, \ldots, M_\ell\}$ such that:

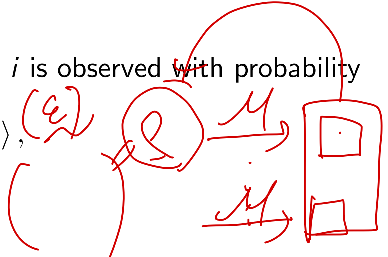
- $M_i \succeq 0$ for all $i = 1, \ldots, \ell$ (each M_i is positive semidefinite)
- $\sum_{i=1}^{\ell} M_i = I$ (completeness relation)

Such a measurement is called a **Positive Operator-Valued Measure (POVM)**.

In this talk, we often assume each M_i is rank-one, i.e., $M_i = w_i \, |\xi_i\rangle \langle\xi_i|$ for some $\|\xi_i\| = 1$ and $w_i \geq 0$.

When measuring a state ρ with this POVM, the outcome i is observed with probability

$$p_i = \mathrm{Tr}(\rho M_i) = w_i \, \langle\xi_i| \, \rho \, |\xi_i\rangle \,,$$



and the state ρ collapses after measurement.

TV $\|\ \|_1$ $\|\ \|_F$ $\|\ \|_2$

Given n samples from an unknown distribution p, infer some property of p:

- **Uniformity testing:** test whether p is uniform or ε-far from uniform
- **Identity testing:** test whether $p = q$ for a known distribution q or ε-far from it
- ✓ **Distribution learning:** learn p to total variation distance error ε

succ. prob $1-\delta$

All guarantees are required to hold with high probability.

fail prob δ

$\geq .99$

$\log(1/\delta)$

Quantum Testing and Learning

$\rho \overset{?}{=} I/d$

Given n copies of an unknown quantum state ρ, infer some property of ρ:

- **Mixedness testing:** test whether ρ is the maximally mixed state or ε-far from it
- **State certification:** test whether $\rho = \sigma$ for a known state σ or ε-far from it
- (Lect 5) **Quantum state tomography:** learn ρ to trace distance error ε $\quad \|\rho - \tilde{\rho}\|_{tr} \leq \varepsilon.$

All guarantees are required to hold with high probability.

$\rho = \begin{pmatrix} \quad \end{pmatrix} \in \mathbb{C}^{d \times d}$

Key Difference 2: Entanglement

Formally, N copies of a quantum state ρ correspond to a single copy of the larger mixed state $\rho^{\otimes N}$.

We can apply a single global measurement to $\rho^{\otimes N}$ — such a measurement is called an entangled measurement.

While the state $\rho^{\otimes N}$ itself is a **product state** and not entangled, it turns out that entangled measurements allow us to extract information **more efficiently** than product (separable) measurements.

To fully explain this phenomenon, we will need tools from **representation theory** (covered in Lectures 2 and 3).

Schur–Weyl Duality

Suppose we are interested in estimating some property of ρ's spectrum.

Then, any optimal measurement on $\rho^{\otimes N}$ should satisfy the following invariance conditions:

1. It should be invariant under permuting the copies of ρ:

$$\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_N \;\Rightarrow\; \rho_{\pi(1)} \otimes \rho_{\pi(2)} \otimes \cdots \otimes \rho_{\pi(N)}$$

2. It should be invariant under applying a shared unitary to all copies:

$$\rho^{\otimes N} \;\Rightarrow\; \left(U^\dagger \rho U\right)^{\otimes N}$$

Schur-Weyl duality (cont.)

Both S_N and U_d have a natural action on $(\mathbb{C}^d)^{\otimes n}$. Let

$$\mathbf{P}(\cdot), \mathbf{Q}(\cdot) : (\mathbb{C}^d)^{\otimes n} \to (\mathbb{C}^d)^{\otimes n}$$

be the associated representations:

$$\mathbf{P}(\pi) x_1 \otimes x_2 \otimes \cdots \otimes x_N = x_{\pi^{-1}(1)} \otimes x_{\pi^{-1}(2)} \otimes \cdots \otimes x_{\pi^{-1}(N)}$$

$$\mathbf{Q}(U) x_1 \otimes x_2 \otimes \cdots \otimes x_N = (U x_1) \otimes (U x_2) \otimes \cdots \otimes (U x_N)$$

Both representations have direct product structure ("irreps") which are naturally identified with partitions of N, and since they commute, they have nice joint structure. This can be formalized by the famous **Schur-Weyl duality**.

Schur-Weyl duality (cont.)

Schur–Weyl Duality:

$$\mathbf{PQ} \cong \bigoplus_{\substack{\lambda \vdash N \\ \ell(\lambda) \leq d}} p_\lambda \otimes q_\lambda^d$$

1. λ is a partition of N into at most d parts, i.e., $\lambda = (\lambda_1, \ldots, \lambda_k)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$, $\sum_i \lambda_i = N$, and $k \leq d$.
2. p_λ is the associated irrep of S_N.
3. q_λ^d is the associated irrep of U_d; moreover, q_λ^d is a matrix polynomial.

Corollary: There exists a fixed unitary U_{Schur} on $(\mathbb{C}^d)^{\otimes N}$ such that for any $\rho \in \mathbb{C}^{d \times d}$,

$$U_{\mathsf{Schur}}^\dagger \rho^{\otimes N} U_{\mathsf{Schur}} = \bigoplus_{\substack{\lambda \vdash N \\ \ell(\lambda) \leq d}} \mathrm{Id}_{\mathsf{dim}(\lambda)} \otimes q_\lambda^d(\rho) \tag{8}$$

Weak Schur Sampling

This unitary transformation gives us a generic way to turn $\rho^{\otimes N}$ into a block diagonal matrix with nice structure.

This also motivates **weak Schur sampling**:

1. Rotate $\rho^{\otimes N}$ with the Schur transformation.
2. Let Π_λ denote the projection onto the (now coordinate) subspace indexed by λ.
3. Measure with the (projective) measurement

$$\{\Pi_\lambda\}_{\substack{\lambda \vdash N \\ \ell(\lambda) \leq d}} .$$

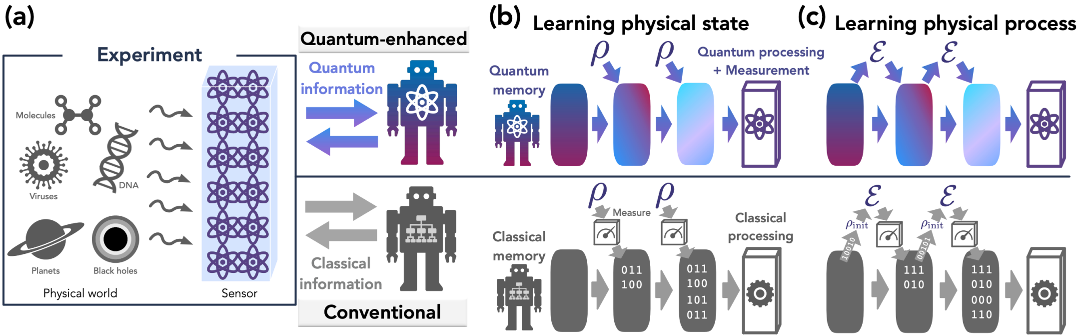
Weak Schur Sampling

Theorem [folklore, see e.g. Wright'16]: Suppose we are interested in measuring a property of ρ which depends only on its spectrum. Then weak Schur sampling is the optimal measurement on $\rho^{\otimes N}$.

Note that since U_{Schur} is a rotation in $(\mathbb{C}^d)^{\otimes N}$, this is a necessarily entangled measurement!

The cost of entanglement

Hsin-Yuan Huang *et al.*, Quantum advantage in learning from experiments. *Science* **376**, 1182-1186 (2022).



The power of quantum memory

For learning d-dimensional states:

$d = 2^n$ fully $2^3 \sim 2^4$ t t-copy $d^{2.5}$

Learning Problem	With entanglement	Without entanglement
State tomography	$\Theta(d^2)$ [HHJW'17, OW'17]	$\Theta(d^3)$ [CHLLS'22]
State certification	$\Theta(d)$ [OW'16]	$\Theta(d^2)$ [CHLL'22]
Shadow tomography with m observables	$\mathcal{O}(\log^2 m \log d)$ [BO'20]	$\tilde{\Theta}(\min(m, d))$ [HKP'20, CCHL'21]
Shadow tomography of Pauli observables	$\mathcal{O}(\log d)$ [HKP'20]	$\tilde{\Theta}(\sqrt{d})$ [HKP'20, CCHL'21]
Purity testing	$\Theta(1)$ [folklore]	$\Theta(\sqrt{d})$ [CCHL'21]

By the way. . .

Doesn't this suggest we *can't* perform quantum learning with our current quantum memory capabilities?

Idea: Our lower bounds yield tasks that provably and unconditionally require quantum memory to be solved efficiently.

Existing quantum computers *do* have some limited amounts of quantum memory.

If we can solve some of these tasks on real quantum computers,
then the computer must be using quantum memory
$\Rightarrow$ the computer must exhibit fundamentally quantum behavior.

Conclusion: Our techniques can be used to demonstrate quantum advantage.

ρ

Shadow tomography [Aaronson'18]:
Let $\mathcal{O} = \{O_j\}_{j=1}^{m} \subseteq \mathbb{C}^{d \times d}$ be a set of m matrices ("observables").

Assume that $\underline{\|O_j\|_{\mathrm{op}} \leq 1}$ for all j.

Given N copies of ρ, estimate $\boxed{\mathrm{tr}(O_j \rho)}$ to error ε for all $j = 1, \ldots, m$.

Upper and lower bounds

[Bădescu–O'Donnell '21]: There is an estimator using entangled measurements that succeeds with high probability when

success. prob $1-\delta$

$$N = \mathcal{O}\left(\frac{\log^2(m)\log d}{\varepsilon^4}\right) \cdot \log(1/\delta)$$

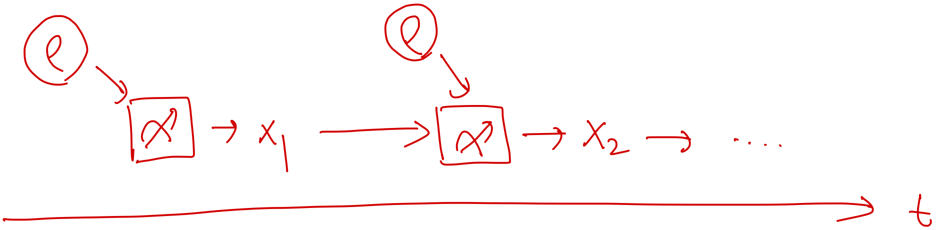
Theorem [CCHL'21]: There is a collection of m observables such that any estimator using unentangled measurements that succeeds with probability ≥ 0.51 requires

$$N = \widetilde{\Omega}\left(\frac{\min(m, d)}{\varepsilon^2}\right).$$

Any estimator for this problem without quantum memory is exponentially worse!

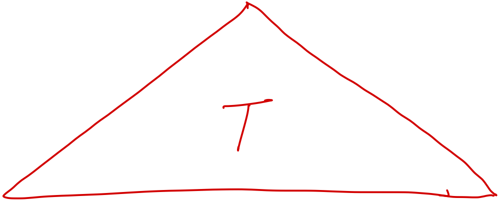
Unentangled but adaptive algorithms

The key technical challenge in proving bounds for unentangled algorithms is proving lower bounds against adaptive algorithms.
Such a lower bound is essential for proving quantum advantage.



The tree representation

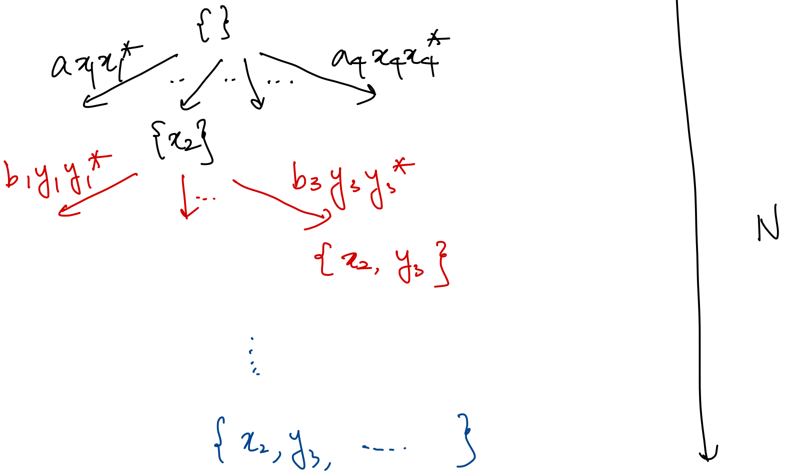
Inspired by the literature on bandit lower bounds, [BCL'20] introduce the tree representation of unentangled learners.



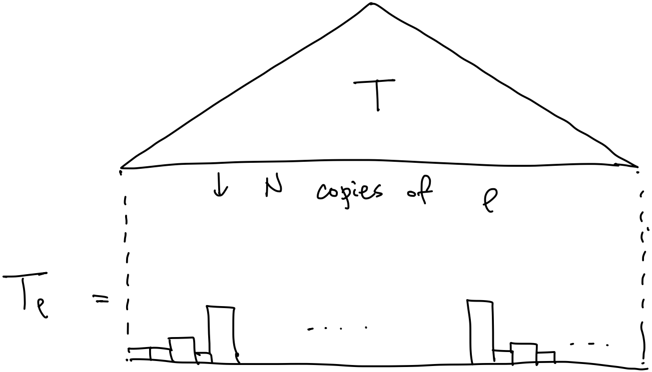
$\longrightarrow$ estimation

From trees to distributions

Consider some estimator that uses N copies.



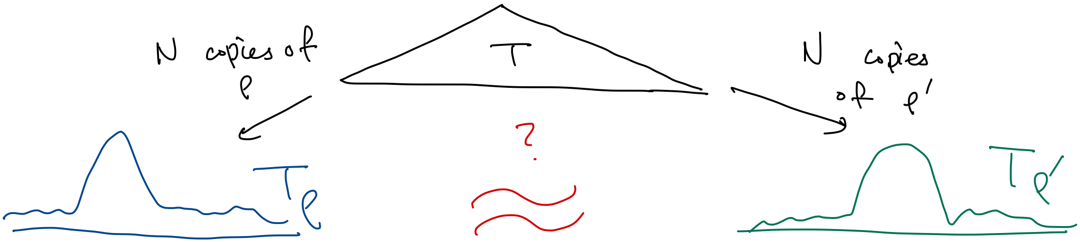
From trees to distributions



for any leaf node $u = \{x_1 \dots x_N\}$:
$$T_\ell(u) = \prod_{i=1}^{N} w_i x_i^* \ell x_i .$$

An estimator gets exactly 1 leaf node from the tree.

To show that an estimator T cannot distinguish two states ρ and ρ', it is equivalent to show that T_ρ & $T_{\rho'}$ are statistically indistinguishable.



The tree distribution

Note: to get our lower bounds, it doesn't suffice to just compare two states.

$\downarrow$

Can we distinguish between T_ρ and $\mathbb{E}_{\rho'}(T_{\rho'})$?

$\rightarrow$ Quantum Le Cam's method. (Lec 5).

Preview: Types of Approaches

Huang Bresler

Here are some advanced techniques that we will study later (around Lecture 5). Each has been influential in recent theoretical work.

- **Jensen's inequality:** something like learning tree formalism [HKP'21, CCHL'21, CCHL'22] $\to$ Uses edge-based analysis of likelihood ratios. Often results in loose bounds.
- **Chain Rule :** Upper bound the total variation (TV) distance by the conditional χ^2-divergence. [BCL'20, CLO'22]
  - Generalizes a technique used in classical online learning lower bounds.
  - Originally developed by Bubeck and Cesa-Bianchi (2012).
- **Anti-concentration:** Directly demonstrate anti-concentration of the posterior distribution. [CHLL'22, CHLLS'23]

C: Sitan Chen. O: O'Donnell

L: Jerry Li

Assignment

Read arXiv:2307.08956, Chapters 1 and 2.

2 ~ 3

Introduction to Haar Measure Tools in Quantum Information: A Beginner's Tutorial

Antonio Anna Mele

Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany

Optional exercise: study about quantum Le Cam's method & try to prove $N = \tilde{\Omega}(\min(n, d)/\varepsilon^2)$ for shadow tomography.

Thanks a lot!