Lecture 10: Noncommutative averages CSE 599I, Spring 2021
Instructor: James R. Lee

1 The quantum data processing inequality

1.1 Quantum channels

A classical stochastic channel is a mapping $\Phi : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Y})$ that sends probability measures over a set $\mathcal{X}$ to probability measures over a set $\mathcal{Y}$. In the discrete setting, such a channel is specified by a mapping of every $x \in \mathcal{X}$ to a probability measure $\nu_x \in \mathcal{P}(\mathcal{Y})$, and then $\Phi(\mu)(y) = \sum_{x\in\mathcal{X}} \nu_x(y)$. In terms of samples, every element $x \in \mathcal{X}$ is mapped to a random element $\boldsymbol{Y}(x)$ of $\mathcal{Y}$. So if $\boldsymbol{X}$ is a random variable taking values in $\mathcal{X}$, then after passing through the channel, the resulting random variable is $\boldsymbol{Y}(\boldsymbol{X})$.

Note that a classical channel can be viewed as a linear map $\Phi : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{Y}}$ that preserves probability measures, and this can be decomposed into two properties:

1. (Positivity). If $p \in \mathbb{R}_+^{\mathcal{X}}$, then $\Phi(p) \in \mathbb{R}_+^{\mathcal{Y}}$.
2. (Trace preservation). $\sum_{i=1}^k \Phi(p)_i = \sum_{i=1}^n p_i$.

Let us now describe the analogous notion of a quantum channel. As in the classical case, one can look for "operational" descriptions, or structural ones. Consider a linear map $\mathcal{E} : \mathbb{M}_n(\mathbb{C}) \to \mathbb{M}_k(\mathbb{C})$; we might impose additional properties as in the classical setting:

1. $\mathcal{E}$ is a *positive map* if it maps positive matrices to positive matrices.
2. $\mathcal{E}$ is *trace-preserving* if $\mathrm{Tr}(\mathcal{E}(A)) = \mathrm{Tr}(A)$ for $A \in \mathbb{M}_n(\mathbb{C})$.

There is an additional property of classical stochastic channels that holds for free: If $\Phi : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{Y}}$ is a stochastic channel, then we can extend Φ to a stochastic channel $\tilde{\Phi} : \mathbb{R}^{\mathcal{X}} \otimes \mathbb{R}^{\mathcal{X}'} \to \mathbb{R}^{\mathcal{Y}} \otimes \mathcal{R}^{\mathcal{X}'}$ that only acts non-trivially on the first coordinate, i.e., $\tilde{\Phi} = \Phi \otimes I_{\mathcal{X}'}$. Indeed, we often take for granted that operating stochastically on some subsystem can also be envisioned as a stochastic map on the whole system.

Remark 1.1 (Failure of channel extension). This property can fail in the quantum setting: The tranpose map $\mathcal{E}(A) = A^T$ is positive and trace-preserving, but $\mathcal{E} \otimes I_2$ fails to be positive. Indeed, operating on a 2×2 block matrix, we have

$$(\mathcal{E} \otimes I_2)\left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

It is straightforward to verify that the first matrix is PSD, while the second has eigenvalue -1 corresponding to the eigenvector $(0, -1, 1, 0)$.

A linear map $\mathcal{E} : \mathbb{M}_n(\mathbb{C}) \to \mathbb{M}_k(\mathbb{C})$ is called ℓ-positive if $\mathcal{E} \otimes I_\ell : \mathbb{M}_{n\ell}(\mathbb{C}) \to \mathbb{M}_{k\ell}(\mathbb{C})$ is positive. If $\mathcal{E}$ is ℓ-positive for every $\ell \geqslant 0$, one says that $\mathcal{E}$ is *completely positive (CP)*. A completely positive trace-preserving map (*CPT map,* for short) is our definition of a quantum channel.

Theorem 1.2 (Choi, Kraus). *A map $\mathcal{E} : \mathbb{M}_n(\mathbb{C}) \to \mathbb{M}_k(\mathbb{C})$ is completely positive if and only if*

$$\mathcal{E}(\rho) = V_1 \rho V_1^* + \cdots + V_m \rho V_m^*$$

for some linear operators $V_1, \ldots, V_m : \mathbb{C}^n \to \mathbb{C}^k$. A map is CPT if additionally $V_1^* V_1 + \cdots + V_m^* V_m = I$.

The Stinespring Dilation Theorem can be used to give a "physical" interpretation of CPT maps.

Theorem 1.3. *Let $\mathcal{E} : \mathbb{M}_n(\mathbb{C}) \to \mathbb{M}_k(\mathbb{C})$ be a CPT map. Then there is an auxiliary Hilbert space $\mathbb{C}^m$, a density $\sigma \in \mathcal{D}(\mathbb{C}^m)$, a unitary U, and a decomposition $\mathbb{C}^n \otimes \mathbb{C}^m = \mathcal{H}_A \otimes \mathcal{H}_B$ such that*

$$\mathcal{E}(\rho) = \mathrm{Tr}_B \left(U(\rho \otimes \sigma) U^* \right).$$

The physical interpretation is analogous to the discussion of state transformations in Lecture 8. One can think of $\rho \otimes \sigma$ as a coupling of ρ with some larger environment, the map $\rho \otimes \sigma \mapsto U(\rho \otimes \sigma) U^*$ as a unitary evolution, and the partial trace $\mathrm{Tr}_B(\cdot)$ as observing only part of the resulting system.

1.2 A data-processing inequality

The classical data processing inequality asserts that stochastic channels can only reduce the relative entropy:

$$\mathsf{D}(\Phi(p) \,\|\, \Phi(q)) \leqslant \mathsf{D}(p \,\|\, q).$$

This is also true in the quantum setting.

Theorem 1.4 (Quantum DPI). *For any quantum channel $\mathcal{E} : \mathbb{M}_n(\mathbb{C}) \to \mathbb{M}_k(\mathbb{C})$ and densities $\rho, \sigma \in \mathcal{D}(\mathbb{C}^n)$, it holds that*

$$\mathsf{S}(\mathcal{E}(\rho) \,\|\, \mathcal{E}(\sigma)) \leqslant \mathsf{S}(\rho \,\|\, \sigma).$$

You will prove this in HW #2 using the following special case.

Theorem 1.5. *If $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is are bipartite states, then*

$$\mathsf{S}(\mathrm{Tr}_A(\rho) \,\|\, \mathrm{Tr}_A(\sigma)) \leqslant \mathsf{S}(\rho \,\|\, \sigma).$$

Proof. In HW #1, you showed that if $X \in \mathbb{M}_k(\mathbb{C})$ is a matrix, and $D_X \in \mathbb{M}_k(\mathbb{C})$ is its diagonal, then we can write

$$D_X = \frac{1}{k} \sum_{j=0}^{k-1} U^j X U^{*j},$$

where U is unitary.

Suppose D is a diagonal matrix. We claim that

$$\frac{\mathrm{Tr}(D)}{k} I = \frac{1}{k} \sum_{j=0}^{k-1} V^j D V^{*j}$$

for some permutation matrix V. Indeed, V can simply correspond to a cyclic permutation of the diagonal so that by averaging over all k shifts, every entry of the diagonal is equal to the average diagonal entry. Summarizing: There are unitaries $U_1, U_2, \ldots, U_r$ such that for any $X \in \mathbb{M}_k(\mathbb{C})$,

$$\frac{\mathrm{Tr}(X)}{k} I = \frac{1}{r} \sum_{j=1}^{r} U_j X U_j^*.$$

A similar construction works for block matrices: Assume that $\mathcal{H}_A = \mathbb{C}^n$ and $\mathcal{H}_B = \mathbb{C}^k$, and think of $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ as an $n \times n$ block matrix where each block is a $k \times k$ matrix. Let M_ρ denote the block matrix where every block contains $\frac{\mathrm{Tr}_A(\rho)}{k} I_k$. Then there are unitaries $V_1, V_2, \dots, V_r$ such that for any $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$,

$$M_\rho = \frac{1}{r}\sum_{j=1}^{r} V_j \rho V_j^*. \tag{1.1}$$

In fact, we can define V_j to be the $n \times n$ block matrix with blocks U_j on the diagonal. One can think of M_ρ in (1.1) as a sort of "operator conditional expectation," and the following argument as an operator variant of Jensen's inequality applied to the jointly convex function $(\rho, \sigma) \mapsto \mathsf{S}(\rho \,\|\, \sigma)$.

Do this now to both ρ and σ, yielding

$$\begin{aligned}
\mathsf{S}(\mathrm{Tr}_A(\rho) \,\|\, \mathrm{Tr}_A(\sigma)) &= \mathsf{S}(M_\rho \,\|\, M_\sigma) \\
&= \mathsf{S}\left(\frac{1}{r}\sum_{j=1}^{r} V_j \rho V_j^* \,\middle\|\, \frac{1}{r}\sum_{j=1}^{r} V_j \sigma V_j^*\right) \\
&\leqslant \frac{1}{r}\sum_{j=1}^{r} \mathsf{S}(V_j \rho V_j^* \,\|\, V_j \sigma V_j^*) \\
&= \mathcal{S}(\rho \,\|\, \sigma),
\end{aligned}$$

where the inequality uses joint convexity of the relative entropy, and the last line uses unitary invariance of the relative entropy: For all unitaries U and density matrices A, B,

$$\begin{aligned}
\mathsf{S}(UAU^* \,\|\, UBU^*) &= \mathrm{Tr}(UAU^*(\log(UAU^*) - \log(UBU^*)) \\
&= \mathrm{Tr}(UA[(\log A)U^* - (\log B)U^*]) \\
&= \mathsf{S}(A \,\|\, B).
\end{aligned}$$

In the preceding equality, we used that $U^*U = I$, that $\log(UAU^*) = U \log(A) U^*$ holds for any unitary U, and that the trace is unitarily invariant. ◻

It is instructive to see that joint convexity of the quantum relative entropy is also a consequence of Theorem 1.5. Indeed, consider $\rho_1, \rho_2, \sigma_1, \sigma_2 \in \mathcal{D}(\mathbb{C}^n)$ and define

$$\rho := \begin{bmatrix} t\rho_1 & 0 \\ 0 & (1-t)\rho_2 \end{bmatrix}, \quad \sigma := \begin{bmatrix} t\sigma_1 & 0 \\ 0 & (1-t)\sigma_2 \end{bmatrix},$$

for some $t \in [0, 1]$. Then,

$$\mathsf{S}(\rho \,\|\, \sigma) = t\mathsf{S}(\rho_1 \,\|\, \sigma_1) + (1-t)\mathsf{S}(\rho_2 \,\|\, \sigma_2),$$

and

$$\mathsf{S}(t\rho_1 + (1-t)\rho_2 \,\|\, t\sigma_1 + (1-t)\sigma_2)$$

is precisely the relative entropy of the states after taking partial trace of the 2×2 block matrices, hence Theorem 1.5 implies that $(\rho, \sigma) \mapsto \mathsf{S}(\rho \,\|\, \sigma)$ is jointly convex.

2 The operator Jensen inequality

If matrices $A_1, A_2, \dots, A_m$ satisfy $A_1^* A_1 + \cdots + A_m^* A_m = I$, we have already said that the map $X \mapsto A_1^* X A_1 + \cdots + A_m^* X A_m$ is like a "noncommutative averaging operation." We can extend this to the notion of a noncommutative convex combination of matrices $X_1, X_2, \dots, X_m$:

$$A_1^* X_1 A_1 + \cdots + A_m^* X_m A_m.$$

It is remarkable that operator convexity of a function f generalizes to the stronger notion of operator convexity with respect to noncommutative convex combinations, as the next theorem asserts. In the next theorem, a rectangular matrix V is called an *isometry* if its columns are orthogonal, i.e., if $V^* V = I$. Note that a unitary matrix is precisely an isometry that is also a square matrix.

Theorem 2.1 (Hansen-Pedersen)**.** *Suppose $f : J \to \mathbb{R}$ is continuous on some interval $J \subseteq \mathbb{R}$. Then the following are equivalent:*

(i) f is operator convex.

(ii) For any square matrices $A_1, \dots, A_m$ with

$$A_1^* A_1 + \cdots + A_m^* A_m = I, \tag{2.1}$$

and Hermitian matrices $X_1, \dots, X_m$ (whose spectrum lies in J),

$$f(A_1^* X_1 A_1 + \cdots + A_m^* X_m A_m) \preceq A_1^* f(X_1) A_1 + \cdots + A_m^* f(X_m) A_m.$$

(iii) $f(V^* X V) \preceq V^* f(X) V$ for every isometry V and Hermitian X with $\operatorname{spec}(X) \subseteq J$*.*

Proof. Let us prove that (i) $\Rightarrow$ (ii). We leave the easier implications (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) as an exercise. Consider $A_1, \dots, A_m \in \mathbb{M}_n(\mathbb{C})$ satisfying (2.1). Note that

$$V := \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} : \mathbb{C}^n \to \mathbb{C}^{mn}$$

is an isometry since $V^* V = A_1^* A_1 + \cdots + A_m^* A_m = I_n$. In particular, we can decompose $\mathbb{C}^{mn} = K \oplus K'$, where $K = V(\mathbb{C}^n)$ is the range of V, and $K' \cong \mathbb{C}^{n(m-1)}$. Let $\{u_1, \dots, u_{n(m-1)}\}$ be an orthonormal basis for K', and define the matrix

$$U := \begin{bmatrix} | & | & & | & A_1 \\ | & | & & | & A_2 \\ u_1 & \cdots & & u_{n(m-1)} & \vdots \\ | & | & & | & A_{m-1} \\ | & | & & | & A_m \end{bmatrix} \in \mathbb{M}_{mn}(\mathbb{C}),$$

where the first $n(m-1)$ columns contain the vectors $u_1, \dots, u_{n(m-1)}$ (as column vectors), and the last n columns contain V. Then $U \in \mathbb{M}_{mn}(\mathbb{C})$ is an isometry, hence a unitary matrix, and if we think of U as an $m \times m$ block matrix, then $U_{km} = A_k$ for each $k = 1, 2, \dots, m$

Define $X \in \mathbb{M}_{mn}(\mathbb{C})$ as the block diagonal matrix with Hermitian matrices $X_1, X_2, \ldots, X_m$ on the diagonal. Let $\omega_m := \exp(2\pi \mathrm{i}/m)$ denote a primitive mth root of unity, and define the block-diagonal matrix $E \in \mathbb{M}_{mn}(\mathbb{C})$ by

$$E := \begin{bmatrix} \omega_m I_n & 0 & 0 & \cdots & 0 \\ 0 & \omega_m^2 I_n & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & 0 \\ 0 & \cdots & 0 & \omega_m^{m-1} I_n & 0 \\ 0 & 0 & \cdots & 0 & I_n \end{bmatrix}$$

Note that, as in HW #1, for any $Y \in \mathbb{M}_{mn}(\mathbb{C})$, we have

$$D_Y = \frac{1}{m} \sum_{j=1}^{m} E^{-j} Y E^{j}, \tag{2.2}$$

where $D_Y \in \mathbb{M}_{mn}(\mathbb{C})$ is the block-diagonal matrix with $(D_Y)_{jj} = Y_{jj}$ and $(D_Y)_{ij} = 0$ otherwise.

Now write

$$\begin{aligned} f\left(\sum_{j=1}^{m} A_j^* X_j A_j\right) = f((U^*XU)_{mm}) &\overset{(2.2)}{=} f\left(\left(\sum_{j=1}^{m} \frac{1}{m} E^{-j} U^* X U E^{j}\right)_{mm}\right) \\ &= \left(f\left(\sum_{j=1}^{m} \frac{1}{m} E^{-j} U^* X U E^{j}\right)\right)_{mm} \\ &\leq \left(\frac{1}{m} \sum_{j=1}^{m} f(E^{-k} U^* X U E^{k})\right)_{mm} = \left(U^* f(X) U\right)_{mm} = \sum_{j=1}^{m} A_j^* f(X_j) A_j. \quad \square \end{aligned}$$

Note that the third equality uses the fact that $\sum_{j=1}^{m} \frac{1}{m} E^{-j} U^* X U E^{j}$ is a block diagonal matrix. If M is a block-diagonal matrix with $M_1, M_2, \ldots, M_m$ on the diagonal, then $f(M)$ is the block-diagonal matrix with $f(M_1), f(M_2), \ldots, f(M_m)$ on the diagonal, hence $f(M_{mm}) = f(M)_{mm}$.

Example 2.2 (Operator convexity of the square)**.** If $A, B \geq 0$ and $S^*S + T^*T = I$, then operator convexity of the square gives

$$(S^*AS + T^*BT)^2 \leq S^*A^2S + T^*B^2T.$$

Apparently there is no simpler proof of this matrix inequality.