8.S372/18.S996 Quantum Information Science III **Fall 2020**

Lecture 10: October 1, 2020

Lecturer: Aram Harrow *Scribe: Matthew Kearney, John Martyn*

10.1 Non-Commutative Union Bound

To begin, recall the statement of the non-commutative union bound:

$$\omega \geq 0, \quad \mathrm{tr}(\omega) \leq 1, \quad P_1, ..., P_L = \text{set of projectors} \quad \Rightarrow$$
$$\mathrm{tr}(\omega) - \mathrm{tr}(P_L...P_1 \omega P_1...P_L) \leq 2\sqrt{\sum_i \mathrm{tr}(\hat{P}_i \omega)}, \quad \hat{P}_i = I - P_i.$$

We will now prove this bound. We will first examine the case where ω is a pure state written as

$$\omega = |\psi\rangle \langle\psi| , \quad \| \, |\psi\rangle \, \| \leq 1$$

We would like to show that

$$\| \, |\psi\rangle \, \|^2 - \|P_L...P_1 \, |\psi\rangle \, \|^2 \leq 2\sqrt{\sum_i \|\hat{P}_i \, |\psi\rangle \, \|^2}.$$

We note that since P_L and $\hat{P_L}$ are orthogonal operators that sum to I we can write $|\psi\rangle$ as

$$|\psi\rangle = P_L \, |\psi\rangle + \hat{P_L} \, |\psi\rangle \, .$$

We will now use a proof by induction on L with the inductive assumption that

$$\| \, |\psi\rangle - P_{L-1}...P_1 \, |\psi\rangle \, \|^2 \leq \sum_{i=1}^{L-1} \|\hat{P}_i \, |\psi\rangle \, \|^2.$$

To begin we have

$$|\psi\rangle - P_L...P_1 \, |\psi\rangle = \hat{P}_L \, |\psi\rangle + P_L(|\psi\rangle - P_{L-1}...P_1 \, |\psi\rangle).$$

Since P_L and $\hat{P}_L$ are orthogonal operators we can then use the Pythagorean theorem

$$\| \left|\psi\right\rangle - P_L ... P_1 \left|\psi\right\rangle \|^2 = \|\hat{P}_L \left|\psi\right\rangle \|^2 + \|P_L(\left|\psi\right\rangle - P_{L-1} ... P_1 \left|\psi\right\rangle)\|^2.$$

Since projection operators do not increase the norm we have

$$\| \left|\psi\right\rangle - P_L ... P_1 \left|\psi\right\rangle \|^2 \leq \|\hat{P}_L \left|\psi\right\rangle \|^2 + \| \left|\psi\right\rangle - P_{L-1} ... P_1 \left|\psi\right\rangle \|^2.$$

We can then use our inductive assumption to get

$$\| \left|\psi\right\rangle - P_L ... P_1 \left|\psi\right\rangle \|^2 \leq \sum_{i=1}^{L} \|\hat{P}_i \left|\psi\right\rangle \|^2 \quad \Rightarrow$$
$$\| \left|\psi\right\rangle - P_L ... P_1 \left|\psi\right\rangle \| \leq \sqrt{\sum_{i=1}^{L} \|\hat{P}_i \left|\psi\right\rangle \|^2}.$$

On the other hand, by the triangle inequality we get

$$\| \left|\psi\right\rangle \| - \|P_L ... P_1 \left|\psi\right\rangle \| \leq \sqrt{\sum_{i=1}^{L} \|\hat{P}_i \left|\psi\right\rangle \|^2}.$$

Let $A = \| \left|\psi\right\rangle \|$ and $B = \|P_L ... P_1 \left|\psi\right\rangle \|$. Since $A, B \leq 1$ we have

$$A^2 - B^2 = (A - B)(A + B) \leq 2(A - B) \leq 2\sqrt{\sum_{i=1}^{L} \|\hat{P}_i \left|\psi\right\rangle \|^2}.$$

This proves the statement of the the theorem for the pure state case. We now discuss the case of mixed states. We note that the left hand side of the inequality is linear in omega. Therefore if $\omega = \sum_i p_i \psi_i$, this gives us

$$\mathrm{tr}(\omega) - \mathrm{tr}(P_L ... P_1 \omega P_1 ... P_L) = \sum_i p_i(\mathrm{tr}(\psi_i) - \mathrm{tr}(P_L ... P_1 \psi_i P_1 ... P_L)).$$

However, the right hand side of the inequality consists of the square root of a linear function of ω. This means the right hand side is concave in ω. This gives us the following property

$$2\sqrt{\sum_i \mathrm{tr}(\hat{P}_i \omega)} = 2\sqrt{\sum_i \sum_j p_j \mathrm{tr}(\hat{P}_i \psi_j)} \geq 2 \sum_j p_j \sqrt{\sum_i \mathrm{tr}(\hat{P}_i \psi_j)}.$$

Putting these facts together we have

$$\begin{aligned}\mathrm{tr}(\omega) - \mathrm{tr}(P_L...P_1\omega P_1...P_L) &= \sum_i p_i(\mathrm{tr}(\psi_i) - \mathrm{tr}(P_L...P_1\psi_i P_1...P_L)) \\ &\leq 2\sum_j p_j \sqrt{\sum_i \mathrm{tr}(\hat{P}_i\psi_j)} \\ &\leq 2\sqrt{\sum_i \mathrm{tr}(\hat{P}_i\omega)}.\end{aligned}$$

This proves the statement of the theorem for mixed states.

10.2 Proving HSW Theorem with Non-Commutative Union Bound

Recall that the failure probability of our sequential decoding scheme is given by

$$p_{\mathrm{err}}(m) = 1 - p_{\mathrm{success}} = 1 - \mathrm{tr}\Big(\Pi_{c_m}\hat{\Pi}_{c_{m-1}}...\hat{\Pi}_{c_1}\Pi\sigma_{c_m}\Pi\hat{\Pi}_{c_1}\hat{\Pi}_{c_{m-1}}\Pi_{c_m}\Big).$$

Remember from the hypothesis of the packing lemma that

$$\mathrm{tr}\Pi\sigma_{c_m}\Pi \geq 1 - \epsilon.$$

Putting these together we have

$$p_{\mathrm{err}}(m) \leq \epsilon + \mathrm{tr}\Pi\sigma_{c_m}\Pi - \mathrm{tr}(\Pi_{c_m}\hat{\Pi}_{c_{m-1}}...\hat{\Pi}_{c_1}\Pi\sigma_{c_m}\Pi\hat{\Pi}_{c_1}\hat{\Pi}_{c_{m-1}}\Pi_{c_m}).$$

Using the non-commutative union bound to $\Pi\sigma_{c_m}\Pi$, we get

$$p_{\mathrm{err}}(m) \leq \epsilon + 2\sqrt{\mathrm{tr}((\hat{\Pi}_{c_m} + \Pi_{c_{m-1}} + ... + \Pi_{c_1})\Pi\sigma_{c_m}\Pi)}$$

Taking the expectation of this quantity over the message and codebook, we establish

$$\mathbb{E}_C\mathbb{E}_m p_{\mathrm{err}}(m|C) \leq \epsilon + 2\mathbb{E}_{m,C}\sqrt{\mathrm{tr}((\hat{\Pi}_{c_m} + \Pi_{c_{m-1}} + ... + \Pi_{c_1})\Pi\sigma_{c_m}\Pi)}.$$

Once again, by the concavity of the square root (i.e. applying Jensen's inequality), we have that

$$\mathbb{E}_C\mathbb{E}_m p_{\text{err}}(m|C) \leq \epsilon + 2\sqrt{\mathbb{E}_{m,C}\text{tr}((\hat{\Pi}_{c_m} + \Pi_{c_{m-1}} + ... + \Pi_{c_1})\Pi\sigma_{c_m}\Pi)}$$

We already showed in section 9.2 that

$$\mathbb{E}_{m,C}\text{tr}(\hat{\Pi}_{c_m}\Pi\sigma_{c_m}\Pi) \leq \epsilon + 2\sqrt{\epsilon}$$
$$\mathbb{E}_{m,C}\sum_{m\neq m'}\text{tr}(\Pi_{c'_m}\Pi\sigma_{c_m}\Pi) \leq \frac{Md}{D}$$

Therefore this gives us

$$\mathbb{E}_{m,C}p_{\text{err}}(m|C) \leq \epsilon + 2\sqrt{\epsilon + 2\sqrt{\epsilon} + Md/D}.$$

Hypothesis Testing

We would like to distinguish $\rho^{\otimes n}$ from $\sigma^{\otimes n}$. Specifically, we want a measurement M such that

$$\text{tr}(\rho^{\otimes n}M) \geq \alpha, \quad \alpha \in (0,1)$$
$$\text{tr}(\sigma^{\otimes n}M) \sim 2^{-nR}$$

We will prove Stein's Lemma, which states the optimal $R = D(\rho\|\sigma) = \text{tr}(\rho(\log\rho - \log\sigma))$. The optimal M is the projector onto $[\alpha^{-1}\rho^{\otimes n} - 2^{nR}\sigma^{\otimes n} \geq 0]$ which is the projector onto the non-negative eigenspace of the given quantity.

We have shown as an exercise that, classically, the best M to distinguish distributions p^n and q^n is given by M being a projector onto $T^n_{p,\delta}$. Specifically,

$$p^n(T^n_{p,\delta}) \to 1 \text{ as } n \to \infty$$
$$q^n(T^n_{p,\delta}) \approx |T^n_{p,\delta}|q(1)^{np(1)}...q(d)^{np(d)} \approx 2^{-nD(p\|q)}$$

so we can distinguish the two stat fairly well, depending on the magnitude of $D(p\|q)$.

We will now explore the quantum version following the proof of Bjelakovic et al. Define ρ and σ as

$$\rho = \sum_x r_x |\alpha_x\rangle\langle\alpha_x| \quad \sigma = \sum_x s_x |\beta_x\rangle\langle\beta_x|$$

We define a new type of typical projector as

$$\Pi^n_{\rho\|\sigma,\delta} = \sum_{x^n : |\frac{1}{n}\sum_{i=1}^{n} \log s_{x_i} - \mathrm{tr}(\rho \log \sigma)| \leq \delta} \beta_{x^n}$$

$$\beta_{x^n} = \beta_{x_1} \otimes ... \otimes \beta_{x_n}$$

We note the following properties of this projector

$$\mathrm{tr}(\rho^{\otimes n} \Pi^n_{\rho\|\sigma,\delta}) \geq 1 - \epsilon \tag{10.1}$$

$$[\Pi^n_{\rho\|\sigma,\delta}, \sigma^{\otimes n}] = 0 \tag{10.2}$$

$$2^{n\mathrm{tr}(\rho\log\sigma - \delta)} \Pi^n_{\rho\|\sigma,\delta} \leq \Pi^n_{\rho\|\sigma,\delta} \sigma^{\otimes n} \Pi^n_{\rho\|\sigma,\delta} \leq 2^{n\mathrm{tr}(\rho\log\sigma + \delta)} \Pi^n_{\rho\|\sigma,\delta} \tag{10.3}$$

Achievability

We will first show that Stein's Lemma is achievable with $M = \Pi^n_{\rho\|\sigma,\delta}\Pi^n_{\rho,\delta}\Pi^n_{\rho\|\sigma,\delta}$. With this definition, we have

$$\begin{aligned} \mathrm{tr}(\rho^{\otimes n}\Pi^n_{\rho,\delta} - \rho^{\otimes n} M) &= \mathrm{tr}(\Pi^n_{\rho,\delta}(\rho^{\otimes n} - \Pi^n_{\rho,\delta}\Pi^n_{\rho\|\sigma,\delta}\rho^{\otimes n}\Pi^n_{\rho\|\sigma,\delta}) \\ &\leq \|\rho^{\otimes n} - \Pi^n_{\rho\|\sigma,\delta}\rho^{\otimes n}\Pi^n_{\rho\|\sigma,\delta}\|_1 \quad \Rightarrow \\ \mathrm{tr}(M\rho^{\otimes n}) &\geq \mathrm{tr}(\rho^{\otimes n}\Pi^n_{\rho,\delta}) - \|\rho^{\otimes n} - \Pi^n_{\rho\|\sigma,\delta}\rho^{\otimes n}\Pi^n_{\rho\|\sigma,\delta}\|_1. \end{aligned}$$

By the gentle measurement lemma we have that

$$\mathrm{tr}(M\rho^{\otimes n}) \geq 1 - \epsilon - 2\sqrt{\epsilon} \geq \alpha.$$

Now we look at how M acts on $\sigma^{\otimes n}$:

$$\mathrm{tr}(M\sigma^{\otimes n}) = \mathrm{tr}(\Pi^n_{\rho,\delta}\Pi^n_{\rho\|\sigma,\delta}\sigma^{\otimes n}\Pi^n_{\rho\|\sigma,\delta}).$$

Using equation 10.3 this gives us

$$\mathrm{tr}(M\sigma^{\otimes n}) \leq \mathrm{tr}(\Pi^n_{\rho,\delta})2^{n\mathrm{tr}(\rho\log\sigma+\delta)} \leq 2^{n(S(\rho)+\delta\mathrm{tr}(\rho\log\sigma)+\delta)} = 2^{-n(D(\rho\|\sigma)-2\delta)},$$

and so we have proven achievability.

Converse

Suppose $\mathrm{tr}(M\rho^{\otimes n}) \geq \alpha$. We will argue that $\mathrm{tr}(M\sigma^{\otimes n})$ is not too small. From 10.2 and 10.3 we have

$$\sigma^{\otimes n} \geq \Pi^n_{\rho\|\sigma,\delta} 2^{n\mathrm{tr}(\rho\log\sigma-\delta)}$$
$$\mathrm{tr}(M\sigma^{\otimes n}) \geq \mathrm{tr}(M\Pi^n_{\rho\|\sigma,\delta})2^{n\mathrm{tr}(\rho\log\sigma-\delta)}.$$

To bound this, we will now show a bound for $\mathrm{tr}(M\Pi^n_{\rho\|\sigma,\delta})$. We note the following

$$\rho^{\otimes n}\Pi^n_{\rho,\delta} = \Pi^n_{\rho,\delta}\rho^{\otimes n}\Pi^n_{\rho,\delta} \leq 2^{(-n(s(\rho)-\delta))}\Pi^n_{\rho,\delta} \tag{10.4}$$

We will compute $\mathrm{tr}(M\Pi^n_{\rho\|\sigma,\delta})$.

$$\begin{aligned}\mathrm{tr}(M\Pi^n_{\rho\|\sigma,\delta}) &= \mathrm{tr}(\Pi^n_{\rho\|\sigma,\delta}M\Pi^n_{\rho\|\sigma,\delta}) \\ &\geq \mathrm{tr}(\Pi^n_{\rho\|\sigma,\delta}M\Pi^n_{\rho\|\sigma,\delta}\Pi^n_{\rho,\delta})\end{aligned}$$

Using equation 10.4 we have

$$\mathrm{tr}(M\Pi^n_{\rho\|\sigma,\delta}) \geq \mathrm{tr}(\Pi^n_{\rho\|\sigma,\delta}M\Pi^n_{\rho\|\sigma,\delta}\Pi^n_{\rho,\delta}\rho^{\otimes n})2^{n(s(\rho)-\delta)}.$$

Let B be the atypical part of $\rho^{\otimes n}$ ($\rho = A + B =$ typical + atypical).

$$\mathrm{tr}(M\Pi^n_{\rho\|\sigma,\delta}) \geq \mathrm{tr}(\Pi^n_{\rho\|\sigma,\delta}M\Pi^n_{\rho\|\sigma,\delta}(\rho^{\otimes n} - B))2^{n(s(\rho)-\delta)}.$$

Once again by gentle measurement we have

$$\mathrm{tr}(M\Pi^n_{\rho\|\sigma,\delta}) \geq (\alpha - 2\sqrt{\epsilon} - \epsilon)2^{n(S(\rho)-\delta))}.$$

This finally brings us to our conclusion that

$$\mathrm{tr}(M\sigma^{\otimes n}) \geq (\alpha - 2\sqrt{\epsilon} - \epsilon)2^{-n(D(\rho\|\sigma)+2\delta))},$$

and the proof of the converse is complete.

Corollary: Monotonicity of $D(\rho\|\sigma)$ under Partial Trace

Given ρ_{AB},σ_{AB} there exists an M such that $\operatorname{tr}(M\rho_A^{\otimes n}) \geq \alpha$ and $\operatorname{tr}(M\sigma_A^{\otimes n}) \approx 2^{-nD(\rho_A\|\sigma_A)}$. This means

$$\operatorname{tr}((M \otimes I_B)^{\otimes n}\rho_{AB}^{\otimes n}) = \operatorname{tr}(M\rho_A^{\otimes n}) \geq \alpha$$
$$\operatorname{tr}((M \otimes I_B)^{\otimes n}\sigma_{AB}^{\otimes n}) = \operatorname{tr}(M\sigma_A^{\otimes n}) \geq 2^{-nD(\rho_A\|\sigma_A)}.$$

Therefore

$$2^{-nD(\rho_A\|\sigma_A)} \geq 2^{-nD(\rho_{AB}\|\sigma_{AB})} \quad \Rightarrow$$
$$D(\rho_{AB}\|\sigma_{AB}) \geq D(\rho_A\|\sigma_A).$$

Evidently, $D(\|)$ is decreases under partial trace.

Corollary: Strong Subadditivity

We can express the conditional mutual information as

$$I(A:C|B) = I(A:BC) - I(A:B) = D(\rho_{ABC}\|\rho_B \otimes \rho_{BC}) - D(\rho_{AB}\|\rho_A \otimes \rho_B).$$

If we let $\sigma_{ABC} = \rho_B \otimes \rho_{BC}$, then the second divergence is simply the first but with both systems traced over C. Thus, the monotonicity of $D(\rho\|\sigma)$ under partial trace gives us that

$$I(A:C|B) = I(A:BC) - I(A:B) \geq 0.$$

This is just strong subadditivity.

Aside: Converse of Schumacher Compression

Recall equation 10.4

$$A = \rho^{\otimes n}\Pi_{\rho,\delta}^n = \Pi_{\rho,\delta}^n\rho^{\otimes n}\Pi_{\rho,\delta}^n \leq 2^{-n(s(\rho)-\delta)}\Pi_{\rho,\delta}^n$$

Let $\rho^{\otimes n} = A + B$ with $\operatorname{tr}(B) \leq \epsilon$. Then we have

$$\alpha \leq \operatorname{tr}(M\rho^{\otimes n}) = tr(MA) + \operatorname{tr}(BM) \leq \operatorname{tr}(AM) + \epsilon.$$

This gives us

$$\alpha - \epsilon \leq \operatorname{tr}(AM) \leq \operatorname{tr}(M\Pi^n_{\rho,\delta}) \exp(-n(S(\rho) - \delta)).$$

So finally we have that

$$\operatorname{tr}(M) \geq \operatorname{tr}(M\Pi^n_{\rho,\delta}) \geq (\alpha - \epsilon) \exp(n(S(\rho) - \delta)),$$

which is the converse of Schumacher compression.