Lecture 4

Regularization of the smoothed max-relative entropy

In this lecture we will prove an important theorem concerning the smoothed max-relative entropy, which is that by regularizing the smoothed max-relative entropy we obtain the ordinary quantum relative entropy:

$$\lim_{n\to\infty} \frac{\mathrm{D}^{\varepsilon}_{\max}\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right)}{n} = \mathrm{D}(\rho\|\sigma) \tag{4.1}$$

for all density operators $\rho, \sigma \in \mathrm{D}(\mathcal{X})$ and all $\varepsilon \in (0,1)$. For the sake of clarity, recall that we define the smoothed max-relative entropy with respect to *trace-distance smoothing*:

$$\mathrm{D}^{\varepsilon}_{\max}(\rho\|\sigma) = \inf_{\xi\in\mathcal{B}_{\varepsilon}(\rho)} \mathrm{D}_{\max}(\xi\|\sigma) \tag{4.2}$$

where

$$\mathcal{B}_{\varepsilon}(\rho) = \left\{\xi \in \mathrm{D}(\mathcal{X}) \,:\, \tfrac{1}{2}\|\rho - \xi\|_1 \leq \varepsilon\right\}. \tag{4.3}$$

Bibliographic remarks

Lemma 4.4, which in some sense is the engine that drives the proof we will discuss, is due to Bjelaković and Siegmund-Schultze (arXiv:quant-ph/0307170), who used it to prove the so-called *quantum Stein lemma*, and through it obtained an alternative proof of the monotonicity of quantum relative entropy.

The more direct route from Bjelaković and Siegmund-Schultze's lemma to the regularization (4.1) to be followed in this lecture appears in the following as-of-yet unpublished manuscript:

Shitikanth Kashyap, Ashwin Nayak, and Michael Saks. Asymptotic equipartition for quantum relative entropy revisited. Manuscript, 2014.

4.1 Strong typicality

The general notion of *typicality* is fundamentally important in information theory, and there is a sense in which it goes hand-in-hand with the concept of entropy. We will begin the lecture with a brief and directed summary of *strong typicality*, which is a particular formulation of typicality that is convenient for the proof.

First let us introduce some notation. Supposing that Σ is an alphabet, for every string $a_1 \cdots a_n \in \Sigma^n$ and symbol $a \in \Sigma$, we write

$$N(a \,|\, a_1 \cdots a_n) = \big|\{k \in \{1, \ldots, n\} \,:\, a_k = a\}\big|, \tag{4.4}$$

which is simply the number of times the symbol a occurs in the string $a_1 \cdots a_n$. With respect to that notation, strong typicality is defined as follows.

Definition 4.1. Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\delta > 0$ be a positive real number. A string $a_1 \cdots a_n \in \Sigma^n$ is *δ-strongly typical with respect to p* if

$$\left| \frac{N(a \,|\, a_1 \cdots a_n)}{n} - p(a) \right| \leq p(a)\delta \tag{4.5}$$

for every $a \in \Sigma$. The set of all δ-strongly typical strings of length n with respect to p is denoted $S_{n,\delta}(p)$.

What the definition expresses is that the proportion of each symbol in a strongly typical string is approximately what one would expect if the individual symbols were chosen independently at random according to the probability vector p. Notice that because it is the quantity $p(a)\delta$, as opposed to δ, that appears on the right-hand side of the inequality in the definition, we have that the error tolerance for the frequency with which each symbol appears shrinks proportionately as the probability for that symbol to appear shrinks—and if $p(a) = 0$ for some $a \in \Sigma$, then a strongly typical string cannot include the symbol a at all.

Next we will prove two basic facts concerning the notion of strong typicality. The two facts are stated as the lemmas that follow.

Lemma 4.2. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\delta > 0$ be a positive real number. It is the case that*

$$\sum_{a_1 \cdots a_n \in S_{n,\delta}(p)} p(a_1) \cdots p(a_n) \geq 1 - 2 \sum_{\substack{a \in \Sigma \\ p(a) > 0}} \exp\big(-2n\delta^2 p(a)^2\big). \tag{4.6}$$

Proof. Suppose first that $a \in \Sigma$ is fixed, and consider the probability that a string $a_1 \cdots a_n \in \Sigma^n$, where each symbol is selected independently at random according to the probability vector p, satisfies

$$\left| \frac{N(a \,|\, a_1 \cdots a_n)}{n} - p(a) \right| > p(a)\delta. \tag{4.7}$$

To upper-bound this probability, one may define $X_1, \ldots, X_n$ to be independent and identically distributed random variables, taking value 1 with probability $p(a)$ and value 0 otherwise, so that the probability of the event (4.7) is equal to

$$\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - p(a) \right| > p(a)\delta \right). \tag{4.8}$$

If it is the case that $p(a) > 0$, then Hoeffding's inequality implies that

$$\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - p(a) \right| > p(a)\delta \right) \leq 2\exp\bigl(-2n\delta^2 p(a)^2\bigr), \tag{4.9}$$

while it is the case that

$$\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - p(a) \right| > p(a)\delta \right) = 0 \tag{4.10}$$

in case $p(a) = 0$. The lemma follows from the union bound. □

Lemma 4.3. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, let $\delta > 0$ be a positive real number, let $a_1 \cdots a_n \in S_{n,\delta}(p)$ be a δ-strongly typical string with respect to p, and let $\phi : \Sigma \rightarrow [0, \infty)$ be a nonnegative real-valued function. The following inequality is satisfied:*

$$\left| \frac{\phi(a_1) + \cdots + \phi(a_n)}{n} - \sum_{a \in \Sigma} p(a)\phi(a) \right| \leq \delta \sum_{a \in \Sigma} p(a)\phi(a). \tag{4.11}$$

Proof. The inequality (4.11) follows from the definition of strong typicality together with the triangle inequality:

$$\begin{aligned}
&\left| \frac{\phi(a_1) + \cdots + \phi(a_n)}{n} - \sum_{a \in \Sigma} p(a)\phi(a) \right| \\
&\qquad = \left| \sum_{a \in \Sigma} \left(\frac{N(a \,|\, a_1 \cdots a_n)}{n} - p(a) \right) \phi(a) \right| \\
&\qquad \leq \sum_{a \in \Sigma} \left| \frac{N(a \,|\, a_1 \cdots a_n)}{n} - p(a) \right| \phi(a) \leq \delta \sum_{a \in \Sigma} p(a)\phi(a),
\end{aligned} \tag{4.12}$$

as required. □

4.2 Lemmas

Next we will prove two lemmas that are needed for the proof of the main theorem to which this lecture is devoted. The first of these lemmas is the one due to Bjelaković and Siegmund-Schultze mentioned at the start of the lecture.

Lemma 4.4. *Let $\rho, \sigma \in \mathrm{D}(\mathcal{X})$ be density operators for which $\mathrm{im}(\rho) \subseteq \mathrm{im}(\sigma)$ and let $\delta > 0$ be a positive real number. There exist positive real numbers K and μ such that, for every positive integer n, there exists a projection operator Π_n acting on $\mathcal{X}^{\otimes n}$ satisfying $\left[\Pi_n, \sigma^{\otimes n}\right] = 0$,*

$$\left\langle \Pi_n, \rho^{\otimes n} \right\rangle \geq 1 - K \exp(-\mu n), \tag{4.13}$$

and

$$2^{(1+\delta) n \operatorname{Tr}(\rho \log(\sigma))} \Pi_n \leq \Pi_n \sigma^{\otimes n} \Pi_n \leq 2^{(1-\delta) n \operatorname{Tr}(\rho \log(\sigma))} \Pi_n. \tag{4.14}$$

Proof. By considering a spectral decomposition of σ, one may select an alphabet Σ, an orthonormal set $\{x_a \,:\, a \in \Sigma\} \subset \mathcal{X}$, and a probability vector $q \in \mathcal{P}(\Sigma)$ such that

$$\sigma = \sum_{a \in \Sigma} q(a) x_a x_a^* \tag{4.15}$$

and $q(a) > 0$ for all $a \in \Sigma$. Define a new probability vector $p \in \mathcal{P}(\Sigma)$ as

$$p(a) = x_a^* \rho x_a \tag{4.16}$$

for every $a \in \Sigma$. The fact that p is indeed a probability vector follows from the assumption that ρ is a density operator with $\mathrm{im}(\rho) \subseteq \mathrm{im}(\sigma)$.

Real numbers K and μ satisfying the requirements of the lemma may now be selected as follows:

$$\begin{aligned} K &= 2 |\operatorname{supp}(p)|, \\ \mu &= 2\delta^2 \min\bigl\{ p(a)^2 \,:\, a \in \Sigma,\ p(a) > 0 \bigr\}. \end{aligned} \tag{4.17}$$

Toward a verification that K and μ satisfying the requirements of the lemma, let

$$\Pi_n = \sum_{a_1 \cdots a_n \in S_{n,\delta}(p)} x_{a_1} x_{a_1}^* \otimes \cdots \otimes x_{a_n} x_{a_n}^*, \tag{4.18}$$

where $S_{n,\delta}(p)$ denotes the set of δ-strongly typical sequences with respect to the probability vector p, for every positive integer n. The condition $\left[\Pi_n, \sigma^{\otimes n}\right] = 0$ is immediate, while the bound

$$\begin{aligned} \left\langle \Pi_n, \rho^{\otimes n} \right\rangle &= \sum_{a_1 \cdots a_n \in S_{n,\delta}(p)} p(a_1) \cdots p(a_n) \\ &\geq 1 - 2 \sum_{\substack{a \in \Sigma \\ p(a) > 0}} \exp\bigl(-2n\delta^2 p(a)^2\bigr) \geq 1 - K \exp(-\mu n) \end{aligned} \tag{4.19}$$

follows directly from Lemma 4.2.

It remains to prove the inequalities in (4.14). As

$$\Pi_n \sigma^{\otimes n} \Pi_n = \sum_{a_1 \cdots a_n \in S_{n,\delta}(p)} q(a_1) \cdots q(a_n) x_{a_1} x_{a_1}^* \otimes \cdots \otimes x_{a_n} x_{a_n}^*, \tag{4.20}$$

these inequalities are equivalent to

$$-(1-\delta) n \operatorname{Tr}(\rho \log(\sigma)) \leq -\sum_{k=1}^{n} \log(q(a_k)) \leq -(1+\delta) n \operatorname{Tr}(\rho \log(\sigma)) \tag{4.21}$$

for every $a_1 \cdots a_n \in S_{n,\delta}(p)$. By taking $\phi(a) = -\log(q(a))$ for every $a \in \Sigma$ in Lemma 4.3, so that

$$\sum_{a \in \Sigma} p(a) \phi(a) = -\operatorname{Tr}(\rho \log(\sigma)), \tag{4.22}$$

the inequalities (4.21) are obtained, which completes the proof. □

The next lemma is just a simple technical fact concerning the inner-product of a product of projection operators with a density operator.

Lemma 4.5. *Let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator, let $\varepsilon > 0$ be a positive real number, and let Π and Δ be projection operators on $\mathcal{X}$ that satisfy $\langle \Pi, \rho \rangle \geq 1 - \varepsilon$ and $\langle \Delta, \rho \rangle \geq 1 - \varepsilon$. It is the case that*

$$\langle \Delta \Pi \Delta, \rho \rangle \geq 1 - 6\varepsilon. \tag{4.23}$$

Proof. Observe first that

$$\langle \Delta \Pi \Delta, \rho \rangle = \operatorname{Tr}(\Pi \Delta \rho \Delta \Pi) = \left\| \Pi \Delta \sqrt{\rho} \right\|_2^2 \geq \left| \langle \sqrt{\rho}, \Pi \Delta \sqrt{\rho} \rangle \right|^2 = |\langle \Delta \Pi, \rho \rangle|^2, \tag{4.24}$$

where the inequality is by the Cauchy–Schwarz inequality. Next, by the identity

$$\mathbb{1} = (\mathbb{1} - \Delta)(\mathbb{1} - \Pi) + \Delta + \Pi - \Delta \Pi, \tag{4.25}$$

one sees that

$$\begin{aligned} \langle \Delta \Pi, \rho \rangle &= \langle \Delta, \rho \rangle + \langle \Pi, \rho \rangle - 1 + \langle (\mathbb{1} - \Delta)(\mathbb{1} - \Pi), \rho \rangle \\ &\geq 1 - 2\varepsilon + \langle (\mathbb{1} - \Delta)(\mathbb{1} - \Pi), \rho \rangle. \end{aligned} \tag{4.26}$$

By the Cauchy–Schwarz inequality, we have

$$\begin{aligned} &\left| \langle (\mathbb{1} - \Delta)(\mathbb{1} - \Pi), \rho \rangle \right| = \left| \langle (\mathbb{1} - \Pi) \sqrt{\rho}, (\mathbb{1} - \Delta) \sqrt{\rho} \rangle \right| \\ &\quad \leq \left\| (\mathbb{1} - \Pi) \sqrt{\rho} \right\|_2 \left\| (\mathbb{1} - \Delta) \sqrt{\rho} \right\|_2 = \sqrt{\langle \mathbb{1} - \Pi, \rho \rangle} \sqrt{\langle \mathbb{1} - \Delta, \rho \rangle} \leq \varepsilon, \end{aligned} \tag{4.27}$$

and therefore

$$\langle (\mathbb{1} - \Delta)(\mathbb{1} - \Pi), \rho \rangle \geq -\varepsilon. \tag{4.28}$$

Consequently,

$$\big|\langle \Delta\Pi, \rho\rangle\big| \geq 1 - 3\varepsilon, \tag{4.29}$$

and therefore

$$\big\langle \Delta\Pi\Delta, \rho\big\rangle \geq (1 - 3\varepsilon)^2 \geq 1 - 6\varepsilon, \tag{4.30}$$

as required. □

Remark 4.6. If Δ commutes with ρ, then the bound established by the previous lemma can be improved to

$$\langle \Delta\Pi\Delta, \rho\rangle \geq 1 - 2\varepsilon. \tag{4.31}$$

To see that this is so, note that

$$\rho - \Delta\rho\Delta = (\mathbb{1} - \Delta)\rho(\mathbb{1} - \Delta), \tag{4.32}$$

and therefore

$$\langle \Delta\Pi\Delta, \rho\rangle = \langle \Pi, \rho\rangle + \langle \Pi, \Delta\rho\Delta - \rho\rangle = \langle \Pi, \rho\rangle - \langle \Pi, (\mathbb{1} - \Delta)\rho(\mathbb{1} - \Delta)\rangle. \tag{4.33}$$

Because $(\mathbb{1} - \Delta)\rho(\mathbb{1} - \Delta)$ is positive semidefinite and $\Pi \leq \mathbb{1}$, it follows that

$$\langle \Delta\Pi\Delta, \rho\rangle \geq \langle \Pi, \rho\rangle - \operatorname{Tr}\big((\mathbb{1} - \Delta)\rho(\mathbb{1} - \Delta)\big) = \langle \Pi, \rho\rangle - \langle \mathbb{1} - \Delta, \rho\rangle \geq 1 - 2\varepsilon. \tag{4.34}$$

Remark 4.7. The proof of the lemma above can be extended to the assumptions $\langle \Pi, \rho\rangle \geq 1 - \varepsilon$ and $\langle \Delta, \rho\rangle \geq 1 - \delta$ to obtain

$$\langle \Delta\Pi\Delta, \rho\rangle \geq \Big(1 - \varepsilon - \delta - \sqrt{\varepsilon\delta}\Big)^2. \tag{4.35}$$

We're also going to make use of Winter's gentle measurement lemma, which is a very useful and well-known fact.

Lemma 4.8 (Winter's gentle measurement lemma)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator, and let $P \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator satisfying $P \leq \mathbb{1}$ and $\langle P, \rho\rangle > 0$. This inequality is satisfied:*

$$\mathrm{F}\left(\rho, \frac{\sqrt{P}\rho\sqrt{P}}{\langle P, \rho\rangle}\right) \geq \sqrt{\langle P, \rho\rangle}. \tag{4.36}$$

Proof. Observe that for any two positive semidefinite operators Q and R, it is necessarily the case that $\mathrm{F}(R, QRQ) = \langle R, Q\rangle$. Indeed,

$$\sqrt{\sqrt{R}QRQ\sqrt{R}} = \sqrt{\big(\sqrt{R}Q\sqrt{R}\big)^2} = \sqrt{R}Q\sqrt{R}, \tag{4.37}$$

and therefore

$$\mathrm{F}(R, QRQ) = \mathrm{Tr}\left(\sqrt{\sqrt{R}QRQ\sqrt{R}}\right) = \mathrm{Tr}\left(\sqrt{R}Q\sqrt{R}\right) = \langle R, Q \rangle. \tag{4.38}$$

By this formula, along with the square root scaling of the fidelity function, one finds that

$$\mathrm{F}\left(\rho, \frac{\sqrt{P}\rho\sqrt{P}}{\langle P, \rho \rangle}\right) = \frac{1}{\sqrt{\langle P, \rho \rangle}}\,\mathrm{F}\left(\rho, \sqrt{P}\rho\sqrt{P}\right) = \frac{\langle \sqrt{P}, \rho \rangle}{\sqrt{\langle P, \rho \rangle}}. \tag{4.39}$$

Finally, under the assumption $0 \leq P \leq \mathbb{1}$, it is the case that $\sqrt{P} \geq P$, and therefore $\langle \sqrt{P}, \rho \rangle \geq \langle P, \rho \rangle$, from which the lemma follows. □

Remark 4.9. We will actually only need the lemma for P being a projection operator, in which case $\sqrt{P} = P$, and so the lemma holds with equality in (4.36).

4.3 Main theorem

Now we're ready for the main theorem and its proof. Here it is.

Theorem 4.10. *Let $\rho, \sigma \in \mathrm{D}(\mathcal{X})$ be density operators. For every $\varepsilon \in (0, 1)$, the following equality holds.*

$$\lim_{n \to \infty} \frac{\mathrm{D}^{\varepsilon}_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n})}{n} = \mathrm{D}(\rho \| \sigma). \tag{4.40}$$

Proof. Let us first consider the case that $\mathrm{im}(\rho) \not\subseteq \mathrm{im}(\sigma)$. In this case the right-hand side of (4.40) is infinite, and so we must prove the same for the left-hand side. This follows from the fact that $\mathrm{D}^{\varepsilon}_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n})$ is infinite for all but finitely many positive integers n. Indeed, let Λ be the projection onto $\mathrm{im}(\sigma)$, so that $\sigma = \Lambda\sigma\Lambda$ and $\langle \Lambda, \rho \rangle < 1$. Now, for a given choice of n, one has that if $\xi_n \in \mathrm{D}(\mathcal{X}^{\otimes n})$ satisfies $\mathrm{im}(\xi_n) \subseteq \mathrm{im}(\sigma^{\otimes n})$, then

$$\mathrm{F}\left(\xi_n, \rho^{\otimes n}\right) = \mathrm{F}\left(\xi_n, \Lambda^{\otimes n}\rho^{\otimes n}\Lambda^{\otimes n}\right) \leq \sqrt{\mathrm{Tr}\left(\Lambda^{\otimes n}\rho^{\otimes n}\Lambda^{\otimes n}\right)} < \langle \Lambda, \rho \rangle^{\frac{n}{2}}, \tag{4.41}$$

and therefore

$$\frac{1}{2}\left\| \xi_n - \rho^{\otimes n} \right\|_1 > 1 - \langle \Lambda, \rho \rangle^{\frac{n}{2}} \tag{4.42}$$

by one of the Fuchs–van de Graaf inequalities. The right-hand side of this inequality must exceed ε for all but finitely many positive integers n, by the fact that $\langle \Lambda, \rho \rangle < 1$. It follows that for all but finitely many positive integers n, there are no elements of $\mathcal{B}_{\varepsilon}(\rho^{\otimes n})$ whose images are contained in $\mathrm{im}(\sigma^{\otimes n})$, which implies for any such n that $\mathrm{D}^{\varepsilon}_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n}) = \infty$.

For the remainder of the proof it will be assumed that $\mathrm{im}(\rho) \subseteq \mathrm{im}(\sigma)$. Let $\delta > 0$ be chosen arbitrarily, and for each positive integer n, let Π_n be the projection whose existence is guaranteed by Lemma 4.4 for ρ, σ, δ, and n, and also let Δ_n be the projection whose existence is guaranteed by Lemma 4.4, again for ρ, δ, and n, but where σ is replaced by ρ. Thus, we have $[\Pi_n, \sigma^{\otimes n}] = 0$ and $[\Delta_n, \rho^{\otimes n}] = 0$, and the following inequalities are satisfied:

$$2^{(1+\delta)n\,\mathrm{Tr}(\rho\log(\sigma))}\Pi_n \leq \Pi_n\sigma^{\otimes n}\Pi_n \leq 2^{(1-\delta)n\,\mathrm{Tr}(\rho\log(\sigma))}\Pi_n, \tag{4.43}$$

$$2^{-(1+\delta)n\,\mathrm{H}(\rho)}\Delta_n \leq \Delta_n\rho^{\otimes n}\Delta_n \leq 2^{-(1-\delta)n\,\mathrm{H}(\rho)}\Delta_n, \tag{4.44}$$

$$\langle \Delta_n, \rho^{\otimes n}\rangle \geq 1 - K\exp(-\mu n), \tag{4.45}$$

$$\langle \Pi_n, \rho^{\otimes n}\rangle \geq 1 - K\exp(-\mu n), \tag{4.46}$$

where K and μ are positive constants independent of n.

It will first be proved that

$$\lim_{n\to\infty} \frac{\mathrm{D}^{\varepsilon}_{\max}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big)}{n} \leq \mathrm{D}(\rho\|\sigma). \tag{4.47}$$

The projection operator Δ_n commutes with $\rho^{\otimes n}$, and therefore

$$\langle \Delta_n\Pi_n\Delta_n, \rho^{\otimes n}\rangle \geq 1 - 2K\exp(-\mu n) \tag{4.48}$$

by Remark 4.6. It is therefore the case that $\langle \Delta_n, \rho^{\otimes n}\rangle$, $\langle \Pi_n, \rho^{\otimes n}\rangle$, and $\langle \Delta_n\Pi_n\Delta_n, \rho^{\otimes n}\rangle$ are all positive for all but finitely many n, and we will restrict our attention to those n for which these values are all positive.

Define a density operator

$$\xi_n = \frac{\Pi_n\Delta_n\rho^{\otimes n}\Delta_n\Pi_n}{\langle \Delta_n\Pi_n\Delta_n, \rho^{\otimes n}\rangle} \tag{4.49}$$

for all values of n under consideration. We will begin by proving a bound on the trace distance between ξ_n and $\rho^{\otimes n}$. To this end, observe that

$$\frac{1}{2}\big\|\xi_n - \rho^{\otimes n}\big\|_1 \leq \frac{1}{2}\big\|\xi_n - \tau_n\big\|_1 + \frac{1}{2}\big\|\tau_n - \rho^{\otimes n}\big\|_1 \tag{4.50}$$

for

$$\tau_n = \frac{\Delta_n\rho^{\otimes n}\Delta_n}{\big\langle \Delta_n, \rho^{\otimes n}\big\rangle}, \tag{4.51}$$

and notice that

$$\xi_n = \frac{\Pi_n\tau_n\Pi_n}{\langle \Pi_n, \tau_n\rangle}. \tag{4.52}$$

By Winter's gentle measurement lemma and one of the Fuchs–van de Graaf inequalities, we find that

$$\frac{1}{2}\left\|\xi_n - \tau_n\right\|_1 = \frac{1}{2}\left\|\frac{\Pi_n \tau_n \Pi_n}{\langle \Pi_n, \tau_n\rangle} - \tau_n\right\|_1 \leq \sqrt{1 - \langle \Pi_n, \tau_n\rangle}, \tag{4.53}$$

and because

$$\langle \Pi_n, \tau_n\rangle = \frac{\langle \Delta_n \Pi_n \Delta_n, \rho^{\otimes n}\rangle}{\langle \Delta_n, \rho^{\otimes n}\rangle} \geq \langle \Delta_n \Pi_n \Delta_n, \rho^{\otimes n}\rangle, \tag{4.54}$$

we obtain

$$\frac{1}{2}\left\|\xi_n - \tau_n\right\|_1 \leq \sqrt{1 - \langle \Delta_n \Pi_n \Delta_n, \rho^{\otimes n}\rangle} \leq \sqrt{2K \exp(-\mu n)}. \tag{4.55}$$

Along similar (although simpler) lines, we find that

$$\frac{1}{2}\left\|\tau_n - \rho^{\otimes n}\right\|_1 \leq \sqrt{1 - \langle \Delta_n, \rho^{\otimes n}\rangle} \leq \sqrt{K \exp(-\mu n)}. \tag{4.56}$$

These upper bounds are decreasing to 0 (exponentially quickly, as it happens), and therefore

$$\frac{1}{2}\left\|\xi_n - \rho^{\otimes n}\right\|_1 \leq \varepsilon, \tag{4.57}$$

or equivalently $\xi_n \in \mathcal{B}_\varepsilon(\rho^{\otimes n})$, implying

$$\mathrm{D}^{\varepsilon}_{\max}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \leq \mathrm{D}_{\max}\big(\xi_n\big\|\sigma^{\otimes n}\big), \tag{4.58}$$

for all but finitely many n. Let us further restrict our attention to these values of n.

Next, we will use the inequalities (4.43) and (4.44) to obtain an upper-bound on $\mathrm{D}_{\max}\big(\xi_n\big\|\sigma^{\otimes n}\big)$. First, by (4.44), together with $\Delta_n \leq \mathbb{1}$ and $\Pi_n^2 = \Pi_n$, we find that

$$\Pi_n \Delta_n \rho^{\otimes n} \Delta_n \Pi_n \leq 2^{-(1-\delta)n\,\mathrm{H}(\rho)} \Pi_n \Delta_n \Pi_n \leq 2^{-(1-\delta)n\,\mathrm{H}(\rho)} \Pi_n. \tag{4.59}$$

Second, by (4.44), together with the fact that $[\Pi_n, \sigma^{\otimes n}] = 0$, we have

$$\Pi_n \leq 2^{-(1+\delta)n\,\mathrm{Tr}(\rho\log(\sigma))} \Pi_n \sigma^{\otimes n} \Pi_n \leq 2^{-(1+\delta)n\,\mathrm{Tr}(\rho\log(\sigma))} \sigma^{\otimes n}. \tag{4.60}$$

Combining (4.59) and (4.60), we obtain

$$\Pi_n \Delta_n \rho^{\otimes n} \Delta_n \Pi_n \leq 2^{n\,\mathrm{D}(\rho\|\sigma) + \delta n(\mathrm{H}(\rho) - \mathrm{Tr}(\rho\log(\sigma)))} \sigma^{\otimes n}. \tag{4.61}$$

Accounting for the normalization of ξ_n and making use of (4.48), we find that

$$\begin{aligned}\mathrm{D}_{\max}\big(\xi_n\big\|\sigma^{\otimes n}\big) \leq n\,\mathrm{D}(\rho\|\sigma) + \delta n(\mathrm{H}(\rho) - \mathrm{Tr}(\rho\log(\sigma)))\\ - \log\big(1 - 2K\exp(-\mu n)\big).\end{aligned} \tag{4.62}$$

At this point we may conclude that

$$\lim_{n\to\infty} \frac{\mathrm{D}^{\varepsilon}_{\max}\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right)}{n} \leq \mathrm{D}(\rho\|\sigma) + \delta(\mathrm{H}(\rho) - \mathrm{Tr}(\rho\log(\sigma))), \tag{4.63}$$

and as δ was an arbitrarily chosen positive real number, we obtain the required inequality (4.47).

Now we will prove the reverse inequality

$$\lim_{n\to\infty} \frac{\mathrm{D}^{\varepsilon}_{\max}\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right)}{n} \geq \mathrm{D}(\rho\|\sigma). \tag{4.64}$$

Let $\delta > 0$ again be chosen arbitrarily. For every positive integer n it is the case that

$$\left\langle \sigma^{\otimes n}, \Pi_n \Delta_n \Pi_n \right\rangle = \left\langle \Delta_n, \Pi_n \sigma^{\otimes n} \Pi_n \right\rangle \leq 2^{(1-\delta) n \,\mathrm{Tr}(\rho \log(\sigma))} \left\langle \Delta_n, \Pi_n \right\rangle \tag{4.65}$$

by (4.43), as well as

$$\left\langle \Delta_n, \Pi_n \right\rangle \leq 2^{(1+\delta) n \,\mathrm{H}(\rho)} \left\langle \Delta_n \rho^{\otimes n} \Delta_n, \Pi_n \right\rangle \leq 2^{(1+\delta) n \,\mathrm{H}(\rho)} \tag{4.66}$$

by (4.44). It therefore follows that the operator

$$Z_n = 2^{n\,\mathrm{D}(\rho\|\sigma) - \delta n (\mathrm{H}(\rho) - \mathrm{Tr}(\rho\log(\sigma)))} \Pi_n \Delta_n \Pi_n \tag{4.67}$$

is positive semidefinite and satisfies $\langle \sigma^{\otimes n}, Z \rangle \leq 1$. By an inspection of the dual form of Optimization Problem 3.2, the conic program for the exponential of the smoothed max-relative entropy, we conclude that

$$\begin{aligned} \mathrm{D}^{\varepsilon}_{\max}\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) \geq{} & n\,\mathrm{D}(\rho\|\sigma) - \delta n(\mathrm{H}(\rho) - \mathrm{Tr}(\rho\log(\sigma))) \\ & + \inf_{\xi_n \in \mathcal{B}_\varepsilon(\rho^{\otimes n})} \log\langle \Pi_n \Delta_n \Pi_n, \xi_n \rangle. \end{aligned} \tag{4.68}$$

For every $\xi_n \in \mathcal{B}_\varepsilon(\rho^{\otimes n})$ we have, by virtue of the fact that $\rho^{\otimes n} - \xi_n$ is traceless and $0 \leq \Pi_n \Delta_n \Pi_n \leq \mathbb{1}$, that

$$\left\langle \Pi_n \Delta_n \Pi_n, \rho^{\otimes n} - \xi_n \right\rangle \leq \frac{1}{2} \left\| \rho^{\otimes n} - \xi_n \right\|_1 \leq \varepsilon \tag{4.69}$$

and therefore

$$\begin{aligned} \langle \Pi_n \Delta_n \Pi_n, \xi_n \rangle &= \langle \Pi_n \Delta_n \Pi_n, \rho^{\otimes n} \rangle + \langle \Pi_n \Delta_n \Pi_n, \xi_n - \rho^{\otimes n} \rangle \\ &\geq 1 - 6K\exp(-\mu n) - \varepsilon \end{aligned} \tag{4.70}$$

by Lemma 4.5. Consequently,

$$\begin{aligned} \mathrm{D}^{\varepsilon}_{\max}\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) \geq{} & n\,\mathrm{D}(\rho\|\sigma) - \delta n(\mathrm{H}(\rho) - \mathrm{Tr}(\rho\log(\sigma))) \\ & + \log\left(1 - 6K\exp(-\mu n) - \varepsilon\right). \end{aligned} \tag{4.71}$$

Given the assumption $\varepsilon \in (0,1)$, one concludes that $\log(1 - 6K\exp(-\mu n) - \varepsilon)$ converges to a constant value as n goes to infinity. It follows that

$$\lim_{n\to\infty} \frac{\mathrm{D}^{\varepsilon}_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n})}{n} \geq \mathrm{D}(\rho\|\sigma) - \delta(\mathrm{H}(\rho) - \mathrm{Tr}(\rho\log(\sigma))). \tag{4.72}$$

Once again, as δ was an arbitrarily chosen positive real number, the required inequality (4.64) follows. □

Remark 4.11. An alternative way to argue the closeness of ξ_n to $\rho^{\otimes n}$ is to use a different known equality concerning the fidelity function, which is that

$$\mathrm{F}\left(AA^*, BB^*\right) = \left\| A^* B \right\|_1 \tag{4.73}$$

for any choice of operators $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$. (This fact is closely connected with Uhlmann's theorem, and can be found as Lemma 3.21 in my book *Theory of Quantum Information*.) In the present case, we obtain

$$\begin{aligned} \mathrm{F}(\Pi_n\Delta_n\rho^{\otimes n}\Delta_n\Pi_n, \rho^{\otimes n}) &= \left\| \sqrt{\rho^{\otimes n}}\Delta_n\Pi_n\sqrt{\rho^{\otimes n}} \right\|_1 \\ &\geq \left| \mathrm{Tr}\left(\sqrt{\rho^{\otimes n}}\Delta_n\Pi_n\sqrt{\rho^{\otimes n}} \right) \right| = \langle \Delta_n\Pi_n\Delta_n, \rho^{\otimes n} \rangle, \end{aligned} \tag{4.74}$$

where the last equality makes use of $[\Delta_n, \rho^{\otimes n}] = 0$. A suitable bound on the trace distance between ξ_n and $\rho^{\otimes n}$ is obtained through the Fuchs–van de Graaf inequalities.

And we will conclude with two corollaries.

It is not important that σ is a density operator in Theorem 4.10—it is true for arbitrary positive semidefinite models. The only part of the proof that depends on the scaling of σ occurs in the proof of Lemma 4.4, where $q \in \mathcal{P}(\Sigma)$ implies that $\phi(a) = -\log(q(a))$ is nonnegative. Although it would not be difficult to modify this portion of the proof slightly to handle arbitrary positive semidefinite models, it is perhaps simpler to observe it as a fairly straightforward corollary of Theorem 4.10.

Corollary 4.12. *If ρ is a density operator and Q is any positive semidefinite operator, we have*

$$\lim_{n\to\infty} \frac{\mathrm{D}^{\varepsilon}_{\max}(\rho^{\otimes n} \| Q^{\otimes n})}{n} = \mathrm{D}(\rho\|Q). \tag{4.75}$$

Proof. Let $\sigma = Q/\mathrm{Tr}(Q)$. Then

$$\begin{aligned} \mathrm{D}^{\varepsilon}_{\max}(\rho^{\otimes n} \| Q^{\otimes n}) &= \mathrm{D}^{\varepsilon}_{\max}(\rho^{\otimes n} \| \mathrm{Tr}(Q)^n \sigma^{\otimes n}) \\ &= \mathrm{D}^{\varepsilon}_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n}) - n\log(\mathrm{Tr}(Q)) \end{aligned}$$

so

$$\begin{aligned}\lim_{n\to\infty}\frac{\mathrm{D}^{\varepsilon}_{\max}\big(\rho^{\otimes n}\big\|Q^{\otimes n}\big)}{n} &= \lim_{n\to\infty}\frac{\mathrm{D}^{\varepsilon}_{\max}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big)}{n} - \log(\operatorname{Tr}(Q))\\ &= \mathrm{D}(\rho\|\sigma) - \log(\operatorname{Tr}(Q))\\ &= \mathrm{D}(\rho\|Q),\end{aligned}$$

as required. □

The second and final corollary is quite spectacular. Of course it is well-known, and we proved it in CS 766/QIC 820, but now we have a completely different alternative proof.

Corollary 4.13 (Monotonicity of quantum relative entropy)**.** *Let $\rho, \sigma \in \mathrm{D}(\mathcal{X})$ be density operators and let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be a channel, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$. It is the case that $\mathrm{D}(\Phi(\rho)\|\Phi(\sigma)) \leq \mathrm{D}(\rho\|\sigma)$.*

Proof. Observe first that the smoothed max-relative entropy is monotonic:

$$\begin{aligned}\mathrm{D}^{\varepsilon}_{\max}(\Phi(\rho)\|\Phi(\sigma)) &\leq \inf_{\xi\in\mathcal{B}_{\varepsilon}(\rho)} \mathrm{D}_{\max}(\Phi(\xi)\|\Phi(\sigma))\\ &\leq \inf_{\xi\in\mathcal{B}_{\varepsilon}(\rho)} \mathrm{D}_{\max}(\xi\|\sigma) = \mathrm{D}^{\varepsilon}_{\max}(\rho\|\sigma)\end{aligned} \tag{4.76}$$

by the monotonicity of the max-relative entropy. Therefore,

$$\begin{aligned}\mathrm{D}(\Phi(\rho)\|\Phi(\sigma)) &= \lim_{n\to\infty}\frac{\mathrm{D}^{\varepsilon}_{\max}\big(\Phi(\rho)^{\otimes n}\big\|\Phi(\sigma)^{\otimes n}\big)}{n}\\ &= \lim_{n\to\infty}\frac{\mathrm{D}^{\varepsilon}_{\max}\big(\Phi^{\otimes n}(\rho^{\otimes n})\big\|\Phi^{\otimes n}(\sigma^{\otimes n})\big)}{n}\\ &\leq \lim_{n\to\infty}\frac{\mathrm{D}^{\varepsilon}_{\max}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big)}{n}\\ &= \mathrm{D}(\rho\|\sigma),\end{aligned} \tag{4.77}$$

as required. □