# Assignment 5

MATH 667 Quantum Information Theory

Mark Girard

12 April 2016

## Problem 1 (Quantum Compression with Ensemble Fidelity)

**Problem 1** (Exercise 12.8, p. 546 in Nielsen and Chuang)**.** Suppose that instead of adopting the definition of a quantum source based on a single density matrix $\rho$ and the entanglement fidelity, we instead adopted the following *ensemble* definition, that an (i.i.d.) quantum source is specified by an ensemble $\{p_j, |\psi_j\rangle\}$ of quantum states, and that consecutive uses of the source are independent and produce a state $|\psi_j\rangle$ with probability $p_j$. A compression-decompression scheme $(\mathcal{C}^n, \mathcal{D}^n)$ is said to be reliable in this definition if the *ensemble average fidelity* approaches 1 as $n \to \infty$:

$$\bar{F}_n = \sum_J p_{j_1} \cdots p_{j_n} \left[F(\rho_J, \mathcal{D}^n \circ \mathcal{C}^n(\rho_J))\right]^2$$

where $J = (j_1, \dots, j_n)$ and $\rho_J = |\psi_{j_1}\rangle\langle\psi_{j_1}| \otimes \cdots \otimes |\psi_{j_n}\rangle\langle\psi_{j_n}|$. Define $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ and show that provided $r > S(\rho)$ there exists a reliable compression scheme of rate $r$ with respect to this definition of fidelity.

**Solution.** Here, $\rho$ is an operator on a space $\mathcal{H} = \mathbb{C}^d$ and each pure state $|\psi_j\rangle$ of the ensemble is a vector $|\psi_j\rangle \in \mathcal{H}$. Let $\rho = \sum_x q_x |x\rangle\langle x|$ be the spectral decomposition of $\rho$. For every positive integer $n$ and real number $\varepsilon > 0$, let $P(n, \varepsilon)$ be the projection operator onto the $\varepsilon$-typical subspace of $\mathcal{H}^{\otimes n}$ with respect to the probabilities $\{q_x\}$ and orthonormal basis $\{|x\rangle\}$ of $\mathcal{H}$ (i.e., it is the typical subspace in the standard sense as in Schumacher's compression theorem).

*Proof.* Let $r > S(\rho)$ and let $\varepsilon > 0$ such that $r - \varepsilon > S(\rho)$. From Schumacher's theorem, there is a (Schumacher-)reliable compression-decompression scheme $(\mathcal{C}^n, \mathcal{D}^n)$ of rate $r$ for the state $\rho$, and this composite compression-decompression channel can have the form

$$(\mathcal{D}^n \circ \mathcal{C}^n)(\sigma) = P(n, \varepsilon)\sigma P(n, \varepsilon) + \mathrm{Tr}[(I - P(n, \varepsilon))\sigma] |\phi_0\rangle\langle\phi_0|$$

for all $\sigma$, where $|\phi_0\rangle \in \mathcal{H}^n$ is some arbitrary normalized pure state. Recall that the fidelity of a pure state $|u\rangle$ and any operator $A$ reduces to

$$F(|u\rangle\langle u|, A) = \sqrt{\langle u|A|u\rangle}.$$

Hence, computing ensemble average fidelity yields

$$\begin{aligned}
\bar{F}_n &= \sum_J p_J \left[F(\rho_J, \mathcal{D}^n \circ \mathcal{C}^n(\rho_J))\right]^2 \\
&= \sum_J p_J \langle\psi_J| \Big(\mathcal{D}^n \circ \mathcal{C}^n(|\psi_J\rangle\langle\psi_J|)\Big) |\psi_J\rangle \\
&= \sum_J p_J \langle\psi_J| \Big(P(n, \varepsilon)|\psi_J\rangle\langle\psi_J|P(n, \varepsilon) + \mathrm{Tr}\Big[(I - P(n, \varepsilon))|\psi_J\rangle\langle\psi_J|\Big] |\phi_0\rangle\langle\phi_0|\Big) |\psi_J\rangle
\end{aligned}$$

$$
\begin{aligned}
&= \sum_J p_J |\langle \psi_J | P(n,\varepsilon) | \psi_J \rangle|^2 + \mathrm{Tr}\Big[(I - P(n,\varepsilon)) |\psi_J\rangle\langle\psi_J|\Big] |\langle \psi_J | \phi_0 \rangle|^2 \\
&\geq \sum_J p_J |\langle \psi_J | P(n,\varepsilon) | \psi_J \rangle|^2 \\
&\geq \sum_J p_J \left(2\langle \psi_J | P(n,\varepsilon) | \psi_J \rangle - 1\right),
\end{aligned}
$$

where in the final line we use the fact that $a^2 \geq 2a - 1$ holds for all real numbers $a$. Continuing, we have

$$
\begin{aligned}
\bar{F}_n &\geq \sum_J p_J \left(2\langle \psi_J | P(n,\varepsilon) | \psi_J \rangle - 1\right) \\
&= 2\,\mathrm{Tr}\Bigg[\underbrace{\sum_J p_J |\psi_J\rangle\langle\psi_J|}_{\rho^{\otimes n}} P(n,\varepsilon)\Bigg] - \underbrace{\sum_J p_J}_{=1} \\
&= 2\,\mathrm{Tr}[\rho^{\otimes n} P(n,\varepsilon)] - 1.
\end{aligned}
$$

Furthermore, recall from the theory of typical sequences that

$$
\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} P(n,\varepsilon)] = 1,
$$

and thus

$$
\lim_{n\to\infty} \bar{F}_n \geq 2 \lim_{n\to\infty} \big(\mathrm{Tr}[\rho^{\otimes n} P(n,\varepsilon)]\big) - 1 = 1.
$$

Since it is clear that $\bar{F}_n \leq 1$ for every $n$, it follows that $\bar{F}_n \to 1$, as desired. $\square$

## Problem 2 (Operational interpretation for the entropy of entanglement)

**Problem 2.** Let $|\psi\rangle$ and $|\phi\rangle$ be two entangled pure bipartite states in $\mathbb{C}^{d_\mathsf{A}} \otimes \mathbb{C}^{d_\mathsf{B}}$. For each integer $n$, define an integer $m$ (depending on $n$) such that $m(n)$ is the maximal integer for which the transformation

$$|\psi\rangle^{\otimes n} \to |\phi\rangle^{\otimes m}$$

is possible by LOCC. Show that

$$\lim_{n\to\infty} \frac{m}{n} = \frac{E(|\psi\rangle)}{E(|\phi\rangle)}$$

where $E(\cdot)$ is the entropy of entanglement, i.e. $E(|\psi\rangle) = S(\rho_\mathsf{A})$ with $S(\cdot)$ being the von-Neumann entropy and $\rho_\mathsf{A} = \mathrm{Tr}_\mathsf{B}|\psi\rangle\langle\psi|$.

**Solution.** The fact that the optimal rate for converting copies of $|\psi\rangle$ into copies of $|\phi\rangle$ is $\frac{E(|\psi\rangle)}{E(|\phi\rangle)}$ will be shown by putting together the following to propositions.

**Proposition 2.1.** *The rate $r = \frac{E(|\psi\rangle)}{E(|\phi\rangle)}$ is achievable.*

*Proof.* From the analysis of asymptotic entanglement manipulation that was done in the lecture, we know that it is possible to convert

$$|\psi\rangle^{\otimes n} \xrightarrow{\text{LOCC}}\approx |\text{Bell}\rangle^{\otimes nE(\psi)}$$

for $n$ sufficiently large. Similarly, we can convert

$$|\text{Bell}\rangle^{\otimes mE(\phi)} \xrightarrow{\text{LOCC}}\approx |\phi\rangle^{\otimes m}$$

for $m$ large enough. Hence, for sufficiently large $n$, one can convert

$$|\psi\rangle^{\otimes n} \xrightarrow{\text{LOCC}}\approx |\phi\rangle^{\otimes n\frac{E(\psi)}{E(\phi)}},$$

and thus the rate $r = \frac{E(\psi)}{E(\phi)}$ is achievable. □

**Proposition 2.2.** *Any rate $r$ above $\frac{E(|\psi\rangle)}{E(|\phi\rangle)}$ is not achievable.*

*Proof.* Suppose that there is an achievable rate $r > \frac{E(\psi)}{E(\phi)}$. This means that it is possible to convert

$$|\psi\rangle^{\otimes n} \xrightarrow{\text{LOCC}}\approx |\phi\rangle^{\otimes nr}$$

for sufficiently large $n$ (up to some small error). Using optimal protocols for entanglement cost and distillation, for sufficiently large $n$ it is possible to convert

$$|\text{Bell}\rangle^{\otimes nE(\psi)} \xrightarrow{\text{LOCC}}\approx |\psi\rangle^{\otimes n} \xrightarrow{\text{LOCC}}\approx |\phi\rangle^{\otimes nr} \xrightarrow{\text{LOCC}}\approx |\text{Bell}\rangle^{\otimes nrE(\phi)}$$

via LOCC (up to some small error). However,

$$\lim_{n\to\infty} \frac{nrE(\phi)}{nE(\psi)} = r\frac{E(\phi)}{E(\psi)} > 1.$$

Hence for sufficiently large $n$ it holds that

$$\lceil nE(\psi)\rceil < \lfloor nrE(\phi)\rfloor.$$

Since the protocol implies that $\lceil nE(\psi)\rceil$ copies of a Bell state can be converted into $\lfloor nrE(\phi)\rfloor$ copies of a Bell state, this implies that it would be possible to produce more Bell states than were started with. This is clearly impossible and thus $r > \frac{E(\psi)}{E(\phi)}$ is not an achievable rate. □

## Problem 3 (Relative entropy of entanglement)

**Problem 3.** The relative entropy of entanglement is a measure of entanglement for mixed bipartite states defined by

$$E_R(\rho_{\mathsf{AB}}) = \min_{\sigma_{\mathsf{AB}} \in \mathfrak{D}_{\mathsf{AB}}} \{S(\rho_{\mathsf{AB}} \| \sigma_{\mathsf{AB}})\}$$

where $S(\rho_{\mathsf{AB}} \| \sigma_{\mathsf{AB}})$ is the relative entropy of entanglement defined by

$$S(\rho \| \sigma) = \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma)$$

and $\mathfrak{D}_{\mathsf{AB}}$ is the convex set of all separable density operators on $\mathbb{C}^{d_{\mathsf{A}}} \otimes \mathbb{C}^{d_{\mathsf{B}}}$.

(a) Show that $S(\rho \| \sigma) \geq 0$ with equality if and only if $\rho = \sigma$.

(b) Show that the relative entropy of entanglement of $\rho_{\mathsf{AB}}$ is an entanglement monotone. To prove it use the fact that $S(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) \leq S(\rho \| \sigma)$ for any CPTP map $\mathcal{E}$.

(c) Show that if $\rho_{\mathsf{AB}}$ is a pure state then $E_R\left(\rho_{\mathsf{AB}}\right) = S\left(\rho_{\mathsf{A}}\right)$, where $\rho_{\mathsf{A}} = \mathrm{Tr}_{\mathsf{B}}\, \rho_{\mathsf{AB}}$ and $S(\cdot)$ is the von-Neumann entropy. That is, for pure states the relative entropy of entanglement is equal to the entropy of entanglement.

(d) Define the function

$$\tau_R(\rho_{\mathsf{AB}}) = \min_{\sigma_{\mathsf{AB}} \in \mathfrak{D}_{\mathsf{AB}}} \left\{ \tilde{D}\left(\rho_{\mathsf{AB}} \| \sigma_{\mathsf{AB}}\right) \right\},$$

where $\tilde{D}\left(\rho \| \sigma\right) = 1 - \left[F\left(\rho, \sigma\right)\right]^2$ and $F(\rho, \sigma)$ is the fidelity. Prove that $\tau_R$ is an entanglement monotone. Bonus: Find what it equals to fore pure states.

**Solution.** .

(a) We first prove the inequality for the classical relative entropy.

**Proposition 3.1.** *For all probability distributions $p$ and $\{q_i\}$, it holds that $H(\{p_i\} \| \{q_i\}) \geq 0$ with equality if and only if $p_i = q_i$ for all $i$.*

*Proof.* First suppose that $\mathrm{supp}(\{p_i\}) \not\subset \mathrm{supp}(\{q_i\})$. It follows that $H(\{p_i\} \| \{q_i\}) = +\infty$ and the inequality is trivially satisfied.

If it holds that $\mathrm{supp}(\{p_i\}) \subset \mathrm{supp}(\{q_i\})$, then $p_i = 0$ whenever $q_i = 0$. Taking the sum of the $q_i$ over all $i$ such that $p_i \neq 0$ yields

$$\sum_{\{i | p_i \neq 0\}} q_i \leq 1.$$

The desired inequality now follows, since

$$\begin{aligned}
D(\{p_i\} \| \{q_i\}) = \sum_{\{i | p_i \neq 0\}} p_i \log \frac{p_i}{q_i} &= -\frac{1}{\ln 2} \sum_{\{i | p_i \neq 0\}} p_i \ln \frac{q_i}{p_i} \\
&\geq \frac{1}{\ln 2} \sum_{\{i | p_i \neq 0\}} p_i \left(1 - \frac{q_i}{p_i}\right) \qquad (3.1) \\
&= \frac{1}{\ln 2} \sum_{\{i | p_i \neq 0\}} (p_i - q_i) \\
&= \frac{1}{\ln 2} \left(1 - \sum_{\{i | p_i \neq 0\}} q_i\right) \geq 0,
\end{aligned}$$

where the inequality in (3.1) is due to the fact that $\ln x \leq x - 1$ for all $x > 0$, which we prove as follows. Define the function $f$ by $f(x) = \ln x + 1 - x$ for all $x > 0$. Then $f$ achieves a global maximum at $x = 1$, since $f'(x) = \frac{1}{x} - 1$ with $f'(1) = 0$ and $f''(x) = -\frac{1}{x^2} \leq 0$, and thus $\ln x + 1 - x \geq 0$ for all $x > 0$. □

**Proposition 3.2.** *For all density operators $\rho$ and $\sigma$, it holds that $S(\rho\|\sigma) \geq 0$ with equality if and only if $\rho = \sigma$.*

*Proof.* Denote the *distinct* eigenvalues of $\rho$ and $\sigma$ by $\{p_i\}$ and $\{q_j\}$ respectively such that we can decompose $\rho$ and $\sigma$ as

$$\rho = \sum_i p_i P_i \qquad \text{and} \qquad \sigma = \sum_j q_j Q_j,$$

where $P_i$ and $Q_j$ are the projection operators onto the eigenspaces whose corresponding eigenvalues are $p_i$ and $q_j$ respectively. The collections of projection matrices, $\{P_i\}$ and $\{Q_j\}$, are orthogonal in the sense that

$$P_i P_{i'} = P_{i'} P_i = \delta_{ii'} P_i \qquad \text{and} \qquad Q_j Q_{j'} = Q_{j'} Q_j = \delta_{jj'} Q_j$$

for all $i, i', j, j'$, and satisfy

$$\sum_i P_i = I = \sum_j Q_j.$$

Let $m_i = \operatorname{Tr} P_i$ and $n_j = \operatorname{Tr} Q_j$ denote the multiplicity of the eigenvalues $p_i$ and $q_j$ respectively. Note that

$$\log \rho = \sum_i \log p_i P_i \qquad \text{and} \qquad \log \sigma = \sum_j \log q_j Q_j.$$

Define the quantities $D_{ij} = \frac{1}{n_i} \operatorname{Tr}[P_i Q_j]$ for each $i$ and $j$. It is clear that each $D_{ij} \geq 0$ is nonnegative and that

$$\sum_j D_{ij} = \frac{1}{n_i} \sum_j \operatorname{Tr}[P_i Q_j] = \frac{1}{n_i} \operatorname{Tr}[P_i] = 1$$

for each $i$. Since the logarithm is a concave function, it holds that

$$\sum_j D_{ij} \log q_j \geq \log\left(\sum_j D_{ij} q_j\right) = \log r_i \tag{3.2}$$

for all $i$, where we define the quantities $r_i = \sum_j D_{ij} q_j$. Then

$$\begin{aligned}
S(\rho\|\sigma) &= \sum_i p_i \log p_i \operatorname{Tr}[P_i] - \sum_{i,j} p_i \log q_j \operatorname{Tr}[P_i Q_j] \\
&= \sum_i n_i p_i \log p_i - \sum_{i,j} n_i p_i \log q_j D_{ij} \\
&= \sum_i n_i p_i \log p_i - \sum_i n_i p_i \sum_j D_{ij} \log q_j \\
&\geq \sum_i n_i p_i \log p_i - \sum_i n_i p_i \log \underbrace{\left(\sum_j D_{ij} q_j\right)}_{r_i} \qquad (3.3) \\
&= \sum_i n_i p_i \log \frac{n_i p_i}{n_i r_i} \\
&= H(\{n_i p_i\}\|\{n_i r_i\}) \qquad (3.4) \\
&\geq 0 \qquad (3.5)
\end{aligned}$$

where we note that $\{n_i p_i\}$ and $\{n_i r_i\}$ are probability distributions since

$$\sum_i n_i p_i = \sum_i p_i \operatorname{Tr}[P_i] = \operatorname{Tr}[\rho] = 1 \qquad \text{and} \qquad \sum_i n_i r_i = \sum_{i,j} n_i D_{ij} q_j = \sum_{i,j} q_j \operatorname{Tr}[P_i Q_j] = \operatorname{Tr}[\sigma] = 1.$$

Hence positivity of $S(\rho\|\sigma)$ reduces to the positivity of the classical relative entropy (3.4) of the distributions $\{n_i p_i\}$ and $\{n_i r_i\}$.

To prove conditions for equality, note that, from the classical relative relative entropy, equality in (3.5) holds if and only if $p_i = r_i$ for all $i$. Since each $q_j$ is distinct, equality in both (3.2) and (3.3) holds if and only if for all $i$ there exists a $j_i$ such that

$$D_{ij} = 1 \text{ if } j = j_i \quad \text{and} \quad D_{ij} = 0 \text{ if } j \neq j_i.$$

(That is, for each $i$ there is exactly one $j$ such that $D_{ij} = 1$ and $D_{ij} = 0$ for all other $j$.)

Suppose $S(\rho\|\sigma) = 0$. Then $p_i = r_i = q_{j_i}$ for each $i$. Finally, we will show that $Q_{j_i} = P_i$ for all $i$ as well. For the sake of obtaining a contradiction, suppose that $Q_{j_i} \neq P_i$ for some $i$. Since $P_i$ and $Q_{j_i}$ are projection operators, it holds that $\mathrm{Tr}[P_i Q_{j_i}] < \mathrm{Tr}[P_i] = n_i$. Hence

$$1 = D_{ij_i} = \frac{1}{n_i}\mathrm{Tr}[P_i Q_{j_i}] < \frac{1}{n_i}\mathrm{Tr}[P_i] = 1$$

a contradiction. Hence

$$\rho = \sum_i p_i P_i = \sum_i q_{j_i} Q_{j_i} = \sum_j q_j Q_j = \sigma$$

as desired. □

(b) If $\sigma_{\mathsf{AB}}$ is a separable density operator, then $\mathcal{E}(\sigma_{\mathsf{AB}})$ is also separable for any LOCC channel $\mathcal{E}$. To show that $E_R$ is an entanglement monotone, it suffices to show that $E_R(\rho) \geq E_R(\mathcal{E}(\rho))$ for any LOCC channel $\mathcal{E}$.

Let $\mathcal{E}$ be an LOCC channel from $\mathsf{AB}$ to $\mathsf{A'B'}$ and denote the image of the separable density operators under the channel $\mathcal{E}$ as

$$\mathcal{E}(\mathfrak{D}_{\mathsf{AB}}) = \{\mathcal{E}(\sigma_{\mathsf{AB}}) \,|\, \sigma_{\mathsf{AB}} \in \mathfrak{D}_{\mathsf{AB}}\}.$$

It follows that $\mathcal{E}(\mathfrak{D}_{\mathsf{AB}}) \subseteq \mathfrak{D}_{\mathsf{A'B'}}$. For an arbitrary density operator $\rho = \rho_{\mathsf{AB}}$, we have

$$\begin{aligned} E_R(\rho) = \min_{\sigma \in \mathfrak{D}_{\mathsf{AB}}} S(\rho\|\sigma) &\geq \min_{\sigma \in \mathfrak{D}_{\mathsf{AB}}} S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) && (3.6) \\ &= \min_{\tau \in \mathcal{E}(\mathfrak{D}_{\mathsf{AB}})} S(\mathcal{E}(\rho)\|\tau) \\ &\geq \min_{\tau \in \mathfrak{D}_{\mathsf{A'B'}}} S(\mathcal{E}(\rho)\|\tau) = E_R(\mathcal{E}(\rho)), && (3.7) \end{aligned}$$

where the inequalities in the above equations hold due to the following observations:

- the inequality in (3.6) follows from the fact that $S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \geq S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma))$ holds for any states $\rho$ and $\sigma$ and any channel $\mathcal{E}$;
- and the inequality in (3.7) follows from the fact that $\mathcal{E}(\mathfrak{D}_{\mathsf{AB}}) \subseteq \mathfrak{D}_{\mathsf{A'B'}}$.

(c) Note that $S(\cdot\|\cdot)$ is strictly convex in the second argument, i.e. for all states $\rho, \sigma$, and $\tau$ and all $t \in [0,1]$ it holds that

$$S(\rho\|(1-t)\sigma + t\tau) \leq (1-t)S(\rho\|\sigma) + tS(\rho\|\tau).$$

A separable density operator $\sigma$ is said to be ***optimal*** for another density operator $\rho$ if and only if $E_R(\rho) = S(\rho\|\sigma)$. Hence $\sigma$ is optimal for $\rho$ if and only if $S(\rho\|\sigma) \leq S(\rho\|\tau)$ holds for all separable density operators $\tau$. The optimality conditions can be restated as follows.

**Proposition 3.3.** *Let $\rho$ and $\sigma$ be bipartite density operators and let $\sigma$ be separable. It holds that $\sigma$ is optimal for $\rho$ if and only if the directional derivative*

$$\frac{d}{dt} S(\rho\|(1-t)\sigma + t\tau) = \lim_{t \to 0^+} \frac{S(\rho\|(1-t)\sigma + t\tau)}{t} \geq 0$$

*is non-negative for all other separable states $\tau$.*

*Proof.* See (Girard 2014)[1]. □

Let $\rho = |\psi\rangle\langle\psi|$ be a pure state. Without loss of generality we may write $|\psi\rangle$ in Schmidt form as

$$|\psi\rangle = \sum_i \sqrt{p_i}|ii\rangle,$$

where $\{p_i\}$ are the Schmidt coefficients. Let $\sigma$ be the separable state defined by

$$\sigma = \sum_i p_i |ii\rangle\langle ii|.$$

We will show that $\sigma$ is optimal for $\rho$. Let $\tau$ be an arbitrary density operator. From (Girard 2014) it follows that the directional derivative can be written as

$$\lim_{t\to 0^+} \frac{S(\rho\|(1-t)\sigma + t\tau)}{t} = 1 - \mathrm{Tr}\left[D_\sigma(\rho)\tau\right] \tag{3.8}$$

where

$$D_\sigma(\rho) = \sum_{i,j} \sqrt{p_i p_j}\Delta(p_i, p_j)|ii\rangle\langle jj|$$

and $\Delta(p, q)$ are the so-called 'divided differences'

$$\Delta(p,q) = \begin{cases} \frac{\log p - \log q}{p-q} & p \neq q \\ \frac{1}{p} & p = q. \end{cases}$$

Note that $\Delta(p,p) = \frac{1}{p}$ and thus $\sqrt{pq}\Delta(p,q) = 1$ if $p = q$. If $p \neq q$, note that

$$\sqrt{pq}\Delta(p,q) = \sqrt{pq}\frac{\log\frac{p}{q}}{p-q} = \sqrt{\frac{p}{q}}\frac{\log\frac{p}{q}}{\frac{p}{q}-1} = f\left(\frac{p}{q}\right)$$

where $f$ is the function defined by $f(t) = \frac{\sqrt{t}\log t}{t-1}$ for all positive $t$ with $t \neq 1$. We may also define $f(1) = 1$ to make $f$ continuous.

**Proposition 3.4.** *For all $t > 0$ it holds that $0 < f(t) \leq 1$, where $f$ is the function defined by $f(1) = 1$ and $f(t) = \frac{\sqrt{t}\log t}{t-1}$ for $t > 0$ with $t \neq 1$.*

*Proof.* The inequalities hold trivially for $t = 1$, so suppose $t \neq 1$. Since $f(t^{-1}) = f(t)$, we may assume without loss of generality that $t > 1$. Define the function $g(t) = \sqrt{t} - \frac{1}{\sqrt{t}} - \log t$ for all $t > 0$. Note that $g(1) = 0$, and that $g$ is monotonically increasing on the interval $(1, \infty)$, since

$$\frac{dg}{dt} = \frac{1}{2\sqrt{t}} + \frac{1}{2t\sqrt{t}} - t = \frac{1}{t\sqrt{t}}\left(\frac{t+1}{2} - \sqrt{t}\right) \geq 0$$

where we use the fact that $\frac{t+1}{2} \geq \sqrt{t}$ holds for all $t > 0$. Therefore $0 \leq g(t) = \sqrt{t} - \frac{1}{\sqrt{t}} - \log t$ holds for all $t > 1$. This is equivalent to $f(t) \leq 1$ for all $t > 1$, as desired.

The inequality $0 < f(t)$ is obvious, since $0 < \frac{\log t}{t-1}$ holds for all $t > 0$ with $t \neq 1$. □

[1] M Girard, G Gour, S Friedland. 2014. *On convex optimization problems in quantum information theory.* Journal of Physics A: Mathematical and Theoretical **47** 505302.

From the above proposition, we see that $0 \leq \sqrt{p_i p_j}\Delta(p_i, p_j) \leq 1$ holds for all $p_i, p_j \geq 0$. Consider an arbitrary separable *pure* state of the form $\tau = |u\rangle\langle u| \otimes |v\rangle\langle v|$ with

$$|u\rangle = \sum_i \alpha_i |i\rangle \qquad \text{and} \qquad |v\rangle = \sum_i \beta_i |i\rangle.$$

Then $\tau$ can be written as

$$\tau = \sum_{i,j} \alpha_i \beta_i \overline{\alpha}_j \overline{\beta}_j |ii\rangle\langle jj|$$

and we have that $1 - \mathrm{Tr}\left[D_\sigma(\rho)\tau\right] \geq 0$ since

$$\begin{aligned} \big|\,\mathrm{Tr}\left[D_\sigma(\rho)\tau\right]\big| &= \left|\sum_{i,j} \sqrt{p_i p_j}\Delta(p_i, p_j)\alpha_i \beta_i \overline{\alpha}_j \overline{\beta}_j\right| \\ &\leq \sum_{i,j} |\alpha_i||\beta_i||\overline{\alpha}_j||\overline{\beta}_j| \\ &= \left(\sum_i |\alpha_i||\beta_i|\right)^2 \\ &\leq \sum_i |\alpha_i|^2 \sum_i |\beta_i|^2 = \langle u|u\rangle\langle v|v\rangle = 1. \end{aligned}$$

From (3.8) and Proposition 3.3, it follows that $S(\rho\|\sigma) \leq S(\rho\|\tau)$ for all arbitrary bipartite pure separable states $\tau$. Any arbitrary separable state $\tau$ can be written as a convex combination of separable pure states

$$\tau = \sum_x t_x \tau_x$$

where each $\tau_x$ is pure and $t_x \geq 0$ and $\sum_x t_x = 1$. By convexity of the relative entropy, it follows that

$$S(\rho\|\tau) = S\left(\rho \,\middle\|\, \sum_x t_x \tau_x\right) \leq \sum_x t_x S(\rho\|\tau_x) \leq S(\rho\|\sigma),$$

and thus $E_R(\rho) = \min_{\tau \in \mathfrak{D}} S(\rho\|\tau) = S(\rho\|\sigma)$.

Finally, computing the relative entropy of $\rho$ and $\sigma$, we find

$$S(\rho\|\sigma) = -\langle\psi|\log\sigma|\psi\rangle = -\sum_i p_i \log p_i = S(\rho_{\mathsf{A}}),$$

as desired.

(d) Recall that $F$ is monotonic under CPTP maps, i.e. for all CPTP maps $\mathcal{E}$ and all density operators $\rho$ and $\sigma$ it holds that $F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq F(\rho, \sigma)$ and thus

$$\tilde{D}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) = 1 - [F(\mathcal{E}(\rho), \mathcal{E}(\sigma))]^2 \leq 1 - [F(\rho, \sigma)]^2 = \tilde{D}(\rho\|\sigma). \tag{3.9}$$

Let $\rho$ be an arbitrary bipartite density operator and let $\mathcal{E}$ be a LOCC channel. As in part (b), we have

$$\begin{aligned} \tau_R(\rho) = \min_{\sigma \in \mathfrak{D}_{\mathsf{AB}}} \tilde{D}(\rho\|\sigma) &\geq \min_{\sigma \in \mathfrak{D}_{\mathsf{AB}}} \tilde{D}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \\ &= \min_{\tau \in \mathcal{E}(\mathfrak{D}_{\mathsf{AB}})} \tilde{D}(\mathcal{E}(\rho)\|\tau) \\ &\geq \min_{\tau \in \mathfrak{D}_{\mathsf{A'B'}}} \tilde{D}(\mathcal{E}(\rho)\|\tau) = \tau_R(\mathcal{E}(\rho)), \end{aligned}$$

and thus $\tau_R(\rho) \geq \tau_R(\mathcal{E}(\rho))$. Hence $\tau_R$ is an entanglement monotone.

Recall that the fidelity of a pure state $\rho = |\psi\rangle\langle\psi|$ with an arbitrary state $\sigma$ reduces to

$$F(\rho, \sigma) = F(|\psi\rangle\langle\psi|, \sigma) = \sqrt{\langle\psi|\sigma|\psi\rangle}.$$

**Proposition 3.5.** *Let $\rho = |\psi\rangle\langle\psi|$ be a bipartite pure state. Then $\tau_R(\rho) = 1 - p_{\max}$, where $p_{\max}$ is the largest Schmidt coefficient of $|\psi\rangle$.*

*Proof.* Let $\rho = |\psi\rangle\langle\psi|$ be an arbitrary pure state. Without loss of generality, we may consider $|\psi\rangle$ is Schmidt form as

$$|\psi\rangle = \sum_i \sqrt{p_i}|ii\rangle,$$

where the Schmidt coefficients are in decreasing order $p_1 \geq p_2 \geq \cdots$. Define $\sigma = |ii\rangle\langle ii|$. We will show that $[F(\rho,\sigma)]^2 \geq [F(\rho,\tau)]^2$ for all separable states $\tau$. Note that $[F(\rho,\sigma)]^2 = |\langle\psi|11\rangle|^2 = p_1 = p_{\max}$. If $\tau$ is pure, we can write it as $\tau = |u\rangle\langle u| \otimes |v\rangle\langle v|$ for some pure states

$$|u\rangle = \sum_i \alpha_i|i\rangle \qquad \text{and} \qquad |v\rangle = \sum_i \beta_i|i\rangle.$$

Then

$$[F(\rho,\tau)]^2 = \langle\psi|\tau|\psi\rangle = \left|\sum_i \sqrt{p_i}\alpha_i\beta_i\right|^2 \leq p_{\max}\left|\sum_i \alpha_i\beta_i\right|^2 \leq p_{\max}\underbrace{\sum_i|\alpha_i|^2}_{\langle u|u\rangle=1}\underbrace{\sum_i|\beta_i|^2}_{\langle v|v\rangle=1} = p_{\max}.$$

If $\tau$ is not pure, it can be written as a convex combination of separable pure states

$$\tau = \sum_x t_x\tau_x$$

where each $\tau_x$ is pure and $t_x \geq 0$ and $\sum_x t_x = 1$. Hence

$$[F(\rho,\tau)]^2 = \langle\psi|\tau|\psi\rangle = \sum_x t_x\langle\psi|\tau_x|\psi\rangle \leq p_{\max}$$

and thus $1 - [F(\rho,\sigma)]^2 = 1 - p_{\max} \leq 1 - [F(\rho,\tau)]^2$ holds for all separable states $\tau$, as desired. $\square$

## Problem 4 (Strong sub-additivity of the Shannon entropy)

**Problem 4.** Prove the strong sub-additivity of the Shannon entropy. That is, prove that

$$H(X\colon Z|Y) \geq 0 \tag{4.1}$$

with equality if and only if $Z \to Y \to X$ forms a Markov chain.

**Solution.** *Proof.* From the last homework assignment, we used the fact that

$$H(X\colon Z|Y) = \sum_y p(y) H(X|_{Y=y}\colon Z|_{Y=y})$$

to prove that $H(X\colon Z|Y) \geq 0$. We now show the conditions for equality in (4.1). Since $p(y) \geq 0$ and $H(X|_{Y=y}\colon Z|_{Y=y}) \geq 0$ for all $y$, it suffices to show that

$$H(X|_{Y=y}\colon Z|_{Y=y}) = 0$$

for all $y$ if and only if $Z \to Y \to X$ forms a Markov chain. Since $H(X|_{Y=y}\colon Z|_{Y=y}) = 0$ if and only if $X|_{Y=y}$ and $Z|_{Y=y}$ independent random variables, we need to show that $p(xz|y) = p(x|y)p(z|y)$ for all $y$ if and only if $Z \to Y \to X$ forms a Markov chain. This is indeed true, since

$$p(xz|y) = p(x|y)p(z|y) \quad \Leftrightarrow \quad \frac{p(xyz)}{p(y)} = \frac{p(xy)}{p(y)}\frac{p(yz)}{p(y)} \quad \Leftrightarrow \quad \frac{p(xyz)}{p(yz)} = \frac{p(xy)}{p(y)} \quad \Leftrightarrow \quad p(x|yz) = p(x|y),$$

as desired. □