CS 766/QIC 820 Theory of Quantum Information (Fall 2011)

Lecture 11: Strong subadditivity of von Neumann entropy

In this lecture we will prove a fundamental fact about the von Neumann entropy, known as *strong subadditivity*. Let us begin with a precise statement of this fact.

Theorem 11.1 (Strong subadditivity of von Neumann entropy)**.** *Let* X, Y, *and* Z *be registers. For every state* $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$ *of these registers it holds that*

$$S(\mathsf{X},\mathsf{Y},\mathsf{Z}) + S(\mathsf{Z}) \leq S(\mathsf{X},\mathsf{Z}) + S(\mathsf{Y},\mathsf{Z}).$$

There are multiple known ways to prove this theorem. The approach we will take is to first establish a property of the quantum relative entropy, known as *joint convexity*. Once we establish this property, it will be straightforward to prove strong subadditivity.

11.1 Joint convexity of the quantum relative entropy

We will now prove that the quantum relative entropy is jointly convex, as is stated by the following theorem.

Theorem 11.2 (Joint convexity of the quantum relative entropy)**.** *Let* $\mathcal{X}$ *be a complex Euclidean space, let* $\rho_0, \rho_1, \sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X})$ *be positive definite density operators, and let* $\lambda \in [0,1]$*. It holds that*

$$S(\lambda\rho_0 + (1-\lambda)\rho_1 \| \lambda\sigma_0 + (1-\lambda)\sigma_1) \leq \lambda\, S(\rho_0\|\sigma_0) + (1-\lambda)\, S(\rho_1\|\sigma_1).$$

The proof of Theorem 11.2 that we will study is fairly standard and has the nice property of being elementary. It is, however, relatively complicated, so we will need to break it up into a few pieces.

Before considering the proof, let us note that the theorem remains true if we allow ρ_0, ρ_1, σ_0, and σ_1 to be arbitrary density operators, provided we allow the quantum relative entropy to take infinite values as we discussed in the previous lecture. Supposing that we do this, we see that if either $S(\rho_0\|\sigma_0)$ or $S(\rho_1\|\sigma_1)$ is infinite, there is nothing to prove. If it is the case that $S(\lambda\rho_0 + (1-\lambda)\rho_1\|\lambda\sigma_0 + (1-\lambda)\sigma_1)$ is infinite, then either $S(\rho_0\|\sigma_0)$ or $S(\rho_1\|\sigma_1)$ is infinite as well: if $\lambda \in (0,1)$, then $\ker(\lambda\rho_0 + (1-\lambda)\rho_1) = \ker(\rho_0) \cap \ker(\rho_1)$ and $\ker(\lambda\sigma_0 + (1-\lambda)\sigma_1) = \ker(\sigma_0) \cap \ker(\sigma_1)$, owing to the fact that ρ_0, ρ_1, σ_0, and σ_1 are all positive semidefinite; and so

$$\ker(\lambda\sigma_0 + (1-\lambda)\sigma_1) \not\subseteq \ker(\lambda\rho_0 + (1-\lambda)\rho_1)$$

implies $\ker(\sigma_0) \not\subseteq \ker(\rho_0)$ or $\ker(\sigma_1) \not\subseteq \ker(\rho_1)$ (or both). In the remaining case, which is that $S(\lambda\rho_0 + (1-\lambda)\rho_1\|\lambda\sigma_0 + (1-\lambda)\sigma_1)$, $S(\rho_0\|\sigma_0)$, and $S(\rho_1\|\sigma_1)$ are all finite, a fairly straightforward continuity argument will establish the inequality from the one stated in the theorem.

Now, to prove the theorem, the first step is to consider a real-valued function $f_{\rho,\sigma} : \mathbb{R} \to \mathbb{R}$ defined as

$$f_{\rho,\sigma}(\alpha) = \operatorname{Tr}\left(\sigma^{\alpha}\rho^{1-\alpha}\right)$$

for all $\alpha \in \mathbb{R}$, for any fixed choice of positive definite density operators $\rho, \sigma \in \mathrm{D}(\mathcal{X})$. Under the assumption that ρ and σ are both positive definite, we have that the function $f_{\rho,\sigma}$ is well defined, and is in fact differentiable (and therefore continuous) everywhere on its domain. In particular, we have

$$f'_{\rho,\sigma}(\alpha) = \operatorname{Tr}\left[\sigma^{\alpha}\rho^{1-\alpha}\left(\ln(\sigma) - \ln(\rho)\right)\right]. \tag{11.1}$$

To verify that this expression is correct, we may consider spectral decompositions

$$\rho = \sum_{i=1}^{n} p_i x_i x_i^* \quad \text{and} \quad \sigma = \sum_{i=1}^{n} q_i y_i y_i^*.$$

We have

$$\operatorname{Tr}\left(\sigma^{\alpha}\rho^{1-\alpha}\right) = \sum_{1\leq i,j\leq n} q_j^{\alpha} p_i^{1-\alpha} \left|\langle x_i, y_j\rangle\right|^2$$

and so

$$f'_{\rho,\sigma}(\alpha) = \sum_{1\leq i,j\leq n} (\ln(q_j) - \ln(p_i)) q_j^{\alpha} p_i^{1-\alpha} \left|\langle x_i, y_j\rangle\right|^2 = \operatorname{Tr}\left[\sigma^{\alpha}\rho^{1-\alpha}\left(\ln(\sigma) - \ln(\rho)\right)\right]$$

as claimed.

The main reason we are interested in the function $f_{\rho,\sigma}$ is that its derivative has an interesting value at 0:

$$f'_{\rho,\sigma}(0) = -\ln(2) S(\rho\|\sigma).$$

We may therefore write

$$S(\rho\|\sigma) = -\frac{1}{\ln(2)} f'_{\rho,\sigma}(0) = -\frac{1}{\ln(2)} \lim_{\alpha\to 0^+} \frac{\operatorname{Tr}\left(\sigma^{\alpha}\rho^{1-\alpha}\right) - 1}{\alpha},$$

where the second equality follows by substituting $f_{\rho,\sigma}(0) = 1$ into the definition of the derivative.

Now consider the following theorem that concerns the relationship among the functions $f_{\rho,\sigma}$ for various choices of ρ and σ.

Theorem 11.3. *Let $\sigma_0, \sigma_1, \rho_0, \rho_1 \in \mathrm{Pd}(\mathcal{X})$ be positive definite operators. For every choice of $\alpha, \lambda \in [0,1]$ we have*

$$\operatorname{Tr}\left((\lambda\sigma_0 + (1-\lambda)\sigma_1)^{\alpha}(\lambda\rho_0 + (1-\lambda)\rho_1)^{1-\alpha}\right) \geq \lambda \operatorname{Tr}\left(\sigma_0^{\alpha}\rho_0^{1-\alpha}\right) + (1-\lambda)\operatorname{Tr}\left(\sigma_1^{\alpha}\rho_1^{1-\alpha}\right).$$

(The theorem happens to be true for all positive definite operators ρ_0, ρ_1, σ_0, and σ_1, but we will really only need it for density operators.)

Before proving this theorem, let us note that it implies Theorem 11.2.

Proof of Theorem 11.2 (assuming Theorem 11.3). We have

$$\begin{aligned}
&S(\lambda\rho_0 + (1-\lambda)\rho_1 \| \lambda\sigma_0 + (1-\lambda)\sigma_1)\\
&\qquad = -\frac{1}{\ln(2)} \lim_{\alpha\to 0^+} \frac{\operatorname{Tr}\left((\lambda\sigma_0 + (1-\lambda)\sigma_1)^{\alpha}(\lambda\rho_0 + (1-\lambda)\rho_1)^{1-\alpha}\right) - 1}{\alpha}\\
&\qquad \leq -\frac{1}{\ln(2)} \lim_{\alpha\to 0^+} \frac{\lambda \operatorname{Tr}\left(\sigma_0^{\alpha}\rho_0^{1-\alpha}\right) + (1-\lambda)\operatorname{Tr}\left(\sigma_1^{\alpha}\rho_1^{1-\alpha}\right) - 1}{\alpha}\\
&\qquad = -\frac{1}{\ln(2)} \lim_{\alpha\to 0^+} \left[\lambda\left(\frac{\operatorname{Tr}\left(\sigma_0^{\alpha}\rho_0^{1-\alpha}\right) - 1}{\alpha}\right) + (1-\lambda)\left(\frac{\operatorname{Tr}\left(\sigma_1^{\alpha}\rho_1^{1-\alpha}\right) - 1}{\alpha}\right)\right]\\
&\qquad = \lambda\, S(\rho_0\|\sigma_0) + (1-\lambda)\, S(\rho_1\|\sigma_1)
\end{aligned}$$

as required. □

Our goal has therefore shifted to proving Theorem 11.3. To prove Theorem 11.3 we require another fact that is stated in the theorem that follows. It is equivalent to a theorem known as *Lieb's concavity theorem*, and Theorem 11.3 is a special case of that theorem, but Lieb's concavity theorem itself is usually stated in a somewhat different form than the one that follows.

Theorem 11.4. *Let $A_0, A_1 \in \operatorname{Pd}(\mathcal{X})$ and $B_0, B_1 \in \operatorname{Pd}(\mathcal{Y})$ be positive definite operators. For every choice of $\alpha \in [0,1]$ we have*

$$(A_0 + A_1)^{\alpha} \otimes (B_0 + B_1)^{1-\alpha} \geq A_0^{\alpha} \otimes B_0^{1-\alpha} + A_1^{\alpha} \otimes B_1^{1-\alpha}.$$

Once again, before proving this theorem, let us note that it implies the main result we are working toward.

Proof of Theorem 11.3 (assuming Theorem 11.4). The substitutions

$$A_0 = \lambda\sigma_0, \quad B_0 = \lambda\rho_0^{\mathsf{T}}, \quad A_1 = (1-\lambda)\sigma_1, \quad B_1 = (1-\lambda)\rho_1^{\mathsf{T}},$$

taken in Theorem 11.4 imply the operator inequality

$$\begin{aligned}
(\lambda\sigma_0 + (1-\lambda)\sigma_1)^{\alpha} &\otimes (\lambda\rho_0^{\mathsf{T}} + (1-\lambda)\rho_1^{\mathsf{T}})^{1-\alpha} \\
&\geq \lambda\, \sigma^{\alpha} \otimes (\rho_0^{\mathsf{T}})^{1-\alpha} + (1-\lambda)\, \sigma_1^{\alpha} \otimes (\rho_1^{\mathsf{T}})^{1-\alpha} \\
&= \lambda\, \sigma^{\alpha} \otimes (\rho_0^{1-\alpha})^{\mathsf{T}} + (1-\lambda)\, \sigma_1^{\alpha} \otimes (\rho_1^{1-\alpha})^{\mathsf{T}}.
\end{aligned}$$

Applying the identity $\operatorname{vec}(\mathbb{1})^*(X \otimes Y^{\mathsf{T}})\operatorname{vec}(\mathbb{1}) = \operatorname{Tr}(XY)$ to both sides of the inequality then gives the desired result. □

Now, toward the proof of Theorem 11.4, we require the following lemma.

Lemma 11.5. *Let $P_0, P_1, Q_0, Q_1, R_0, R_1 \in \operatorname{Pd}(\mathcal{X})$ be positive definite operators that satisfy these conditions:*

1. $[P_0, P_1] = [Q_0, Q_1] = [R_0, R_1] = 0,$
2. $P_0^2 \geq Q_0^2 + R_0^2$, *and*
3. $P_1^2 \geq Q_1^2 + R_1^2$.

It holds that $P_0 P_1 \geq Q_0 Q_1 + R_0 R_1$.

Remark. Notice that in the conclusion of the lemma, $P_0 P_1$ is positive definite given the assumption that $[P_0, P_1] = 0$, and likewise for $Q_0 Q_1$ and $R_0 R_1$.

Proof. The conclusion of the lemma is equivalent to $X \leq \mathbb{1}$ for

$$X = P_0^{-1/2} P_1^{-1/2} (Q_0 Q_1 + R_0 R_1) P_1^{-1/2} P_0^{-1/2}.$$

As X is positive definite, and therefore Hermitian, this in turn is equivalent to the condition that every eigenvalue of X is at most 1.

To establish that every eigenvalue of X is at most 1, let us suppose that u is any eigenvector of X whose corresponding eigenvalue is λ. As P_0 and P_1 are invertible and u is nonzero, we have that $P_0^{-1/2}P_1^{1/2}u$ is nonzero as well, and therefore we may define a unit vector v as follows:

$$v = \frac{P_0^{-1/2}P_1^{1/2}u}{\left\| P_0^{-1/2}P_1^{1/2}u \right\|}.$$

It holds that v is an eigenvector of $P_0^{-1}(Q_0Q_1 + R_0R_1)P_1^{-1}$ with eigenvalue λ, and because v is a unit vector it follows that

$$v^* P_0^{-1}(Q_0Q_1 + R_0R_1)P_1^{-1}v = \lambda.$$

Finally, using the fact that v is a unit vector, we can establish the required bound on λ as follows:

$$\begin{aligned}
\lambda &= v^* P_0^{-1}(Q_0Q_1 + R_0R_1)P_1^{-1}v \\
&\leq \left| v^* P_0^{-1} Q_0 Q_1 P_1^{-1} v \right| + \left| v^* P_0^{-1} R_0 R_1 P_1^{-1} v \right| \\
&\leq \sqrt{v^* P_0^{-1} Q_0^2 P_0^{-1} v}\sqrt{v^* P_1^{-1} Q_1^2 P_1^{-1} v} + \sqrt{v^* P_0^{-1} R_0^2 P_0^{-1} v}\sqrt{v^* P_1^{-1} R_1^2 P_1^{-1} v} \\
&\leq \sqrt{v^* P_0^{-1}(Q_0^2 + R_0^2) P_0^{-1} v}\sqrt{v^* P_1^{-1}(Q_1^2 + R_1^2) P_1^{-1} v} \\
&\leq 1.
\end{aligned}$$

Here we have used the triangle inequality once and the Cauchy-Schwarz inequality twice, along with the given assumptions on the operators. □

Finally, we can finish of the proof of Theorem 11.2 by proving Theorem 11.4.

Proof of Theorem 11.4. Let us define a function $f : [0,1] \rightarrow \text{Herm}(\mathcal{X} \otimes \mathcal{Y})$ as

$$f(\alpha) = (A_0 + A_1)^{\alpha} \otimes (B_0 + B_1)^{1-\alpha} - \left(A_0^{\alpha} \otimes B_0^{1-\alpha} + A_1^{\alpha} \otimes B_1^{1-\alpha} \right),$$

and let $K = \{\alpha \in [0,1] : f(\alpha) \in \text{Pos}(\mathcal{X} \otimes \mathcal{Y})\}$ be the pre-image under f of the set $\text{Pos}(\mathcal{X} \otimes \mathcal{Y})$. Notice that K is a closed set, given that f is continuous and $\text{Pos}(\mathcal{X} \otimes \mathcal{Y})$ is closed. Our goal is to prove that $K = [0,1]$.

It is obvious that 0 and 1 are elements of K. For an arbitrary choice of $\alpha, \beta \in K$, consider the following operators:

$$\begin{aligned}
P_0 &= (A_0 + A_1)^{\alpha/2} \otimes (B_0 + B_1)^{(1-\alpha)/2}, \\
P_1 &= (A_0 + A_1)^{\beta/2} \otimes (B_0 + B_1)^{(1-\beta)/2}, \\
Q_0 &= A_0^{\alpha/2} \otimes B_0^{(1-\alpha)/2}, \\
Q_1 &= A_0^{\beta/2} \otimes B_0^{(1-\beta)/2}, \\
R_0 &= A_1^{\alpha/2} \otimes B_1^{(1-\alpha)/2}, \\
R_1 &= A_1^{\beta/2} \otimes B_1^{(1-\beta)/2}.
\end{aligned}$$

The conditions $[P_0, P_1] = [Q_0, Q_1] = [R_0, R_1] = 0$ are immediate, while the assumptions that $\alpha \in K$ and $\beta \in K$ correspond to the conditions $P_0^2 \geq Q_0^2 + R_0^2$ and $P_1^2 \geq Q_1^2 + R_1^2$, respectively. We may therefore apply Lemma 11.5 to obtain

$$(A_0 + A_1)^\gamma \otimes (B_0 + B_1)^{1-\gamma} \geq A_0^\gamma \otimes B_0^{1-\gamma} + A_1^\gamma \otimes B_1^{1-\gamma}$$

for $\gamma = (\alpha + \beta)/2$, which implies that $(\alpha + \beta)/2 \in K$.

Now, given that $0 \in K$, $1 \in K$, and $(\alpha + \beta)/2 \in K$ for any choice of $\alpha, \beta \in K$, we have that K is dense in $[0, 1]$. In particular, K contains every number of the form $m/2^n$ for n and m nonnegative integers with $m \leq 2^n$. As K is closed, this implies that $K = [0, 1]$ as required. □

11.2 Strong subadditivity

We have worked hard to prove that the quantum relative entropy is jointly convex, and now it is time to reap the rewards. Let us begin by proving the following simple theorem, which states that mixed unitary channels cannot increase the relative entropy of two density operators.

Theorem 11.6. *Let $\mathcal{X}$ be a complex Euclidean space and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a mixed unitary channel. For any choice of positive definite density operators $\rho, \sigma \in \mathrm{D}(\mathcal{X})$ we have*

$$S(\Phi(\rho) \| \Phi(\sigma)) \leq S(\rho \| \sigma).$$

Proof. As Φ is mixed unitary, we may write

$$\Phi(X) = \sum_{j=1}^{m} p_j U_j X U_j^*$$

for a probability vector $(p_1, \ldots, p_m)$ and unitary operators $U_1, \ldots, U_m \in \mathrm{U}(\mathcal{X})$. By Theorem 11.2 we have

$$S(\Phi(\rho) \| \Phi(\sigma)) = S\left(\sum_{j=1}^{m} p_j U_j \rho U_j^* \,\middle\|\, \sum_{j=1}^{m} p_j U_j \sigma U_j^* \right) \leq \sum_{j=1}^{m} p_j S\left(U_j \rho U_j^* \,\middle\|\, U_j \sigma U_j^* \right).$$

The quantum relative entropy is clearly unitarily invariant, meaning

$$S(\rho \| \sigma) = S(U \rho U^* \| U \sigma U^*)$$

for all $U \in \mathrm{U}(\mathcal{X})$. This implies that

$$\sum_{j=1}^{m} p_j S\left(U_j \rho U_j^* \,\middle\|\, U_j \sigma U_j^* \right) = S(\rho \| \sigma),$$

and therefore completes the proof. □

Notice that for any choice of positive definite density operators $\rho_0, \rho_1, \sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X})$ we have

$$S(\rho_0 \otimes \rho_1 \| \sigma_0 \otimes \sigma_1) = S(\rho_0 \| \sigma_0) + S(\rho_1 \| \sigma_1).$$

This fact follows easily from the identity $\log(P \otimes Q) = \log(P) \otimes \mathbb{1} + \mathbb{1} \otimes \log(Q)$, which is valid for all $P, Q \in \mathrm{Pd}(\mathcal{X})$. Combining this observation with the previous theorem yields the following corollary.

Corollary 11.7. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces. For any choice of positive definite density operators $\rho, \sigma \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ it holds that*

$$S(\operatorname{Tr}_{\mathcal{Y}}(\rho) \| \operatorname{Tr}_{\mathcal{Y}}(\sigma)) \leq S(\rho \| \sigma).$$

Proof. The completely depolarizing operation $\Omega \in \mathrm{C}(\mathcal{Y})$ is mixed unitary, as we proved in Lecture 6, which implies that $\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Omega$ is mixed unitary as well. For every $\xi \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ we have

$$(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Omega)(\xi) = \operatorname{Tr}_{\mathcal{Y}}(\xi) \otimes \frac{\mathbb{1}_{\mathcal{Y}}}{m}$$

where $m = \dim(\mathcal{Y})$, and therefore

$$\begin{aligned} S(\operatorname{Tr}_{\mathcal{Y}}(\rho) \| \operatorname{Tr}_{\mathcal{Y}}(\sigma)) &= S\left(\operatorname{Tr}_{\mathcal{Y}}(\rho) \otimes \frac{\mathbb{1}_{\mathcal{Y}}}{m} \,\middle\|\, \operatorname{Tr}_{\mathcal{Y}}(\sigma) \otimes \frac{\mathbb{1}_{\mathcal{Y}}}{m}\right) \\ &= S\left((\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Omega)(\rho) \,\|\, (\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Omega)(\sigma)\right) \\ &\leq S(\rho \| \sigma) \end{aligned}$$

as required. □

Note that the above theorem and corollary extend to arbitrary density operators given that the same is true of Theorem 11.2. Making use of the Stinespring representation of quantum channels, we obtain the following fact.

Corollary 11.8. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\rho, \sigma \in \mathrm{D}(\mathcal{X})$ be density operators, and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It holds that*

$$S(\Phi(\rho) \| \Phi(\sigma)) \leq S(\rho \| \sigma).$$

Finally we are prepared to prove strong subadditivity, which turns out to be very easy now that we have established Corollary 11.7.

Proof of Theorem 11.1. We need to prove that the inequality

$$S(\rho^{\mathsf{XYZ}}) + S(\rho^{\mathsf{Z}}) \leq S(\rho^{\mathsf{XZ}}) + S(\rho^{\mathsf{YZ}})$$

holds for all choices of $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$. It suffices to prove this inequality for all positive definite ρ, as it then follows for arbitrary density operators ρ by the continuity of the von Neumann entropy.

Let $n = \dim(\mathcal{X})$, and observe that the following two identities hold: the first is

$$S\left(\rho^{\mathsf{XYZ}} \,\middle\|\, \frac{\mathbb{1}_{\mathcal{X}}}{n} \otimes \rho^{\mathsf{YZ}}\right) = -S(\rho^{\mathsf{XYZ}}) + S(\rho^{\mathsf{YZ}}) + \log(n),$$

and the second is

$$S\left(\rho^{\mathsf{XZ}} \,\middle\|\, \frac{\mathbb{1}_{\mathcal{X}}}{n} \otimes \rho^{\mathsf{Z}}\right) = -S(\rho^{\mathsf{XZ}}) + S(\rho^{\mathsf{Z}}) + \log(n).$$

By Corollary 11.7 we have

$$S\left(\rho^{\mathsf{XZ}} \,\middle\|\, \frac{\mathbb{1}_{\mathcal{X}}}{n} \otimes \rho^{\mathsf{Z}}\right) \leq S\left(\rho^{\mathsf{XYZ}} \,\middle\|\, \frac{\mathbb{1}_{\mathcal{X}}}{n} \otimes \rho^{\mathsf{YZ}}\right),$$

and therefore

$$S(\rho^{\mathsf{XYZ}}) + S(\rho^{\mathsf{Z}}) \leq S(\rho^{\mathsf{XZ}}) + S(\rho^{\mathsf{YZ}})$$

as required. □

To conclude the lecture, let us prove a statement about quantum mutual information that is equivalent to strong subadditivity.

Corollary 11.9. *Let* X, Y, *and* Z *be registers. For every state* $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$ *of these registers we have*

$$S(\mathsf{X} : \mathsf{Y}) \leq S(\mathsf{X} : \mathsf{Y}, \mathsf{Z}).$$

Proof. By strong subadditivity we have

$$S(\mathsf{X}, \mathsf{Y}, \mathsf{Z}) + S(\mathsf{Y}) \leq S(\mathsf{X}, \mathsf{Y}) + S(\mathsf{Y}, \mathsf{Z}),$$

which is equivalent to

$$S(\mathsf{Y}) - S(\mathsf{X}, \mathsf{Y}) \leq S(\mathsf{Y}, \mathsf{Z}) - S(\mathsf{X}, \mathsf{Y}, \mathsf{Z}).$$

Adding $S(\mathsf{X})$ to both sides gives

$$S(\mathsf{X}) + S(\mathsf{Y}) - S(\mathsf{X}, \mathsf{Y}) \leq S(\mathsf{X}) + S(\mathsf{Y}, \mathsf{Z}) - S(\mathsf{X}, \mathsf{Y}, \mathsf{Z}).$$

This inequality is equivalent to

$$S(\mathsf{X} : \mathsf{Y}) \leq S(\mathsf{X} : \mathsf{Y}, \mathsf{Z}),$$

which establishes the claim. □