11 Entropy and information

Exercises: 11.1, 11.2, 11.3, 11.4, 11.5, 11.6, 11.7, 11.8, 11.9, 11.10, 11.11, 11.12, 11.13, 11.14, 11.15, 11.16, 11.17, 11.18, 11.19, 11.20, 11.21, 11.22, 11.23, 11.24, 11.25, 11.26.

11.1

Fair coin:

$$H(X) = 2\left(-\frac{1}{2}\log\frac{1}{2}\right) = 1.$$

Fair die:

$$H(X) = 6\left(-\frac{1}{6}\log\frac{1}{6}\right) = \log 6 \approx 2.585.$$

If they were unfair, the entropy would be smaller. Since there would be at least one outcome more probable than another, the average information gain after each toss would be smaller.

11.2

From property (1) we know that p and q are real values in the range $[0,1]$, so there are always numbers a and b in the range $(-\infty, 0]$ such that $p = 2^a$ and $q = 2^b$. So we have $I(pq) = I(2^a 2^b) = I(2^{a+b})$, thus from property (3) we obtain that

$$I(2^{a+b}) = I(2^a) + I(2^b).$$

If we define the function $f(x) \equiv I(2^x)$, we get $f(a+b) = f(a) + f(b)$. Because of property (2), we know that f must be continuous and smooth, and therefore, has a Taylor series form. For $a = b = 0$ we get $f(0) = 0$, and for $b = -a$ we get $f(-a) = -f(a)$, so it is clearly an odd function, hence

$$f(x) = \sum_{j=0}^{\infty} k_j x^{2j+1}.$$

Considering $x = a + b$ we obtain

$$f(a+b) = \sum_{j=0}^{\infty} k_j (a+b)^{2j+1} = k_0(a+b) + \sum_{j=1}^{\infty} k_j \sum_{n=0}^{2j+1} \binom{2j+1}{n} a^n b^{2j+1-n},$$

but using the property that f must satisfy, we get

$$f(a) + f(b) = \sum_{j=0}^{\infty} k_j (a^{2j+1} + b^{2j+1}) = k_0(a+b) + \sum_{j=1}^{\infty} k_j (a^{2j+1} + b^{2j+1}).$$

Therefore, the only way for the property to be true is if $k_j = 0$ for all $j > 0$, meaning $f(x) = kx$ for some constant k. Now, since $f(x) = I(2^x)$, it is also true that $I(x) = f(\log x)$, and this gives us

$$I(p) = k \log p.$$

The average of this function over the values of the set $\{p_1, \cdots, p_n\}$ is precisely the Shannon entropy up to a multiplicative constant k

$$H(X) \equiv \langle I(p) \rangle = \sum_{i=1}^{n} p_i I(p_i) = k \sum_{j=1}^{n} p_i \log p_i.$$

11.3

$$\begin{aligned}
\frac{\mathrm{d}H_{\mathrm{bin}}}{\mathrm{d}p} &= -\frac{\mathrm{d}}{\mathrm{d}p}\left[p\log(p)\right] - \frac{\mathrm{d}}{\mathrm{d}p}\left[(1-p)\log(1-p)\right] \\
&= -\log(p) - 1 + \log(1-p) + 1 \\
&= \log(1-p) - \log(p).
\end{aligned}$$

Considering $p \in [0,1]$, $\mathrm{d}H_{\mathrm{bin}}/\mathrm{d}p = 0$ only for $p = 1/2$. The second derivative yields

$$\frac{\mathrm{d}^2 H_{\mathrm{bin}}}{\mathrm{d}p^2} = -\frac{1}{1-p} - \frac{1}{p} = -\frac{1}{p(1-p)} \quad \Longrightarrow \quad \frac{\mathrm{d}^2 H_{\mathrm{bin}}}{\mathrm{d}p^2} < 0 \quad \forall\, p \in (0,1),$$

corresponding to constant negative concavity, so $p = 1/2$ corresponds to a maximum.

11.4

The second derivative of the binary entropy is negative for all $p \in (0,1)$ (see Exercise 11.3). It follows that for all, $p, x_1, x_2 \in [0,1]$, it holds

$$H_{\mathrm{bin}}(px_1 + (1-p)x_2) \geq pH_{\mathrm{bin}}(x_1) + (1-p)H_{\mathrm{bin}}(x_2),$$

with the innequality being strict for all values $p \in (0,1)$. Since $H_{\mathrm{bin}}(0) = H_{\mathrm{bin}}(1) \equiv 0$, we get equality in the trivial cases: $x_1 = x_2$, or $p = 0$, or $p = 1$.

11.5

$$\begin{aligned}
H(p(x,y)||p(x)p(y)) &= \sum_{x,y} p(x,y) \log\left[\frac{p(x,y)}{p(x)p(y)}\right] \\
&= \sum_{x,y} p(x,y)\log[p(x,y)] - \sum_{x,y} p(x,y)\log[p(x)p(y)] \\
&= \sum_{x,y} p(x,y)\log[p(x,y)] - \sum_{x} p(x)\log[p(x)] - \sum_{y} p(y)\log[p(y)] \\
&= H(p(x)) + H(p(y)) - H(p(x,y)).
\end{aligned}$$

Considering that x are the possible outcomes of the random variable X, and y the possible outcomes of the random variable Y, we have $H(p(x)) \equiv H(X)$, $H(p(y)) \equiv H(Y)$, and $H(p(x,y)) \equiv H(X,Y)$. From the non-negativity of the relative entropy, we get $H(X,Y) \leq H(X) + H(Y)$. If we consider equality, then $H(p(x,y)||p(x)p(y)) = 0$, which means $p(x,y) = p(x)p(y)$, hence X and Y are independent random variables. The converse is immediate.

11.6

Let us use the probability distribution given by $p(x|y)p(z|y)p(y)$. The relative entropy between this distribution and $p(x,y,z)$ will be

$$H(p(x,y,z)||p(x|y)p(z|y)p(y)) = \sum_{x,y,z} p(x,y,z) \log \left[\frac{p(x,y,z)}{p(x|y)p(z|y)p(y)} \right].$$

From Bayes' rule we can write $p(x|y) = p(x,y)/p(y)$, and $p(z|y) = p(y,z)/p(y)$, hence

$$\begin{aligned} H(p(x,y,z)||p(x|y)p(z|y)p(y)) &= \sum_{x,y,z} p(x,y,z) \log \left[\frac{p(x,y,z)p(y)}{p(x,y)p(y,z)} \right] \\ &= \sum_{x,y,z} p(x,y,z) \log[p(x,y,z)] + \sum_{y} p(y) \log[p(y)] \\ &\quad - \sum_{x,y} p(x,y) \log[p(x,y)] - \sum_{y,z} p(y,z) \log[p(y,z)] \\ &= -H(X,Y,Z) - H(Y) + H(X,Y) + H(Y,Z). \end{aligned}$$

Using the fact that $H(p(x,y,z)||p(x|y)p(z|y)p(y)) \geq 0$ we get strong subadditivity

$$H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z).$$

If we consider equality, then $H(p(x,y,z)||p(x|y)p(z|y)p(y)) = 0$, which means (see Exercise **??**)

$$p(x,y,z) = p(x|y)p(z|y)p(y) = p(z)p(y|z)p(x|y),$$

hence $Z \to Y \to X$ forms a Markov chain. The converse is immediate.

11.7

Let n_Y be the number of possible outcomes for the variable Y. The uniform distribution $u(y)$ over Y is then given by $u(y) = 1/n_Y$ for all y. The relative entropy $H(p(x,y)||p(x)u(y))$ yields

$$\begin{aligned} H(p(x,y)||p(x)u(y)) &= \sum_{x,y} p(x,y) \log \left[\frac{p(x,y)}{p(x)\frac{1}{n_Y}} \right] \\ &= \sum_{x,y} p(x,y) \log[p(x,y)] - \sum_{x} p(x) \log[p(x)] - \log \tfrac{1}{n_Y} \\ &= -H(X,Y) + H(X) + \log n_Y. \end{aligned}$$

By definition, we have $H(Y|X) = H(X,Y) - H(X)$, thus

$$H(Y|X) = \log n_Y - H(p(x,y)||p(x)u(y)).$$

Notice that $\log n_Y$ corresponds to the Shannon entropy of Y when it is completely independent of X and its probability distribution is the uniform distribution $u(y)$, which is where it has its maximum value, thus $0 \leq H(p(x,y)||p(x)u(y)) \leq \log n_Y$ and therefore $H(Y|X) \geq 0$. Equality will hold when $p(x,y)$ is the farthest possible from $p(x)u(y)$, which is when Y is given by a deterministic function f of X, that is, $p(x,y) = p(x)\delta(y - f(x))$. To verify that, we can explicitly calculate

$$H(p(x)\delta(y-f(x))||p(x)u(y)) = \sum_{x,y} p(x)\delta(y-f(x)) \log\left[\frac{\delta(y-f(x))}{\frac{1}{n_Y}}\right] = \sum_x p(x) \log n_Y = \log n_Y$$
$$\implies \quad H(Y|X) = 0.$$

11.8

There are only four possible joint outcomes, which are given by: $(x = 0, y = 0, z = 0)$, $(x = 1, y = 0, z = 1)$, $(x = 0, y = 1, z = 1)$, and $(x = 1, y = 1, z = 0)$, all with equal probability of $1/4$. So we can calculate the entropies

$$H(X) = H(Y) = H(Z) = 2\left(-\frac{1}{2}\log\frac{1}{2}\right) = 1,$$
$$H(X,Y) = H(X,Z) = H(Y,Z) = H(X,Y,Z) = 4\left(-\frac{1}{4}\log\frac{1}{4}\right) = 2.$$

By definition, we obtain

$$H(X,Y:Z) = H(X,Y) + H(Z) - H(X,Y,Z) = 1,$$
$$H(X:Z) = H(X) + H(Z) - H(X,Z) = 0,$$
$$H(Y:Z) = H(Y) + H(Z) - H(Y,Z) = 0,$$

and thus conclude that $H(X,Y:Z) \nleq H(X:Z) + H(Y:Z)$.

11.9

There are only two possible joint outcomes, which are given by: $(x_1 = x_2 = y_1 = y_2 = 0)$, and $(x_1 = x_2 = y_1 = y_2 = 1)$, all with equal probability of $1/2$. So we calculate the entropies

$$H(X_i) = H(Y_i) = H(X_i, Y_i) = H(X_1, X_2) = H(Y_1, Y_2) = H(X_1, X_2, Y_1, Y_2) = 2\left(-\frac{1}{2}\log\frac{1}{2}\right) = 1,$$

for $i \in \{1, 2\}$. By definition, we obtain

$$H(X_1 : Y_1) = H(X_1) + H(Y_1) - H(X_1, Y_1) = 1,$$
$$H(X_2 : Y_2) = H(X_2) + H(Y_2) - H(X_2, Y_2) = 1,$$

$$H(X_1, X_2 : Y_1, Y_2) = H(X_1, X_2) + H(Y_1, Y_2) - H(X_1, X_2, Y_1, Y_2) = 1,$$

and thus conclude that $H(X_1 : Y_1) + H(X_2 : Y_2) \nleq H(X_1, X_2 : Y_1, Y_2)$.

11.10

If $X \to Y \to Z$ is a Markov chain, then the probability of $Y = y$ is conditioned on $X = x$, and the probability of $Z = z$ is conditioned on $Y = y$. In practice, this means that the probability of the joint result $(X, Y, Z) = (x, y, z)$ is given by $p(x)p(y|x)p(z|y)$. But from Bayes' rule (see Exercise **??**)

$$p(x)p(y|x)p(z|y) = p(x|y)p(y)p(z|y) = p(x|y)p(y|z)p(z),$$

which is the probability of $(X, Y, Z) = (x, y, z)$ when $Z \to Y \to X$ is a Markov chain.

11.11

$$\begin{aligned}
\rho &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} : \quad S(\rho) = -1 \times \log 1 - 0 \times \log 0 = 0. \\
\rho &= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} : \quad \text{this is } \rho = |+\rangle\langle+| \text{ in the } X \text{ basis, so} \\
&\qquad S(\rho) = -1 \times \log 1 - 0 \times \log 0 = 0. \\
\rho &= \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} : \quad \det(\rho - \lambda I) = \lambda^2 - \lambda + \frac{1}{9} = 0 \quad \Longrightarrow \quad \text{eigenvalues} = \left\{ \frac{3-\sqrt{5}}{6}, \frac{3+\sqrt{5}}{6} \right\}, \text{ so} \\
&\qquad S(\rho) = -\left(\frac{3-\sqrt{5}}{6}\right)\log\left(\frac{3-\sqrt{5}}{6}\right) - \left(\frac{3+\sqrt{5}}{6}\right)\log\left(\frac{3+\sqrt{5}}{6}\right) \approx 0.55.
\end{aligned}$$

11.12

$$\begin{aligned}
\rho = p\,|0\rangle\langle 0| + (1-p)\,|+\rangle\langle+| &= p\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1-p}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} 1+p & 1-p \\ 1-p & 1-p \end{bmatrix}.
\end{aligned}$$

$$\begin{aligned}
&\det(\rho - \lambda I) = \lambda^2 - \lambda + \frac{p(1-p)}{2} = 0 \\
\Longrightarrow \quad &\text{eigenvalues} = \left\{ \frac{1-\sqrt{1-2p(1-p)}}{2}, \frac{1+\sqrt{1-2p(1-p)}}{2} \right\}.
\end{aligned}$$

If we define $q \equiv \frac{1-\sqrt{1-2p(1-p)}}{2}$, these eigenvalues can be rewritten as $\{q, 1-q\}$, so the von Neumann entropy will be

$$S(\rho) = -q\log(q) - (1-q)\log(1-q).$$

The Shannon entropy $H(p, 1-p)$ will be

$$H(p, 1-p) = -p\log(p) - (1-p)\log(1-p).$$

11.13

Let $\{|i\rangle\}$ be a basis for which ρ is diagonal, so we can write $\rho = \sum_i p_i |i\rangle\langle i|$. Using the joint entropy theorem we have

$$\begin{aligned} S(\rho \otimes \sigma) \equiv S\left(\sum_i p_i |i\rangle\langle i| \otimes \sigma\right) &= H(p_i) + \sum_i p_i S(\sigma) \\ &= H(p_i) + S(\sigma). \end{aligned}$$

Using the definition of entropy, we have

$$\begin{aligned} S(\rho) &= -\operatorname{tr}(\rho\log\rho) \\ &= -\operatorname{tr}\left(\sum_{i,j} p_i |i\rangle\langle i| \log p_j |j\rangle\langle j|\right) \\ &= -\operatorname{tr}\left(\sum_i p_i \log p_i |i\rangle\langle i|\right) \\ &= -\sum_i p_i \log p_i \equiv H(p_i). \end{aligned}$$

Substituting this result in the expression for $S(\rho \otimes \sigma)$, we obtain $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$.

11.14

Given that $|AB\rangle$ is pure, it is immediate that $S(A, B) = 0$. If $|AB\rangle$ is entangled then it can not be written as a product state $|\psi_A\rangle \otimes |\psi_B\rangle$, meaning it has a Schmidt decomposition

$$|AB\rangle = \sum_i \lambda_i |i_A\rangle \otimes |i_B\rangle,$$

where $\{|i_A\rangle\}$ and $\{|i_B\rangle\}$ are basis for A and B respectively, and $\lambda_i \neq 0$ for at least two different values of i. We can calculate the density matrix ρ^A of system A as

$$\rho^A \equiv \operatorname{tr}_B(\rho^{AB}) = \operatorname{tr}_B\left(\sum_{i,j} \lambda_i\lambda_j^* |i_A\rangle\langle j_A| \otimes |i_B\rangle\langle j_B|\right) = \sum_i |\lambda_i|^2 |i_A\rangle\langle i_A|,$$

and so, the entropy of system A will be

$$S(A) = -\sum_i |\lambda_i|^2 \log|\lambda_i|^2.$$

It holds that $\sum_i |\lambda_i|^2 = 1$, and $|\lambda_i|^2 < 1$ for all i, thus $S(A) > 0$, which results in

$$S(B|A) = -S(A) < 0.$$

The converse is immediate.

11.15

Let us consider that in the computational basis ρ has the general form

$$\rho = \sum_{i,j=0}^{1} a_{ij} |i\rangle\langle j| ,$$

with $a_{00} + a_{11} = 1$. The generalized measurement described by the operators M_1 and M_2 is such that the post measurement state is

$$\begin{aligned}\rho' &= M_1 \rho M_1^\dagger + M_2 \rho M_2^\dagger \\ &= |0\rangle\langle 0| \left(\sum_{i,j=0}^{1} a_{ij} |i\rangle\langle j| \right) |0\rangle\langle 0| + |0\rangle\langle 1| \left(\sum_{i,j=0}^{1} a_{ij} |i\rangle\langle j| \right) |1\rangle\langle 0| \\ &= a_{00} |0\rangle\langle 0| + a_{11} |0\rangle\langle 0| \\ &= |0\rangle\langle 0| .\end{aligned}$$

We see that this measurement process always results in the pure state $\rho' = |0\rangle\langle 0|$, meaning $S(\rho') = 0$, which is the smallest value possible for the entropy. Since the entropy of the original state could have be non-negative value between 0 and $\log 2$, we have that $S(\rho') \leq S(\rho)$, which means that this generalized measurement process can decrease the entropy of the qubit.

11.16

-

11.17

-

11.18

Let us first consider that all ρ_is are the same, that is, $\rho_i \equiv \rho$ for all i. Then the left-hand side of (11.79) becomes

$$S\left(\sum_i p_i \rho \right) = S(\rho),$$

and the right-hand side becomes

$$\sum_i p_i S(\rho) = S(\rho),$$

where in both cases we used the fact that $\sum_i p_i = 1$, thus equality holds. Conversely, if we consider that equality holds then we will have

$$S\left(\sum_i p_i \rho_i\right) = \sum_i p_i S(\rho_i).$$

Using the definition of entropy we can write

$$\begin{aligned}
\mathrm{tr}\left[\sum_i p_i \rho_i \log\left(\sum_j p_j \rho_j\right)\right] &= \sum_i p_i \,\mathrm{tr}[\rho_i \log(\rho_i)] \\
\implies \quad \sum_i p_i \,\mathrm{tr}\left[\rho_i \log\left(\sum_j p_j \rho_j\right)\right] &= \sum_i p_i \,\mathrm{tr}[\rho_i \log(\rho_i)].
\end{aligned}$$

This equality only holds if

$$\sum_j p_j \rho_j = \rho_i$$

for all i. Since the left-hand side does not depend on i, all ρ_i are the same.

11.19

-

11.20

-

11.21

For any probability distribution p_i, consider a density operator written in its diagonal basis such that the elements p_i correspond to its eigenvalues. Its von Neumann entropy is

$$S(\rho) = -\sum_i p_i \log p_i \equiv H(p_i).$$

Since $S(\rho)$ is always concave, so is the Shannon entropy $H(p_i)$.

11.22

Defining $f(p) \equiv S(p\rho + (1-p)\sigma)$, we have that S is concave if and only if $S(p\rho + (1-p)\sigma) \leq pS(\rho) + (1-p)S(\sigma)$ for $p \in [0,1]$. So S is concave if and only if $f(p) \leq pS(\rho) + (1-p)S(\sigma)$.

Differentiating both sides with respect to p we obtain

$$f'(p) \leq S(\rho) - S(\sigma),$$

and differentiating again we obtain $f''(p) \leq 0$. So if this is proven to be satisfied, we obtain that the von Neumann entropy is concave.

For the proof, it will be convenient to define the function $M(p) \equiv p\rho + (1-p)\sigma$. Notice that

$$M' = \rho - \sigma \quad \text{and} \quad M'' = 0.$$

From the definition of entropy we can explicitly write

$$f(p) = -\operatorname{tr}[M(p)\log M(p)] \quad \Longrightarrow \quad f'(p) = -\operatorname{tr}\left[\frac{\mathrm{d}}{\mathrm{d}p}\left(M(p)\log M(p)\right)\right]$$

Since ρ and σ are non-negative matrices and $p \in [0,1]$, $M(p)$ is also a non-negative matrix and thus $g(M(p)) \equiv M(p)\log M(p)$ is analytic, meaning it admits a series representation

$$g(M) = \sum_{j=0}^{\infty} a_j M^j,$$

which prompts us to write

$$\begin{aligned}
\operatorname{tr}\left[\frac{\mathrm{d}}{\mathrm{d}p} g(M)\right] &= \operatorname{tr}\left[\sum_{j=1}^{\infty} a_j \sum_{k=0}^{j-1} M^k \frac{\mathrm{d}M}{\mathrm{d}p} M^{j-1-k}\right] \\
&= \operatorname{tr}\left[\sum_{j=1}^{\infty} a_j \sum_{k=0}^{j-1} M^{j-1} \frac{\mathrm{d}M}{\mathrm{d}p}\right] \\
&= \operatorname{tr}\left[\sum_{j=1}^{\infty} a_j (j-1) M^{j-1} \frac{\mathrm{d}M}{\mathrm{d}p}\right] \\
&= \operatorname{tr}\left[\frac{\mathrm{d}g}{\mathrm{d}M}\frac{\mathrm{d}M}{\mathrm{d}p}\right],
\end{aligned}$$

where we have used the cyclic invariance of the trace. We have $\mathrm{d}g/\mathrm{d}M = I + \log M(p)$, thus

$$f'(p) = -\operatorname{tr}[(I + \log M)\, M'].$$

Since $\operatorname{tr}[M'] = \operatorname{tr}[\rho] - \operatorname{tr}[\sigma] = 0$, we can drop the first term, and for a more clean notation, we will denote $M' \equiv C$ since it is a constant matrix. With that, the first derivative of $f(p)$ becomes

$$f'(p) = -\operatorname{tr}[C\log M(p)] \quad \Longrightarrow \quad f''(p) = -\operatorname{tr}\left[C\frac{\mathrm{d}}{\mathrm{d}p}\log M(p)\right].$$

For the second derivative we can use the operator identity

$$\log M = \int_0^{\infty} \mathrm{d}t \left[\frac{1}{1+t} I - (M + tI)^{-1}\right].$$

This expression requires $M + tI$ to be invertible for all $t \geq 0$, which is not satisfied at $t = 0$ in the case where ρ and σ have null eigenvalues. We will address this problem later, for now, let us consider that ρ and σ are both invertible matrices. Differentiating with respect to p on both sides yields

$$\frac{\mathrm{d}}{\mathrm{d}p} \log M = -\int_0^\infty \mathrm{d}t \, \frac{\mathrm{d}}{\mathrm{d}p} \left(M + tI\right)^{-1}.$$

The quantity in the integrand can be obtained by implicitly differentiating the relation $(M+tI)(M+tI)^{-1} = I$, which will yield

$$\frac{\mathrm{d}}{\mathrm{d}p} \log M = \int_0^\infty \mathrm{d}t \, (M + tI)^{-1} \, C \, (M + tI)^{-1}.$$

Substituting back in the expression for $f''(p)$ yields

$$\begin{aligned} f''(p) &= -\operatorname{tr}\left[\int_0^\infty \mathrm{d}t \, C \, (M + tI)^{-1} \, C \, (M + tI)^{-1}\right] \\ &= -\operatorname{tr}\left[\int_0^\infty \mathrm{d}t \, (M + tI)^{-1/2} \, C \, (M + tI)^{-1/2} \, (M + tI)^{-1/2} \, C \, (M + tI)^{-1/2}\right], \end{aligned}$$

where again, we have used the cyclic invariance of the trace. Notice that the quantity $X(p,t) \equiv (M(p) + tI)^{-1/2} \, C \, (M(p) + tI)^{-1/2}$ has real eigenvalues since $M(p)$ always has real eigenvalues, thus this integral is always a non-negative matrix, meaning

$$f''(p) = -\operatorname{tr}\left[\int_0^\infty \mathrm{d}t \, X^2(p,t)\right] \leq 0.$$

For the case where ρ and σ are not invertible, we can define a "regularized" version of entropy as $S_\varepsilon(x) \equiv -\operatorname{tr}[(x + \varepsilon I)\log(x + \varepsilon I)]$. Notice that the only difference would be that, instead of having $M(p)$ defined as it is, we would have $M_\varepsilon(p) \equiv p\rho + (1-p)\sigma + \varepsilon I$, which satisfies $M'_\varepsilon = M' \equiv C$ and equals $M(p)$ in the limit where $\varepsilon \to 0$, thus the same result is obtained when this limit is taken at the end of the process.

11.23

Helding B fixed means to consider $B_1 = B_2 \equiv B$ in the inequality, which will become

$$f(\lambda A_1 + (1-\lambda)A_2, B) \geq \lambda f(A_1, B) + (1-\lambda) f(A_2, B),$$

meaning $f(A,B)$ is concave in A.

Take the function $f(A,B) \equiv \operatorname{tr}(AB)$. Since the trace is linear, we have

$$\begin{aligned} f(\lambda A_1 + (1-\lambda)A_2, B) &= \lambda f(A_1, B) + (1-\lambda) f(A_2, B), \\ f(A, \lambda B_1 + (1-\lambda)B_2) &= \lambda f(A, B_1) + (1-\lambda) f(A, B_2), \end{aligned}$$

so f is clearly concave in each input (in fact, it is also convex since the inequality is saturated). But notice that

$$f(\lambda A_1 + (1-\lambda)A_2, \lambda B_1 + (1-\lambda)B_2) = \lambda^2 f(A_1, B_1) + \lambda(1-\lambda)\left[f(A_1, B_2) + f(A_2, B_1)\right] + (1-\lambda)^2 f(A_2, B_2),$$

which is not always greater than or equal to $\lambda f(A_1, B_1) + (1-\lambda) f(A_2, B_2)$. Considering $n \times n$ matrices, one example would be $A_1 = B_1 = I$ and $A_2 = B_2 = 0$. We would obtain $f(A_1, B_1) = n$ and $f(A_1, B_2) = f(A_2, B_1) = f(A_2, B_2) = 0$, thus

$$f(\lambda A_1 + (1-\lambda)A_2, \lambda B_1 + (1-\lambda)B_2) = \lambda^2 n,$$
$$\lambda f(A_1, B_1) + (1-\lambda) f(A_2, B_2) = \lambda n.$$

Since $0 < \lambda < 1$, we clearly have $f(\lambda A_1 + (1-\lambda)A_2, \lambda B_1 + (1-\lambda)B_2) < \lambda f(A_1, B_1) + (1-\lambda) f(A_2, B_2)$, and therefore, $f(A, B) = \mathrm{tr}(AB)$ is an example of a function that is concave in each of its inputs but is not jointly concave (in fact, it is also not jointly convex).

11.24

Consider a joint system BCD. Strong subadditivity implies

$$S(B, C, D) + S(B) \leq S(B, D) + S(B, C).$$

Let us introduce another system A that purifies BCD, that is, such that $ABCD$ is pure. We have that $S(B, C, D) = S(A)$ and $S(B, D) = S(A, C)$. Substituting these results in the inequality yields

$$S(A) + S(B) \leq S(A, C) + S(B, C).$$

11.25

Consider an ensemble of density operators ρ_i^{AB} of a bipartite system AB, then let

$$\rho^{AB} \equiv \sum_i p_i \rho_i^{AB},$$

with $\rho^A = \sum_i p_i \rho_i^A$ and $\rho^B = \sum_i p_i \rho_i^B$. We can then introduce an auxiliary system C such that

$$\rho^{ABC} \equiv \sum_i p_i \rho_i^{AB} \otimes |i\rangle\langle i|^C.$$

Let us now explicitly calculate the entropies $S(A, B, C)$, and $S(B, C)$. We obtain

$$\begin{aligned} S(A, B, C) &= -\,\mathrm{tr}\left[\rho^{ABC} \log \rho^{ABC}\right] \\ &= -\,\mathrm{tr}\left[\sum_i p_i \rho_i^{AB} \otimes |i\rangle\langle i|^C \sum_j \log\left(p_j \rho_j^{AB}\right) \otimes |j\rangle\langle j|^C\right] \end{aligned}$$

$$
\begin{aligned}
&= -\operatorname{tr}\left[\sum_{i,j} p_i \rho_i^{AB} \left(\log p_j + \log \rho_j^{AB}\right) \otimes |i\rangle\langle i|^C \, |j\rangle\langle j|^C\right] \\
&= -\sum_i p_i \log p_i \operatorname{tr}\left[\rho_i^{AB}\right] - \sum_i p_i \operatorname{tr}\left[\rho_i^{AB} \log \rho_i^{AB}\right] \\
&= H(p_i) + \sum_i p_i S(\rho_i^{AB}),
\end{aligned}
$$

and equivalently,

$$
S(B, C) = H(p_i) + \sum_i p_i S(\rho_i^B).
$$

Notice also that $S(A, B) \equiv S(\rho^{AB})$ and $S(B) \equiv S(\rho^B)$. Substituting these results in the strong subadditivity innequality we obtain

$$
H(p_i) + \sum_i p_i S(\rho_i^{AB}) + S(\rho^B) \leq S(\rho^{AB}) + H(p_i) + \sum_i p_i S(\rho_i^B).
$$

We can rearrange the terms and get

$$
\sum_i p_i \left[S(\rho_i^{AB}) - S(\rho_i^B)\right] \leq S(\rho^{AB}) - S(\rho^B).
$$

The differences of entropies are, by definition, the conditional entropies ($S(A|B) = S(A, B) - S(B)$), so this inequality indicates that the conditional entropy is concave.

11.26

From strong subadditivity we may write

$$
S(B) + S(C) - S(A, B) - S(A, C) \leq 0.
$$

Adding $2S(A)$ on both sides we get

$$
2S(A) + S(B) + S(C) - S(A, B) - S(A, C) \leq 2S(A).
$$

Now we must only identify the mutual information functions on the left-hand side, which are $S(A : B) \equiv S(A) + S(B) - S(A, B)$ and $S(A : C) \equiv S(A) + S(C) - S(A, C)$, so

$$
S(A : B) + S(A : C) \leq 2S(A).
$$