Solutions: Homework Set # 9

Problem 1 (Conditional differential entropy of gaussian random vectors)

(a)

$$
\begin{aligned}
f(x|y) &= \frac{f(x,y)}{f(y)} \\
&= \frac{\frac{1}{2\pi\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_y\sqrt{1-\rho^2}}\exp -\frac{1}{2\sqrt{1-\rho^2}}(\frac{x^2}{\boldsymbol{\Sigma}_x^2}+\frac{y^2}{\boldsymbol{\Sigma}_y^2}-\frac{2\rho xy}{\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_y})}{\frac{1}{\sqrt{2\pi}\boldsymbol{\Sigma}_y}\exp -\frac{y^2}{2\boldsymbol{\Sigma}_y^2}} \\
&= \frac{1}{\sqrt{2\pi}\boldsymbol{\Sigma}_x\sqrt{1-\rho^2}}\exp -\frac{1}{2(1-\rho^2)\boldsymbol{\Sigma}_x^2}(x-\frac{\rho\boldsymbol{\Sigma}_x y}{\boldsymbol{\Sigma}_y})^2
\end{aligned}
$$

Note that for a fixed y, the obtained pdf is a normal one with variance $\boldsymbol{\Sigma}_x^2(1-\rho^2)$ and mean $\frac{\rho\boldsymbol{\Sigma}_x y}{\boldsymbol{\Sigma}_y}$. so

(b)

$$
\begin{aligned}
h(X|Y) &= \int_x f(y)h(X|Y=y)d_y \\
&= \int_x f(y)\ln(\boldsymbol{\Sigma}_x\sqrt{1-\rho^2}\sqrt{2\pi e})d_y \\
&= \ln(\boldsymbol{\Sigma}_x\sqrt{1-\rho^2}\sqrt{2\pi e}).
\end{aligned}
$$

where $h(X|Y=y)=\ln(\boldsymbol{\Sigma}_x\sqrt{1-\rho^2}\sqrt{2\pi e})$ holds because of the note in (a).
You could as well calculate $h(X|Y)$ by $h(X|Y)=\frac{h(X,Y)}{h(Y)}$.

(c)
- $\rho = 0$: $h(X|Y)=\ln(\boldsymbol{\Sigma}_x\sqrt{2\pi e})$ which is in fact $h(X)$ and we know $h(X|Y)=h(X)$ holds for X and Y being independent ($\rho = 0$)
- $\rho = 1$: $h(X|Y) = -\infty$. when $\rho = 1$, $X = Y$ and this means $X|Y = y$ is just a constant and the differential entropy would thus be $-\infty$.

(d) Call the covariance matrix, $\boldsymbol{\Sigma}$ and assume $\mathbf{X}$ and $\mathbf{Y}$ random vectors are both of length N.

$$
f(\mathbf{x},\mathbf{y}) = \frac{1}{(2\pi)^N\det\boldsymbol{\Sigma}^{\frac{1}{2}}}\exp -\frac{1}{2}\,[\mathbf{x}\quad\mathbf{y}]\,\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathbf{x}^t\\ \mathbf{y}^t\end{bmatrix}
$$

So,

$$
\begin{aligned}
f(\mathbf{x}|\mathbf{y}) &= \frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{y})} \\
&= \frac{\frac{1}{(2\pi)^N\det\boldsymbol{\Sigma}^{\frac{1}{2}}}\exp -\frac{1}{2}\,[\mathbf{x}\quad\mathbf{y}]\,\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathbf{x}^t\\ \mathbf{y}^t\end{bmatrix}}{\frac{1}{(2\pi)^{N/2}\det\mathbf{K}_{22}{}^{\frac{1}{2}}}\exp -\frac{1}{2}\mathbf{y}\mathbf{K}_{22}^{-1}\mathbf{y}^t}
\end{aligned}
$$

Applying Matrix Inversion Lemma:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1}+D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$

,

$$f(\mathbf{x}|\mathbf{y}) = \frac{\sqrt{\det \mathbf{K}_{11}}}{(2\pi e)^{N/2}\sqrt{\det \mathbf{\Sigma}}} \exp -\frac{1}{2}(\mathbf{x}-\mathbf{y}\mathbf{K}_{22}^{-1}\mathbf{K}_{12})(\mathbf{K}_{11}-\mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{12})^{-1}(\mathbf{x}-\mathbf{y}\mathbf{K}_{22}^{-1}\mathbf{K}_{12})^t$$

And again, for a fixed vector $\mathbf{y}$, this is a multivariate Gaussian pdf with non-zero mean and covariance matrix $(\mathbf{K}_{11}-\mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{12})$.

(e) You can either find $h(\mathbf{X}|\mathbf{Y})$ similar to part (b), or use $h(\mathbf{Y}|\mathbf{Y}) = \frac{h(\mathbf{X},\mathbf{Y})}{h(\mathbf{Y})}$;

$$\begin{aligned} h(\mathbf{Y}|\mathbf{Y}) &= \frac{h(\mathbf{X},\mathbf{Y})}{h(\mathbf{Y})} \\ &= \ln((2\pi e)^N\sqrt{\det \mathbf{\Sigma}}) - \ln((2\pi e)^{\frac{N}{2}}\sqrt{\det \mathbf{K}_{22}}) \\ &= \ln((2\pi e)^{\frac{N}{2}}\sqrt{\frac{\det \mathbf{K}_{22} \det \mathbf{K}_{11}-\mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{12}}{\det \mathbf{K}_{22}}}) \\ &= \ln((2\pi e)^{\frac{N}{2}}\sqrt{\det \mathbf{K}_{11}-\mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{12}}) \end{aligned}$$

where we have used that $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det D \det(A - CD^{-1}B)$.

Problem 2 (Parallel Gaussian Channel)

(a) The capacity expression for this channel would be

$$C = \max I(X_1, X_2; Y_1Y_2)$$

subject to

$$\mathbb{E}[X_1^2] \le P_1, \quad \mathbb{E}[X_2^2] \le P_2, \quad \beta_1 P_1 + \beta_2 P_2 \le \beta.$$

It is clear that C is achieved if $\beta_1 P_1 + \beta_2 P_2 = \beta$, otherwise one can increase the power, and therefore the mutual information. increase Note that

$$\begin{aligned} I(X_1, X_2; Y_1Y_2) &= h(Y_1Y_2) - h(Y_1, Y_2|X_1, X_2) \\ &= h(Y_1Y_2) - h(Z_1, Z_2|X_1, X_2) \\ &= h(Y_1Y_2) - h(Z_1, Z_2) \\ &= h(Y_1Y_2) - h(Z_1) - h(Z_2) \\ &\le h(Y_1) + h(y_2) - h(Z_1) - h(Z_2) \\ &= I(X_1; Y_1) + I(X_2; Y_2) \end{aligned}$$

where the inequality is tight if and only if Y_1 and Y_2 are independent. If we choose independent input X_1 and X_2 for the channels, then the outputs would be also independent, and we can achieve the maximum mutual information. So,

$$\begin{aligned} C &= \max_{P_1,P_2:\beta_1P_1+\beta_2P_2=\beta} \frac{1}{2}\log\left(1+\frac{P_1}{N_1}\right) + \frac{1}{2}\log\left(1+\frac{P_2}{N_2}\right) \\ &= \max_{P_1,P_2:\beta_1P_1+\beta_2P_2=\beta} \frac{1}{2}\log\left(\frac{(N_1+P_1)(N_2+P_2)}{N_1N_2}\right). \end{aligned}$$

Note that instead of maximizing the whole expression above, we can only maximize $f(P_1, P_2) = (N_1 + P_1)(N_2 + P_2)$ subject to $g(P_1, P_2) = \beta_1 P_1 + \beta_2 P_2 - \beta = 0$, since the rest does not depend on P_1 and P_2. Using the KKT conditions we have,

$$\frac{\partial f}{\partial P_i} = \lambda \frac{\partial g}{\partial P_i} \qquad \forall i : P_i^* > 0.$$

The channel start acting like a pair of channels if $P_1^* > 0$ and $P_2^* > 0$. Therefore,

$$\begin{aligned} P_2 + N_2 &= \lambda \beta_1 \\ P_1 + N_1 &= \lambda \beta_2, \end{aligned}$$

or

$$\beta_1(P_1 + N_1) = \beta_2(P_2 + N_2)$$

. Solving the system of equations

$$\begin{cases} \beta_1 P_1 - \beta_2 P_2 = -\beta_1 N_1 + \beta_2 N_2 \\ \beta_1 P_1 + \beta_2 P_2 = \beta \end{cases}$$

results in

$$P_1^* = \frac{\beta - (\beta_1 N_1 - \beta_2 N_2)}{2\beta_1} \tag{1}$$

$$P_2^* = \frac{\beta + (\beta_1 N_1 - \beta_2 N_2)}{2\beta_2}. \tag{2}$$

Note that P_1^* and P_2^* are positive if and only if $\beta > |\beta_1 N_1 - \beta_2 N_2|$. Therefore, the critical value for β at which the channel starts acting like a pair channel is $\beta^* = |\beta_1 N_1 - \beta_2 N_2|$.

(b) Having the optimal values for P_1^* and P_2^* from (1) and (2), we have

$$P_1^* = \frac{11}{2}, \qquad P_2^* = \frac{9}{4}$$

and

$$C = \frac{1}{2}\log(1 + \frac{11/2}{3}) + \frac{1}{2}\log(1 + \frac{9/4}{2}) = \frac{1}{2}\log(\frac{289}{48}) \simeq 1.295.$$

Problem 3 (Two look Gaussian channel)

The input distribution that achieves capacity is $X \sim \mathcal{N}(0, P)$. Evaluating the mutual information for this distribution we get:

$$\begin{aligned} C_2 =& \max I(X; Y_1, Y_2) \\ =& h(Y_1, Y_2) - h(Y_1, Y_2|X) \\ =& h(Y_1, Y_2) - h(Z_1, Z_2|X) \\ =& h(Y_1, Y_2) - h(Z_1, Z_2). \end{aligned}$$

From the noise covariance matrix we get

$$h(Z_1, Z_2) = \frac{1}{2}\log(2\pi e)^2 |K_Z| = \frac{1}{2}\log(2\pi e)^2 N^2(1 - \rho^2).$$

Since $Y_1 = X + Z_1$ and $Y_2 = X + Z_2$, we have

$$(Y_1, Y_2) \sim \mathcal{N}\left(\mathbf{0}, \left[\begin{array}{cc} P+N & P+\rho N \\ P+\rho N & P+N \end{array}\right]\right)$$

and

$$h(Y_1, Y_2) = \frac{1}{2}\log(2\pi e)^2 |K_Y| = \frac{1}{2}\log(2\pi e)^2 \left(N^2(1-\rho^2) + 2PN(1-\rho)\right).$$

Hence the capacity is

$$C_2 = \frac{1}{2}\log\left(1 + \frac{2P}{N(1+\rho)}\right)$$

(a) For $\rho = 1$ we get $C_2 = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$ which is the single channel capacity. The reason is that $Y_1 = Y_2$ so the additional output symbol is not giving us any extra information.

(b) For $\rho = 0$ the capacity is $C_2 = \frac{1}{2}\log\left(1 + \frac{2P}{N}\right)$ which corresponds to using twice the power in a single look.

(c) For $\rho = -1$ we get $C_2 = \infty$. If we compute $Y_1 + Y_2$ we can perfectly recover X.

Note that in all the cases above the capacity is the same as the capacity of the channel $X \to (Y_1 + Y_2)$.

Problem 4 (Intermittent additive noise channel)

(a) With finite probability the channel is a noise-free channel. In this situation we can guess that the capacity is infinite.

(b) An infinite sequence of bits $b_0 b_1 b2 \ldots$ can be represented by a real number $x = \sum_{i=0}^{\infty} b_i 2^{-(i+1)}$ such that $x \in [0,1]$. Let $x' = 2\sqrt{P}x - \sqrt{P}$ so that $x' \in [-\sqrt{P}, \sqrt{P}]$ be a real number whose amplitude squared is less than P and consider a communication strategy that always sends x' across the channel. Then, when $Z_i = 0$, the receiver gets exactly x'. However, when $Z_i = Z^*$, the receiver observes a corrupted version of x', so in general the receiver cannot know when the channel is noise-free.

Notice that $Pr(Y_i = Y_j | Z_i, Z_j \sim \mathcal{N}) = 0.$, i.e. whenever the signal is corrupted by Gaussian noise, the receiver will never observe the same output symbol twice. The decoding strategy can be the following: observe Y_i's until you receive two identical Y_i's, which happens with probability 1 in finite number of trials. This communication strategy may not be the optimal one, but it does achieve infinite number of bits per symbol on average. Hence, the capacity of this channel is indeed infinity.