Solutions: Homework Set # 9

Problem 1

(a)

$$\max \frac{I(X^n;Y^n)}{n} \overset{a}{\leq} \max \frac{I(X_1;Y_1)}{n} \overset{b}{\leq} \frac{\frac{1}{2}\log(1+\frac{nP}{n})}{n}.$$

where (a) comes from the constraint that all our power, nP , be used at time 1 and (b) comes from that fact that given Gaussian noise and a power constraint nP, $I(X;Y) \leq \frac{1}{2}\log(1+\frac{nP}{N})$.

(b) We have

$$\max \frac{I(X^n;Y^n)}{n} \overset{a}{\leq} \max \frac{nI(X;Y)}{n} = \max I(X;Y) \overset{b}{\leq} \frac{1}{2}\log(1+\frac{P}{N}).$$

where (a) comes from the fact that the channel is memoryless. Notice that the quantity in part (a) goes to zero as $n \to \infty$ while the quantity in part (b) stays constant. Hence the impulse scheme is suboptimal.

Problem 2

This problem has been posed in two different ways in the exercise session and we give the solution to both questions:

(1) $\mathbf{X}$ is not assumed to be Gaussian, $h(\hat{\mathbf{X}}|\mathbf{X})$ is asked to be calculated in part (b).

(a) In order to have $\mathbb{E}(\|\hat{\mathbf{X}}-\mathbf{X}\|^2)$ minimized, it can be shown that $\mathbb{E}\left((\hat{\mathbf{X}}-\mathbf{X})\mathbf{Y}^t\right) = \mathbf{0}$. So,

$$\mathbb{E}(\hat{\mathbf{X}}\mathbf{Y}^t) = \mathbb{E}(\mathbf{X}\mathbf{Y}^t) \quad (1)$$
$$\Rightarrow \mathbf{F}\mathbb{E}(\mathbf{Y}\mathbf{Y}^t) = \mathbb{E}(\mathbf{X}\mathbf{Y}^t) \quad (2)$$
$$\Rightarrow \mathbf{F} = \mathbb{E}(\mathbf{X}\mathbf{Y}^t)\left(\mathbb{E}(\mathbf{Y}\mathbf{Y}^t)\right)^{-1} \quad (3)$$
$$\Rightarrow \mathbf{F} = \mathbb{E}(\mathbf{X}(\mathbf{X}^t+\mathbf{Z}^t))\left(\mathbb{E}((\mathbf{X}+\mathbf{Z})(\mathbf{X}^t+\mathbf{Z}^t))\right)^{-1} \quad (4)$$
$$\Rightarrow \mathbf{F} = \mathbf{K}_X\left(\mathbf{K}_X+\mathbf{I}\right)^{-1} \quad (5)$$
$$(6)$$

The last line follows because X and Z are independent and thus $\mathbb{E}(\mathbf{X}\mathbf{Z}^t) = 0$. Furthermore, $\mathbf{K}_X$ is defined to be the covariance matrix of X, $\mathbf{K}_X = \mathbb{E}(\mathbf{X}\mathbf{X}^t)$.
Then,

$$
\begin{aligned}
\mathbb{E}(\|\hat{\mathbf{X}} - \mathbf{X}\|^2) &= \mathbb{E}(\|\mathbf{F}\mathbf{Y} - \mathbf{X}\|^2) \\
&= \mathbb{E}((\mathbf{F}\mathbf{Y} - \mathbf{X})^t(\mathbf{F}\mathbf{Y} - \mathbf{X})) \\
&= \mathbb{E}(Tr(\mathbf{F}\mathbf{Y} - \mathbf{X})(\mathbf{F}\mathbf{Y} - \mathbf{X})^t) \\
&= Tr\mathbb{E}((\mathbf{F}\mathbf{Y} - \mathbf{X})(\mathbf{F}\mathbf{Y} - \mathbf{X})^t) \\
&= Tr\mathbb{E}(\mathbf{F}\mathbf{Y}(\mathbf{F}\mathbf{Y} - \mathbf{X})^t) - Tr\mathbb{E}(\mathbf{X}(\mathbf{F}\mathbf{Y} - \mathbf{X})^t) \\
&= Tr\mathbf{F}\mathbb{E}(\mathbf{Y}\mathbf{Y}^t)\mathbf{F}^t - Tr\mathbf{F}\mathbb{E}(\mathbf{Y}\mathbf{X}^t) - Tr\mathbb{E}(\mathbf{X}\mathbf{Y}^t)\mathbf{F}^t + Tr\mathbb{E}(\mathbf{X}\mathbf{X})^t \\
&= -Tr\mathbf{F}\mathbb{E}(\mathbf{Y}\mathbf{X}^t) + Tr\mathbb{E}(\mathbf{X}\mathbf{X})^t \quad \text{from equation(2)} \\
&= Tr\left(-\mathbf{F}\mathbb{E}((\mathbf{X} + \mathbf{Z})\mathbf{X}^t) + \mathbb{E}(\mathbf{X}\mathbf{X})^t\right) \\
&= Tr\left(-\mathbf{K}_X\left(\mathbf{K}_X + \mathbf{I}\right)^{-1}\mathbf{K}_X + \mathbf{K}_X\right)
\end{aligned}
$$

(b)

$$
\begin{aligned}
h(\hat{\mathbf{X}}|\mathbf{X}) &= h(\mathbf{F}\mathbf{Y}|\mathbf{X}) \\
&= h(\mathbf{F}(\mathbf{X} + \mathbf{Z})|\mathbf{X}) \\
&= h(\mathbf{F}\mathbf{Z}|\mathbf{X}) \\
&= h(\mathbf{F}\mathbf{Z}) \quad \text{Since } X \text{ and } Y \text{ are independent}
\end{aligned}
$$

$Z \sim \mathcal{N}(0, \mathbf{I})$, Thus $\mathbf{F}\mathbf{Z}$ is also Gaussian, with covariance matrix $\mathbf{K} = \mathbb{E}\left((\mathbf{F}\mathbf{Z})^t\mathbf{F}\mathbf{Z}\right) = \mathbb{E}\left(\mathbf{F}^t\mathbf{Z}^t\mathbf{Z}\mathbf{F}\right) = \mathbf{F}^t\mathbf{F}$. Thus,

$$
\begin{aligned}
h(\hat{\mathbf{X}}|\mathbf{X}) &= h(\mathbf{F}\mathbf{Z}) \\
&= \frac{1}{2}\log(2\pi e \det(\mathbf{F}^t\mathbf{F})) \\
&= \log\left(2\pi e \det\left(\mathbf{K}_X\left(\mathbf{K}_X + \mathbf{I}\right)^{-1}\right)\right).
\end{aligned}
$$

(c) The relation between part (a) and (b) is just through $\mathbf{K}_X$ as calculated above.

(2) $\mathbf{X}$ is assumed to be Gaussian, $h(\mathbf{X}|\hat{\mathbf{X}})$ is asked to be calculated in part (b).

(a) In order to have $\mathbb{E}(\|\hat{\mathbf{X}} - \mathbf{X}\|^2)$ minimized, it can be shown that

$$\mathbb{E}\left((\hat{\mathbf{X}} - \mathbf{X})\mathbf{Y}^t\right) = \mathbf{0}. \tag{7}$$

So,

$$
\begin{aligned}
& \mathbb{E}(\hat{\mathbf{X}}\mathbf{Y}^t) = \mathbb{E}(\mathbf{X}\mathbf{Y}^t) \\
\Rightarrow\ & \mathbf{F}\mathbb{E}(\mathbf{Y}\mathbf{Y}^t) = \mathbb{E}(\mathbf{X}\mathbf{Y}^t) \\
\Rightarrow\ & \mathbf{F} = \mathbb{E}(\mathbf{X}\mathbf{Y}^t)\left(\mathbb{E}(\mathbf{Y}\mathbf{Y}^t)\right)^{-1} \\
\Rightarrow\ & \mathbf{F} = \mathbb{E}(\mathbf{X}(\mathbf{X}^t + \mathbf{Z}^t))\left(\mathbb{E}((\mathbf{X} + \mathbf{Z})(\mathbf{X}^t + \mathbf{Z}^t))\right)^{-1} \\
\Rightarrow\ & \mathbf{F} = \mathbf{K}_X\left(\mathbf{K}_X + \mathbf{I}\right)^{-1}
\end{aligned}
$$

The last line follows because X and Z are independent and thus $\mathbb{E}(\mathbf{X}\mathbf{Z}^t) = 0$. Furthermore, $\mathbf{K}_X$ is defined to be the covariance matrix of X, $\mathbf{K}_X = \mathbb{E}(\mathbf{X}\mathbf{X}^t)$.
Then,

$$\begin{aligned}
\mathbb{E}(\|\hat{\mathbf{X}} - \mathbf{X}\|^2) &= \mathbb{E}((\hat{\mathbf{X}} - \mathbf{X})^t(\hat{\mathbf{X}} - \mathbf{X})) \\
&= \mathbb{E}Tr((\hat{\mathbf{X}} - \mathbf{X})(\hat{\mathbf{X}} - \mathbf{X})^t) \\
&= Tr\mathbf{K}_e.
\end{aligned}$$

where $\mathbf{K}_e$ is the covariance matrix of the random variable $\mathbf{e} = \mathbf{X} - \hat{\mathbf{X}}$.

(b)

$$\begin{aligned}
h(\mathbf{X}|\hat{\mathbf{X}}) &= h(\mathbf{X} - \hat{\mathbf{X}}|\hat{\mathbf{X}}) \\
&= h(\mathbf{e}|\mathbf{F}\mathbf{Y}) \quad \text{define } \mathbf{e} = \mathbf{X} - \hat{\mathbf{X}} \\
&\overset{\star}{=} h(\mathbf{e}) \\
&= \frac{1}{2}\log(2\pi e \det(\mathbf{K}_e)) \quad \mathbf{X} \text{ and } \hat{\mathbf{X}} \text{ are both Gaussian and thus so is } \mathbf{e}
\end{aligned}$$

In this set of equalities, $(\star)$ follows because:
From (7), we know that $\mathbb{E}(\mathbf{e}\mathbf{Y}^t) = 0$. This says that $\mathbb{E}(\mathbf{e}\mathbf{F}\mathbf{Y}^t) = 0$ and thus $\mathbf{e}$ and $\mathbf{F}\mathbf{Y}$ are uncorrelated. At the same time, we know that for Gaussian random variables, being uncorrelated means being independent. Thus $\mathbf{e}$ and $\mathbf{F}\mathbf{Y}$ are independent and thus $h(\mathbf{e}|\mathbf{F}\mathbf{Y}) = h(\mathbf{e})$.

(c) The relation between (a) and (b) is through $\mathbf{K}_e$ as calculated above. i.e., $\mathbb{E}(\|\hat{\mathbf{X}} - \mathbf{X}\|^2) = Tr\mathbf{K}_e$ and $h(\mathbf{X}|\hat{\mathbf{X}}) = \frac{1}{2}\log(2\pi e \det(\mathbf{K}_e))$.

Problem 3

(a) Following the hint, what remains is to maximize the quantity $h(Y_1, U_1) - h(Z_1, U_1)$. Since Z_1 and U_1 are quassian, then the vector (Z_1, U_1) is a gaussian vector with covariance matrix $\left[\begin{smallmatrix} N & \mu N \\ \mu N & N \end{smallmatrix}\right]$. As a result, $h(Z_1, U_2) = \frac{1}{2}\log(2\pi e)^2 N^2(1-\mu^2)$. Also the vector (Y_1, Z_1) has a covariance matrix $\left[\begin{smallmatrix} P+N & \mu N \\ \mu N & N \end{smallmatrix}\right]$ so the maximum possible value of its entropy is when it is a gaussian vector and is $h(Z_1, U_2) = \frac{1}{2}\log(2\pi e)^2(PN + N^2(1-\mu^2))$. As a result the capacity is given by $\frac{1}{2}\log(1 + \frac{P}{(1-\mu^2 N)})$.

(b) $I(\tilde{Y}_1; X_1) \leq I(Y_1, U_1; X_1)$ holds by data processing inequality.
To have $I(\tilde{Y}_1; X_1) = I(Y_1, U_2; X_1)$, let calculate each term separately:

$$\begin{aligned}
I(\tilde{Y}_1; X_1) &= I(Y_1 + \lambda U_1; X_1) \\
&= H(Y_1 + \lambda U_1) + H(X_1) - H(X_1, Y_1 + \lambda U_1) \\
&= \frac{1}{2}\log \frac{(P + N + \lambda^2 N + \lambda\mu_1 N)(P)}{\det\begin{bmatrix} P & P \\ P & P + N + \lambda^2 N + \lambda\mu_1 N) \end{bmatrix}}
\end{aligned}$$

and you have already calculated $I(U_1, Y_1; X_1)$ in part (a). Setting them equal gives the answer for λ

(c) Following the hint,

$$\begin{aligned}
\mathbb{E}[(\hat{X}_1 - X_1)^2] &= \mathbb{E}[(\alpha_1 Y_1 + \beta_1 U_1 - X_1)^2] \\
&= \mathbb{E}[((\alpha_1 - 1)X_1 + \alpha_1 Z_1 + \beta_1 U_1)^2] \\
&= \mathbb{E}[((\alpha_1 - 1)X_1)^2 + (\alpha_1 Z_1 + \beta_1 U_1)^2] \\
&= (\alpha_1 - 1)^2 \mathbb{E}[X_1^2] + \alpha_1^2 \mathbb{E}[Z_1^2] + \beta_1^2 \mathbb{E}[U_1^2] + 2\alpha_1\beta_1 \mathbb{E}[Z_1 U_1] \\
&= (\alpha_1 - 1)^2 P + \alpha_1^2 N + \beta_1^2 N + 2\alpha_1\beta_1 \mu N
\end{aligned}$$

To minimize this value, we take the derivative of it with respect to α_1 and β_1 and set it to zero:

$$\begin{cases} 2(\alpha_1 - 1)P + 2\alpha_1 N + 2\beta_1 \mu N = 0 \\ 2\beta_1 N + 2\alpha_1 \mu N = 0. \end{cases}$$

Solving this set of equations gives the optimal α_1 and β_1.

(d) Since X_1 and X_2 are chosen independently to maximize $I(X_1; U_1, Y_1)$ and $I(X_2; U_2, Y_2)$ respectively, we know that both of them should be Gaussian and thus:

$$\begin{aligned}
C &= H(U_1, Y_1) - H(U_1, Y_1 | X_1) + H(U_2, Y_2) - H(U_2, Y_2 | X_2) && (8) \\
&= H(U_1, Y_1) - H(U_1, Z_1) + H(U_2, Y_2) - H(U_2, Z_2). && (9)
\end{aligned}$$

since (U_1, Z_1) is independent of X_1, and (U_2, Z_2) is independent of X_2. From the correlation between the Gaussian noises that is given in the question, and knowing the fact that the optimizing X_1, X_2 has a Gaussian distribution of power P_1, P_2 respectively, each term can now be computed:

$$\begin{aligned}
H(U_1, Y_1) &= \frac{1}{2} \log \det \left(2\pi e \begin{bmatrix} N + P_1 & \mu_1 N \\ \mu_1 N & N \end{bmatrix} \right) \\
H(U_1, Z_1) &= \frac{1}{2} \log \det \left(2\pi e \begin{bmatrix} N & \mu_1 N \\ \mu_1 N & N \end{bmatrix} \right) \\
H(U_2, Y_2) &= \frac{1}{2} \log \det \left(2\pi e \begin{bmatrix} N + P_2 & \mu_2 N \\ \mu_2 N & N \end{bmatrix} \right) \\
H(U_2, Z_2) &= \frac{1}{2} \log \det \left(2\pi e \begin{bmatrix} N & \mu_2 N \\ \mu_2 N & N \end{bmatrix} \right)
\end{aligned}$$

Thus

$$C = \max_{P_1, P_2:\ P_1 + P_2 \le P} \frac{1}{2} \log \frac{(1 - \mu_1)^2 N^2 + P_1 N}{(1 - \mu_1)^2 N^2} + \frac{1}{2} \log \frac{(1 - \mu_2)^2 N^2 + P_2 N}{(1 - \mu_2)^2 N^2}$$

If you define $N_1 = (1 - \mu_1)^2 N^2$ and $N_2 = (1 - \mu_2)^2 N^2$,It would become just a simple water filling problem to find P_1, P_2 and in fact you have already seen that

$$P_1 = (\nu - N_1)^+,$$

and

$$P_2 = (\nu - N_2)^+,$$

where ν is found by $P_1 + P_2 = P$.

Problem 4

The channel reduces to $Y = 2X + Z_1 + Z_2$. The power constraint on the input $2X$ is $4P$. Z_1 and Z_2 are zero mean, and therefore so is $Z1 + Z2$. Then

$$\begin{aligned} \mathrm{Var}(Z_1 + Z_2) &= E[Z_1^2 + Z_2^2 + 2Z_1 Z_2] \\ &= 2\sigma^2 + 2\rho\sigma^2. \end{aligned}$$

Thus the noise distribution is $\mathcal{N}(0; 2\sigma^2(1+\rho))$.

(a) Plugging the noise and power values into the formula for the one-dimensional $(P; N)$ channel capacity, $C = \frac{1}{2}\log(1 + \frac{P}{N})$, we get $C = \frac{1}{2}\log(1 + \frac{2P}{\sigma^2(1+\rho)})$.

(b) When $\rho = 0$, $C = \frac{1}{2}\log(1 + \frac{2P}{\sigma^2})$. When $\rho = 1$, $C = \frac{1}{2}\log(1 + \frac{P}{\sigma^2})$. When $\rho = -1$, $C = \infty$.