Outstanding Reading

- Read all chapters assigned from IT-I (EE514, Winter 2012).
- Read chapter 8 in the book.
- Read chapter 9 in the book.
Please do use our discussion board (https://catalyst.uw.edu/gopost/board/bilmes/27386/) for all questions, comments, so that all will benefit from them being answered.
Capacity in Band limited case

Next, the signal power per input sample is $TP/2WT = P/2W$. 
Capacity in Band limited case

- Next, the signal power per input sample is $TP/2WT = P/2W$.
- And recall, the noise power per output sample is $\sigma^2/2$. 
Next, the signal power per input sample is $TP/2WT = P/2W$.

And recall, the noise power per output sample is $\sigma^2/2$.

We can then plug this into our existing formula for capacity as a function if SNR, getting

$$C = \frac{1}{2} \log(1 + \frac{P/2W}{\sigma^2/2}) = \frac{1}{2} \log(1 + \frac{P}{\sigma^2W}) \text{ bits/sample} \quad (1)$$
Capacity in Band limited case

- Next, the signal power per input sample is $TP/2WT = P/2W$.
- And recall, the noise power per output sample is $\sigma^2/2$.
- We can then plug this into our existing formula for capacity as a function if SNR, getting

$$C = \frac{1}{2} \log(1 + \frac{P/2W}{\sigma^2/2}) = \frac{1}{2} \log(1 + \frac{P}{\sigma^2W}) \text{ bits/sample} \quad (1)$$

- But then when considering that there are $2W$ samples per second, we get

$$C = W \log(1 + \frac{P}{\sigma^2W}) \text{ bits per second} \quad (2)$$
Capacity in Band limited case

- Next, the signal power per input sample is $TP/2WT = P/2W$.
- And recall, the noise power per output sample is $\sigma^2/2$.
- We can then plug this into our existing formula for capacity as a function if SNR, getting

$$C = \frac{1}{2} \log(1 + \frac{P/2W}{\sigma^2/2}) = \frac{1}{2} \log(1 + \frac{P}{\sigma^2W}) \text{ bits/sample} \quad (1)$$

- But then when considering that there are $2W$ samples per second, we get

$$C = W \log(1 + \frac{P}{\sigma^2W}) \text{ bits per second} \quad (2)$$

- This is a famous formula.
Capacity in Band limited case

- Next, the signal power per input sample is $TP/2WT = P/2W$.
- And recall, the noise power per output sample is $\sigma^2/2$.
- We can then plug this into our existing formula for capacity as a function if SNR, getting
  \[
  C = \frac{1}{2} \log\left(1 + \frac{P}{2W \sigma^2/2}\right) = \frac{1}{2} \log\left(1 + \frac{P}{\sigma^2 W}\right) \text{ bits/sample} \quad (1)
  \]
- But then when considering that there are $2W$ samples per second, we get
  \[
  C = W \log\left(1 + \frac{P}{\sigma^2 W}\right) \text{ bits per second} \quad (2)
  \]
- This is a famous formula.
- We can increase throughput capacity either by increasing bandwidth $W$, increasing signal power $P$, or decreasing noise variance $\sigma^2$. 
Band limited case

\[ C = W \log(1 + \frac{P}{\sigma^2 W}) \text{ bits per second} \] (3)

- We see that \( C \) increases rapidly from \( W \in [0, P/N_0] \) after which \( C \) increases more slowly (diminishing returns) with \( C_\infty = \left( \frac{P}{\sigma^2} \right) \log e \).
Suppose we have $k$ independent Gaussian channels with a common power constraint.

So the noises are uncorrelated since $Z_i \perp \perp Z_j$ for $i \neq j$.

Power constraint (is common to all). I.e.,

$$E \sum_{j=1}^{k} X_j^2 = \sum_{j=1}^{k} EX_j = \sum_{i} P_i \leq P \quad (4)$$
Parallel Channels

- Goal: find maximum capacity in this case capacity. I.e., find

\[ C = \max_{f(x_{1:k}): \sum E X_i^2 \leq P} I(X_{1:k}; Y_{1:k}) \]  \tag{5}  

- We have the following:

\[ I(X_{1:k}) \leq \sum_j (h(Y_j) - h(Z_j)) \leq \sum_j 1/2 \log(1 + P_i N_i) \]  \tag{7}  

with \( P_i = E X_i^2 \) and \( \sum_i P_i = P \).
Parallel Channels

- Goal: find maximum capacity in this case capacity. I.e., find

\[
C = \max_{f(x_1:k): \sum EX_i^2 \leq P} I(X_1:k; Y_1:k) \tag{5}
\]

- We have the following:

\[
I(X_1:k) = h(Y_1:k) - h(Y_1:k | X_1:k) \tag{7}
\]

\[
\leq \sum_j \left( h(Y_j) - h(Z_j) \right) \leq \sum_j \frac{1}{2} \log(1 + P_i N_i) \tag{8}
\]

with \( P_i = EX_i^2 \) and \( \sum_i P_i = P \).
Parallel Channels

- Goal: find maximum capacity in this case capacity. I.e., find

\[
C = \max_{f(x_1:k): \sum E X_i^2 \leq P} I(X_1:k; Y_1:k)
\]  \hspace{1cm} (5)

- We have the following:

\[
I(X_1:k) = h(Y_1:k) - h(Y_1:k | X_1:k) = h(Y_1:k) - h(Z_1:k | X_1:k)
\]  \hspace{1cm} (6)

\[
\leq \sum_j (h(Y_j) - h(Z_j))
\]  \hspace{1cm} (7)

\[
\leq \sum_j 1/2 \log(1 + P_i N_i)
\]  \hspace{1cm} (8)

with \( P_i = EX_i^2 \) and \( \sum_i P_i = P \).
Parallel Channels

- Goal: find maximum capacity in this case capacity. I.e., find

\[ C = \max_{f(x_1:k) : \sum E X_i^2 \leq P} I(X_1:k; Y_1:k) \]  \hspace{1cm} (5)

- We have the following:

\[ I(X_1:k) = h(Y_1:k) - h(Y_1:k|X_1:k) = h(Y_1:k) - h(Z_1:k|X_1:k) \]  \hspace{1cm} (6)

\[ = h(Y_1:k) - h(Z_1:k) \]  \hspace{1cm} (7)

\[ \leq \sum_i P_i = \sum E X_i^2 \]  \hspace{1cm} (8)

with \( P_i = E X_i^2 \) and \( \sum_i P_i = P \).
Goal: find maximum capacity in this case capacity. I.e., find

$$C = \max_{f(x_1:k) : \sum EX_i^2 \leq P} I(X_1:k; Y_1:k)$$

(5)

We have the following:

$$I(X_1:k) = h(Y_1:k) - h(Y_1:k | X_1:k) = h(Y_1:k) - h(Z_1:k | X_1:k)$$

(6)

$$= h(Y_1:k) - h(Z_1:k) = h(Y_1:k) - \sum_{j=1}^{k} h(Z_j)$$

(7)

$$\leq \sum_{i} P_i = P. \text{ with } P_i = EX_i^2$$

(8)
Parallel Channels

- Goal: find maximum capacity in this case capacity. I.e., find

\[ C = \max_{f(x_1:k) : \sum E X_i^2 \leq P} I(X_1:k; Y_1:k) \]  

(5)

- We have the following:

\[ I(X_1:k) = h(Y_1:k) - h(Y_1:k | X_1:k) = h(Y_1:k) - h(Z_1:k | X_1:k) \]

(6)

\[ = h(Y_1:k) - h(Z_1:k) = h(Y_1:k) - \sum_{j=1}^{k} h(Z_j) \]

(7)

\[ \leq \sum_{j} (h(Y_j) - h(Z_j)) \]

(8)

with \( P_i = E X_i^2 \) and \( \sum_i P_i = P \).
Parallel Channels

- Goal: find maximum capacity in this case capacity. I.e., find

\[
C = \max_{f(x_{1:k}) \text{ such that } \sum EX^2_i \leq P} I(X_{1:k}; Y_{1:k}) \tag{5}
\]

- We have the following:

\[
I(X_{1:k}) = h(Y_{1:k}) - h(Y_{1:k}|X_{1:k}) = h(Y_{1:k}) - h(Z_{1:k}|X_{1:k}) \tag{6}
\]

\[
= h(Y_{1:k}) - h(Z_{1:k}) = h(Y_{1:k}) - \sum_{j=1}^{k} h(Z_j) \tag{7}
\]

\[
\leq \sum_{j} (h(Y_j) - h(Z_j)) \leq \sum_{j} \frac{1}{2} \log(1 + \frac{P_i}{N_i}) \tag{8}
\]

with \( P_i = EX^2_i \) and \( \sum_i P_i = P \).
The way to solve this is to solve

\[
\text{maximize} \quad \sum_{j} \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right) \\
\text{subject to} \quad \sum_{i} P_i = P
\]  

(9)

Or in Lagrangian form

\[
J(P_{1:n}) = \sum_{j} \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right) + \lambda \left(\sum_{j} P_i - P\right)
\]  

(11)
Lagrangian optimization

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0 \text{ for } i = 1, \ldots, m \\
& \quad h_i(x) = 0 \text{ for } i = 1, \ldots, p
\end{align*}
\]
Lagrangian optimization

minimize \[ f_0(x) \] \hspace{1cm} (12)
subject to \[ f_i(x) \leq 0 \] for \( i = 1, \ldots, m \) \hspace{1cm} (13)
\[ h_i(x) = 0 \] for \( i = 1, \ldots, p \) \hspace{1cm} (14)

Lagrangian form

\[ L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \] \hspace{1cm} (15)
Lagrangian optimization

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0 \text{ for } i = 1, \ldots, m \\
& \quad h_i(x) = 0 \text{ for } i = 1, \ldots, p
\end{align*}
\] (12), (13), (14)

Lagrangian form

\[
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
\] (15)

and define dual objective

\[
g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)
\] (16)
Strong duality

- **Notation:** $\lambda_{\text{opt}} = \lambda^*$, $\nu_{\text{opt}} = \nu^*$
Strong duality

- Notation: \( \lambda_{opt} = \lambda^* \), \( \nu_{opt} = \nu^* \)

- Strong duality means \( f(x_{opt}) = g(\lambda_{opt}, \nu_{opt}) \)
Strong duality

- **Notation:** $\lambda_{opt} = \lambda^*$, $\nu_{opt} = \nu^*$

- **Strong duality means** $f(x_{opt}) = g(\lambda_{opt}, \nu_{opt})$

- **If this holds,** last time we derived KKT conditions for optimality, namely that:

\[
\begin{align*}
    f_i(x_{opt}) &\leq 0 \text{ for } i = 1, \ldots, m \\
    h_i(x_{opt}) &= 0 \text{ for } i = 1, \ldots, p \\
    \lambda_i^* &\geq 0 \text{ for } i = i, \ldots, m \\
    \lambda_i^* f_i(x_{opt}) &= 0 \text{ for } i = 1, \ldots, m
\end{align*}
\]

and

\[
\nabla_x L |_{x=x_{opt}} = 0
\]
Back to our problem

- Consider again our problem from before, on the next slide.
Parallel Channels

The way to solve this is to solve

\[
\begin{align*}
\text{maximize} & \quad \sum_j \frac{1}{2} \log(1 + \frac{P_i}{N_i}) \\
\text{subject to} & \quad \sum_i P_i = P
\end{align*}
\] (9)

Or in Lagrangian form

\[
J(P_1:n) = \sum_j \frac{1}{2} \log(1 + \frac{P_i}{N_i}) + \lambda(\sum_j P_i - P)
\] (11)
Back to our problem

- Consider again our problem from before, on the next slide.
- $x = (P_1, P_2, \ldots, P_m)$ vector of power values.
Back to our problem

- Consider again our problem from before, on the next slide.
- \( x = (P_1, P_2, \ldots, P_m) \) vector of power values.
- \( N_i \) noises given (before we called these \( \sigma_i^2 \)).
Consider again our problem from before, on the next slide.

\( x = (P_1, P_2, \ldots, P_m) \) vector of power values.

\( N_i \) noises given (before we called these \( \sigma_i^2 \)).

We want to do

\[
\text{minimize } - \sum_{i=1}^{k} \log(1 + P_i/N_i)
\]  \hspace{1cm} (22)

subject to inequality constraints \( P_i \geq 0 \) \( \forall i \) (so \( f_i(P_i) = -P_i \)), and equality constraints \( \sum_{i=1}^{k} P_i = P \) (i.e., \( h = (\sum_{j=1}^{n} P_i - P) = 0 \)).
Consider again our problem from before, on the next slide.

\[ x = (P_1, P_2, \ldots, P_m) \text{ vector of power values.} \]

\[ N_i \text{ noises given (before we called these } \sigma_i^2). \]

We want to do

\[
\text{minimize } - \sum_{i=1}^{k} \log(1 + P_i/N_i)
\]

subject to inequality constraints \( P_i \geq 0 \ \forall i \) (so \( f_i(P_i) = -P_i \)), and equality constraints \( \sum_{i=1}^{k} P_i = P \) (i.e., \( h = (\sum_{j=1}^{n} P_i - P) = 0 \)).

This problem is convex.
Back to our problem

- Consider again our problem from before, on the next slide.
- \( x = (P_1, P_2, \ldots, P_m) \) vector of power values.
- \( N_i \) noises given (before we called these \( \sigma_i^2 \)).
- We want to do

\[
\text{minimize } - \sum_{i=1}^{k} \log(1 + P_i/N_i)
\]

subject to inequality constraints \( P_i \geq 0 \ \forall i \) (so \( f_i(P_i) = -P_i \)), and equality constraints \( \sum_{i=1}^{k} P_i = P \) (i.e., \( h = (\sum_{j=1}^{n} P_i - P) = 0 \)).

- This problem is convex.
- Also, there exists a strictly feasible point, so strong duality holds, as do the KKT conditions for optimality.
Back to our problem

We get Lagrangian

\[ L(x, \lambda, \nu) = - \sum_{i=1}^{k} \log(1 + P_i/N_i) - \sum_{i=1}^{k} \lambda_i P_i + \nu \left( \sum_{i=1}^{k} P_i - P \right) \]

(23)
Back to our problem

- We get Lagrangian

\[ L(x, \lambda, \nu) = - \sum_{i=1}^{k} \log(1 + P_i/N_i) - \sum_{i=1}^{k} \lambda_i P_i + \nu \left( \sum_{i=1}^{k} P_i - P \right) \]  

(23)

- KKT conditions are:

\[ \forall i : P_i^* \geq 0, \sum_i P_i^* = P, \lambda_i^* \geq 0 \forall i, \text{ and } \lambda_i^* P_i^* = 0 \]  

(24)

and also \( \forall i \)

\[- \frac{1}{(1 + P_i/N_i)} \frac{1}{N_i} - \lambda_i^* + \nu^* = 0 \]  

(25)
KKT Conditions

- From the Lagrangian gradient conditions we can further get

\[- \frac{1}{N_i + P_i} - \lambda_i^* + \nu^* = 0\]  \hspace{1cm} (26)

\[\Rightarrow - \frac{1}{N_i + P_i} + \nu^* = \lambda_i^* \geq 0\]  \hspace{1cm} (27)

\[\nu^* \geq \frac{1}{N_i + P_i}\]  \hspace{1cm} (28)
KKT Conditions

- From the Lagrangian gradient conditions we can further get

\[- \frac{1}{N_i + P_i} - \lambda_i^* + \nu^* = 0 \]  

(26)

\[
\Rightarrow - \frac{1}{N_i + P_i} + \nu^* = \lambda_i^* \geq 0
\]  

(27)

(28)

- We then eliminate \( \lambda_i^* \) (the slack variable) to get KKT conditions in form

\[
\forall i : P_i^* \geq 0,
\]

(29)

\[
\sum_i P_i^* = P
\]

\[
\left( \nu^* - \frac{1}{N_i + P_i} \right) P_i^* = 0
\]

\[
\nu^* \geq \frac{1}{N_i + P_i^*}
\]  

(30)
We have two cases

- **Case 1**: From condition $\nu^* \geq \frac{1}{N_i+P_i^*}$, if $\nu^* < 1/N_i$, then we must have $P_i^* > 0$ to achieve the condition.
We have two cases

- **Case 1:** From condition $\nu^* \geq \frac{1}{N_i + P_i^*}$, if $\nu^* < 1/N_i$, then we must have $P_i^* > 0$ to achieve the condition.

- In such case, since $(\nu^* - \frac{1}{N_i + P_i^*}) P_i^* = 0$, we must have
  \[ \nu^* = \frac{1}{N_i + P_i^*}. \]
We have two cases

- **Case 1:** From condition $\nu^* \geq \frac{1}{N_i + P_i^*}$, if $\nu^* < 1/N_i$, then we must have $P_i^* > 0$ to achieve the condition.

- In such case, since $\left(\nu^* - \frac{1}{N_i + P_i}\right) P_i^* = 0$, we must have $\nu^* = \frac{1}{N_i + P_i^*}$.

- Hence, $P_i^* = \frac{1}{\nu^*} - N_i$
We have two cases

- **Case 1**: From condition $\nu^* \geq \frac{1}{N_i + P_i^*}$, if $\nu^* < 1/N_i$, then we must have $P_i^* > 0$ to achieve the condition.

- In such case, since $\left(\nu^* - \frac{1}{N_i + P_i}\right) P_i^* = 0$, we must have
  $$\nu^* = \frac{1}{N_i + P_i^*}.$$  

- Hence, $P_i^* = \frac{1}{\nu^*} - N_i$

- **Case 2**: If $\nu^* \geq 1/N_i$, then $P_i^* = 0$ since otherwise ($P_i^* > 0$) would mean
  $$(\nu^* - \frac{1}{N_i + P_i}) P_i^* > 0 \quad \text{(31)}$$
  >0 since $\nu^* \geq 1/N_i$ and $P_i^* > 0$
So, $P_i$ must have form

$$P_i^* = \begin{cases} 
1/\nu^* - N_i & \text{if } \nu^* < 1/N_i \\
0 & \text{if } \nu^* \geq 1/N_i 
\end{cases}$$

(32)

$$= \max \left\{ 0, 1/\nu^* - N_i \right\}$$

(33)
So, $P_i$ must have form

$$P_i^* = \begin{cases} 
\frac{1}{\nu^*} - N_i & \text{if } \nu^* < \frac{1}{N_i} \\
0 & \text{if } \nu^* \geq \frac{1}{N_i} 
\end{cases}$$  \hspace{1cm} (32)

$$= \max \{0, \frac{1}{\nu^*} - N_i\}$$  \hspace{1cm} (33)

With the last constraint, we have that

$$\sum_i \left( \frac{1}{\nu^*} - N_i \right)^+ = P$$  \hspace{1cm} (34)

where $a^+ = \max \{0, a\}$. 
Condition on $P_i$

- So, $P_i$ must have form

$$P_i^* = \begin{cases} 
\frac{1}{\nu^*} - N_i & \text{if } \nu^* < \frac{1}{N_i} \\
0 & \text{if } \nu^* \geq \frac{1}{N_i} 
\end{cases} \quad (32)$$

$$= \max \left\{ 0, \frac{1}{\nu^*} - N_i \right\} \quad (33)$$

- With the last constraint, we have that

$$\sum_i \left( \frac{1}{\nu^*} - N_i \right)^+ = P \quad (34)$$

where $a^+ = \max \{0, a\}$.

- This leads to the famous water filling idea for parallel channels.
Water Filling

\[ \sum_i \left( \frac{1}{\nu^*} - N_i \right)^+ = P \]  

- Think of this as gradually increasing $1/\nu^*$ until Equation 36 is satisfied.
Water Filling

\[ \sum_i \left( \frac{1}{\nu^*} - N_i \right)^+ = P \] (36)

- We can think of this as filling up a bin with uniformly dispersed rain water and stopping the rain until we achieve equality in Eq. 36.
Water Filling

\[ \sum_i \left( \frac{1}{\nu^*} - N_i \right)^+ = P \] (36)

- We can think of this as filling up a bin with uniformly dispersed rain water and stopping the rain until we achieve equality in Eq. 36.

- the red area indicated above is equal to \( P \).
Capacity of Parallel Channels

- So the final capacity is then

\[
C_n = \frac{1}{2} \sum_{j=1}^{k} \log \left( 1 + \frac{P_i}{N_i} \right)
\]  \hspace{1cm} (37)

\[
= \frac{1}{2} \sum_{j=1}^{k} \log \left( 1 + \frac{1/\nu^* - N_i}{N_i} \right) \text{ bits per parallel channel use}
\]  \hspace{1cm} (38)
Capacity of Parallel Channels

- So the final capacity is then

\[
C_n = \frac{1}{2} \sum_{j=1}^{k} \log(1 + P_i/N_i) \tag{37}
\]

\[
= \frac{1}{2} \sum_{j=1}^{k} \log \left( 1 + \frac{(1/\nu^* - N_i)^+}{N_i} \right) \text{ bits per parallel channel use} \tag{38}
\]

- In units of bits per transmission (bits per single channel transmission, take the average):

\[
C_n = \frac{1}{2n} \sum_{j=1}^{k} \log \left( 1 + \frac{(\nu^* - N_i)^+}{N_i} \right) \text{ bits per transmission} \tag{39}
\]
Before we had $k$ independent Gaussian channels with a common power constraint and uncorrelated noise.

Now, have colored noise, i.e.,

$$Z_1: k \sim N(0, K_Z)$$

where $K_Z$ is not necessarily diagonal (i.e., noise is correlated from one time step to the next).
Before we had $k$ independent Gaussian channels with a common power constraint and uncorrelated noise.

Now, have colored noise, i.e.,

$$Z_1:k \sim N(0, K_Z)$$

(40)

where $K_Z$ is not necessarily diagonal (i.e., noise is correlated from one time step to the next).
Before and After

- Before we had $k$ independent Gaussian channels with a common power constraint and uncorrelated noise.

\[
\begin{align*}
X_1 &\rightarrow \left(\begin{array}{c}
z_1 \\
z_2 \\
\vdots \\
z_k \\
X_k
\end{array}\right) \rightarrow \quad \left(\begin{array}{c}
Y_1 \\
Y_2 \\
\vdots \\
Y_k
\end{array}\right)
\end{align*}
\]

- Now, have colored noise, i.e.,

\[
Z_{1:k} \sim N(0, K_Z)
\]  

(40)

where $K_Z$ is not necessarily diagonal (i.e., noise is correlated from one time step to the next).

- We still assume $X_{1:n} \perp \perp Z_{1:n}$ (so noise and signal are independent).
Colored Noise

- Define $K_X = EXX^\top$
Colored Noise

- Define \( K_X = EXX^\top \)
- We have power constraint of the form

\[
\frac{1}{n} \sum_i EX_i^2 = \frac{1}{n} \text{tr}(K_X) \leq P
\] (41)
Colored Noise

- Define $K_X = EXX^\top$
- We have power constraint of the form
  \[
  \frac{1}{n} \sum_i E X_i^2 = \frac{1}{n} \text{tr}(K_X) \leq P
  \] (41)

- Let's first form upper bound:
  \[
  I(X_1:n; Y_1:n) = h(Y_1:n) - h(Z_1:n)
  \] (42)
Colored Noise

- Define $K_X = EXX^\top$
- We have power constraint of the form

$$\frac{1}{n} \sum_i E X_i^2 = \frac{1}{n} \text{tr}(K_X) \leq P \quad (41)$$

- Lets first form upper bound:

$$I(X_1:n; Y_1:n) = h(Y_1:n) - h(Z_1:n) \quad (42)$$

- and

$$h(Y_1:n) \leq \frac{1}{2} \log ((2\pi e)^n |K_X + K_Z|) \text{ since } X \perp \!\!\!\!\!\perp Z \quad (43)$$
Colored Noise

Now, $K_Z$ is fixed, we wish to maximize $K_X$ in Equation 43
Colored Noise

- Now, $K_Z$ is fixed, we wish to maximize $K_X$ in Equation 43
- I.e., we want to do

$$\max_{K_X \succeq 0} |K_X + K_Z| \quad \text{subject to} \quad \frac{1}{n} \text{tr}(K_X) \leq P \quad (44)$$
Colored Noise

Now, $K_Z$ is fixed, we wish to maximize $K_X$ in Equation 43

I.e., we want do do

$$\maximize_{K_X \succeq 0} |K_X + K_Z| \quad \text{subject to} \quad \frac{1}{n} \text{tr}(K_X) \leq P$$  \hspace{1cm} (44)

Note that $K_X \succeq 0$ is a positive semidefinite constraint
Colored Noise

- Now, $K_Z$ is fixed, we wish to maximize $K_X$ in Equation 43.
- i.e., we want do do do

$$\text{maximize} \left| K_X + K_Z \right| \quad \text{subject to} \quad \frac{1}{n} \text{tr}(K_X) \leq P \quad (44)$$

- Note that $K_X \succeq 0$ is a positive semidefinite constraint.
- Now via singular value decomposition (spectral eigenvalue factorization) $K_Z = Q\Lambda Q^\top$ with $QQ^\top = I$ and $|Q| = 1$. Hence
Colored Noise

Now, $K_Z$ is fixed, we wish to maximize $K_X$ in Equation 43

I.e., we want to do

$$\max_{K_X \succeq 0} |K_X + K_Z| \quad \text{subject to} \quad \frac{1}{n} \text{tr}(K_X) \leq P \quad (44)$$

Note that $K_X \succeq 0$ is a positive semidefinite constraint.

Now via singular value decomposition (spectral eigenvalue factorization) $K_Z = Q \Lambda Q^T$ with $QQ^T = I$ and $|Q| = 1$. Hence

$$|K_X + K_Z| = |K_X + Q \Lambda Q^T| \quad (45)$$

$$= |Q||Q^T K_X Q + \Lambda||Q^T| \quad (46)$$

$$= |Q^T K_X Q + \Lambda| \quad (47)$$

$$= |A + \Lambda| \quad (48)$$
Colored Noise

- Now, $K_Z$ is fixed, we wish to maximize $K_X$ in Equation 43.
- I.e., we want to do:

$$\max_{K_X \succeq 0} |K_X + K_Z| \text{ subject to } \frac{1}{n} \text{tr}(K_X) \leq P \quad (44)$$

- Note that $K_X \succeq 0$ is a positive semidefinite constraint.
- Now via singular value decomposition (spectral eigenvalue factorization) $K_Z = Q\Lambda Q^\top$ with $QQ^\top = I$ and $|Q| = 1$. Hence

$$|K_X + K_Z| = |K_X + Q\Lambda Q^\top|$$

$$= |Q||Q^\top K_X Q + \Lambda||Q^\top|$$

$$= |Q^\top K_X Q + \Lambda|$$

$$= |A + \Lambda| \quad (46)$$

- So, $K_X = QAQ^\top$.
But trace is preserved, since

\[ \text{tr}(A) = \text{tr}(Q^T K_X Q) = \text{tr}(QQ^T K_X) = \text{tr}(K_X) \quad (49) \]
Colored Noise

- But trace is preserved, since

\[ \text{tr}(A) = \text{tr}(Q^T K_X Q) = \text{tr}(QQ^T K_X) = \text{tr}(K_X) \]  

(49)

- So the optimization problem we get is:

\[ \max_{A} |A + \Lambda| \quad \text{subject to} \quad \text{tr}(A) \leq nP \]  

(50)
But trace is preserved, since

$$\text{tr}(A) = \text{tr}(Q^TKXQ) = \text{tr}(QQ^TKX) = \text{tr}(KX)$$  \hspace{1cm} (49)

So the optimization problem we get is:

$$\max_A |A + \Lambda| \quad \text{subject to} \quad \text{tr}(A) \leq nP$$  \hspace{1cm} (50)

Recall, for any jointly Gaussian random vector $U_{1:n}$, we have

$$h(U_{1:n}) = \frac{1}{2} \log((2\pi e)^n |K_U|)$$  \hspace{1cm} (51)

$$\leq \sum_i h(U_i)$$  \hspace{1cm} (52)

$$= \frac{1}{2} \log((2\pi e)^n \det(K_U))$$  \hspace{1cm} (53)

where $|\det(K_U)| = \prod_{i=1}^n K_U(i,i)$
Thus

\[ |A + \Lambda| \leq \prod_{i} (A_{ii} + \lambda_i) \]  \hspace{1cm} (54)

again noting that \( \{\lambda_i\}_i \) are the eigenvalues of \( K_Z \).
Thus

\[ |A + \Lambda| \leq \prod_{i} (A_{ii} + \lambda_i) \]  \hspace{1cm} (54)

again noting that \( \{\lambda_i\}_i \) are the eigenvalues of \( K_Z \).

This achieves equality when \( A = \text{diag}(A) \), reducing our optimization problem to:

\[
\text{maximize } \prod_i (A_{ii} + \lambda_i) \text{ subject to } \sum_i A_{ii} = nP
\]  \hspace{1cm} (55)
Colored Noise

Thus

$$|A + \Lambda| \leq \prod_i (A_{ii} + \lambda_i)$$  \hspace{1cm} (54)

again noting that \(\{\lambda_i\}_i\) are the eigenvalues of \(K_Z\).

This achieves equality when \(A = \text{diag}(A)\), reducing our optimization problem to:

$$\text{maximize} \prod_i (A_{ii} + \lambda_i) \quad \text{subject to} \quad \sum_i A_{ii} = nP$$  \hspace{1cm} (55)

Why can we say \(\sum_i A_{ii} = nP\) rather than \(\sum_i A_{ii} \leq nP\)?
Thus

\[ |A + \Lambda| \leq \prod_i (A_{ii} + \lambda_i) \]  \hspace{1cm} (54)

again noting that \( \{\lambda_i\}_i \) are the eigenvalues of \( K_Z \).

This achieves equality when \( A = \text{diag}(A) \), reducing our optimization problem to:

\[
\max_{\{A_{ii} \geq 0\}_i} \prod_i (A_{ii} + \lambda_i) \quad \text{subject to} \quad \sum_i A_{ii} = nP
\]  \hspace{1cm} (55)

Why can we say \( \sum_i A_{ii} = nP \) rather than \( \sum_i A_{ii} \leq nP \)?

Note, we use equality in the constraint since \( \lambda_i \geq 0 \) and also we have constraint \( A_{ii} \geq 0 \) in the maximum process.
Like before, we have KKT conditions to get:

\[ A_{ii} = (\nu - \lambda_i)^+ \]  \hspace{1cm} (56)

and

\[ \sum_i (\nu - \lambda_i)^+ = nP \]  \hspace{1cm} (57)
Colored Noise - KKT

- Like before, we have KKT conditions to get:

\[ A_{ii} = (\nu - \lambda_i)^+ \quad (56) \]

and

\[ \sum_i (\nu - \lambda_i)^+ = nP \quad (57) \]

- And the total capacity becomes

\[ C_n = \frac{1}{2} \sum_{i=1}^{n} \log(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i}) \text{ bits per transmission} \quad (58) \]
Colored Noise - KKT

- Like before, we have KKT conditions to get:

\[ A_{ii} = (\nu - \lambda_i)^+ \]  

(56)

and

\[ \sum_i (\nu - \lambda_i)^+ = nP \]  

(57)

- And the total capacity becomes

\[ C_n = \frac{1}{2} \sum_{i=1}^n \log(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i}) \] \text{ bits per transmission}  

(58)

- So this is water filling again, but on the eigenvalues of \( K_Z \).
Like before, we have KKT conditions to get:

\[ A_{ii} = (\nu - \lambda_i)^+ \]  \hspace{1cm} (56)

and

\[ \sum_i (\nu - \lambda_i)^+ = nP \]  \hspace{1cm} (57)

And the total capacity becomes

\[ C_n = \frac{1}{2} \sum_{i=1}^{n} \log(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i}) \]  \hspace{1cm} \text{bits per transmission} \hspace{1cm} (58)

So this is water filling again, but on the eigenvalues of \( K_Z \).

I.e., we are filling a diagonalized version of the problem. Once we get \( A_{ii} \), we can “re-correlate” using \( K_X = QAQ^\top \) to get final constraint on \( X \).
So now we have positive semidefinite kernel functions $f_X$ and $f_Z$ such that (stationary case)

$$EX_iX_j = f_X(|i - j|)$$ and $$EZ_iZ_j = f_Z(|i - j|)$$ (59)

so that the matrices are symmetric Toeplitz (i.e., symmetric same along each diagonal)
Stationary Stochastic (Gaussian) Processes

- It can be shown that we have water filling in this case as well, but this time in the frequency domain.
Stationary Stochastic (Gaussian) Processes

- It can be shown that we have water filling in this case as well, but this time in the frequency domain.
- Let $R_{ZZ}(e^{j\omega})$ be the power spectrum of the noise and signal discrete FT $X(e^{j\omega})$. 
It can be shown that we have water filling in this case as well, but this time in the frequency domain.

Let $R_{ZZ}(e^{j\omega})$ be the power spectrum of the noise and signal discrete FT $X(e^{j\omega})$.

Then the capacity in this case becomes:

$$C = \int_{-\pi}^{\pi} \frac{1}{2} \log \left( 1 + \frac{(\nu - R_{ZZ}(e^{j\omega}))^+}{R_{ZZ}(e^{j\omega})} \right) d\omega$$  \hspace{1cm} (60)$$

where $\nu$ is chosen to make

$$\int_{-\pi}^{\pi} (\nu - R_{ZZ}(e^{j\omega}))^+ d\omega = P$$  \hspace{1cm} (61)$$
Spectral Water filling

- Pictured:

\[ R_x(e^{j\omega}) \]

So put energy first where the noise is low.
Feedback

- For discrete memoryless channel, feedback does not increase capacity (since in finding $C$, you are already exploiting the randomness to the extent possible).
Feedback

- For discrete memoryless channel, feedback does not increase capacity (since in finding $C$, you are already exploiting the randomness to the extent possible).
- For discrete non-memoryless channels, feedback could help (but not by too much).
Feedback

- For discrete memoryless channel, feedback does not increase capacity (since in finding $C$, you are already exploiting the randomness to the extent possible).
- For discrete non-memoryless channels, feedback could help (but not by too much).
- Here, the analogy is that the noise is correlated, i.e.,

$$Z_{1:n}^\top \sim \mathcal{N}(0, K_{Z}^{(n)})$$

(62)

where again $K_{Z}^{(n)}$ is not a diagonal matrix.
Feedback

- For discrete memoryless channel, feedback does not increase capacity (since in finding $C$, you are already exploiting the randomness to the extent possible).
- For discrete non-memoryless channels, feedback could help (but not by too much)
- Here, the analogy is that the noise is correlated, i.e.,

$$Z_{1:n}^\top \sim \mathcal{N}(0, K_Z^{(n)})$$

where again $K_Z^{(n)}$ is not a diagonal matrix.
- Also, we have $Y_{1:n} = X_{1:n} + Z_{1:n}$ but not $X_{1:n} \perp\!\!\!\!\!\!\!\perp Z_{1:n}$ due to feedback (i.e., independence doesn’t hold in this case).
Feedback

- For discrete memoryless channel, feedback does not increase capacity (since in finding $C$, you are already exploiting the randomness to the extent possible).
- For discrete non-memoryless channels, feedback could help (but not by too much).
- Here, the analogy is that the noise is correlated, i.e.,

$$Z_1:n^\top \sim \mathcal{N}(0, K_{Z}^{(n)})$$

where again $K_{Z}^{(n)}$ is not a diagonal matrix.

- Also, we have $Y_1:n = X_1:n + Z_1:n$ but not $X_1:n \perp \perp Z_1:n$ due to feedback (i.e., independence doesn’t hold in this case).

- Since noise is correlated, can use previous received values to help predict the noise.
**Feedback**

**Definition 5.1**

We define a $(2^{nR}, n)$ code as a set of mappings of the form

$$x_i(w, Y_{1:i-1}) : \{1, 2, \ldots, 2^{nR}\} \times \mathbb{R}^{i-1} \rightarrow \mathbb{R}$$

where $w \in \{1, 2, \ldots, 2^{nR}\}$ is the input message and $Y_{1:i-1}$ is the sequence of past output values (feedback). We think of $x_i(w, \cdot)$ this as a “code function” rather than a mere codeword.

- We also have a power constraint of the form

$$E \left[ \frac{1}{n} \sum_{i=1}^{n} x_i^2(w, Y_{1:i-1}) \right] \leq P \text{ for all } w \in \{1, 2, \ldots, 2^{nR}\}$$

(64)
Feedback

- We already saw that

\[ C_n = \frac{1}{2} \sum_{i=1}^{n} \log \left( 1 + \frac{(\nu - \lambda_i)^+}{\lambda_i} \right) \]  \text{bits per transmission}  \quad (65)  

is the capacity for colored noise channel w/o feedback.
Feedback

- We already saw that

\[ C_n = \frac{1}{2} \sum_{i=1}^{n} \log(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i}) \]  \text{bits per transmission}  \quad (65)

is the capacity for colored noise channel w/o feedback.

- We want to see, in various different ways, how much better \( C_{n,FB} \) (the capacity with colored noise and feedback) is going to be.
We already saw that

\[ C_n = \frac{1}{2} \sum_{i=1}^{n} \log\left(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i}\right) \text{ bits per transmission} \quad (65) \]

is the capacity for colored noise channel w/o feedback.

We want to see, in various different ways, how much better \( C_{n,FB} \) (the capacity with colored noise and feedback) is going to be.

Feedback is useful if we have channel asymmetric communication costs or abilities. E.g.,
Feedback

We already saw that

\[ C_n = \frac{1}{2} \sum_{i=1}^{n} \log(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i}) \]  
bits per transmission \hspace{1cm} (65)

is the capacity for colored noise channel w/o feedback.

We want to see, in various different ways, how much better \( C_{n,FB} \) (the capacity with colored noise and feedback) is going to be.

Feedback is useful if we have channel asymmetric communication costs or abilities. E.g.,

- Cheap to communicate in one direction, expensive to communicate in the other direction (e.g., Mars rover, some sensor networks, internet service providers).
Feedback

- We already saw that

\[
C_n = \frac{1}{2} \sum_{i=1}^{n} \log \left( 1 + \frac{(\nu - \lambda_i)^+}{\lambda_i} \right) \text{ bits per transmission} \quad (65)
\]

is the capacity for colored noise channel w/o feedback.

- We want to see, in various different ways, how much better \( C_{n,FB} \) (the capacity with colored noise and feedback) is going to be.

- Feedback is useful if we have channel asymmetric communication costs or abilities. E.g.,
  - Cheap to communicate in one direction, expensive to communicate in the other direction (e.g., Mars rover, some sensor networks, internet service providers).
  - Feasible to consider case when feedback is partly noisy (but asymmetries exist in the noise in different directions).
Bayesian Network View

Top: without feedback and uncorrelated noise
Bottom: without feedback and correlated noise
Bayesian Network View

With feedback and correlated noise:

Note, \( Z_i = Y_i - X_i \) which means that

\[
I(X_1:n; Y_1:n) = h(Y_1:n) - h(Y_1:n|X_1:n)
\]

\[
= h(X_1:n + Z_1:n) - h(Z_1:n)
\]

max when Gaussian

Gaussian

(66)

(67)
Bayesian Network View

With feedback and correlated noise:

\[ Z_i = Y_i - X_i \]

which means that

\[ I(X_1: n; Y_1: n) = h(Y_1: n) - h(Y_1: n | X_1: n) \]

\[ = h(X_1: n + Z_1: n) - h(Z_1: n) \]

(66)

(67)

max when Gaussian

Gaussian

\[ H(Y_1 | X_1, Z_1) = 0 \]

\[ H(Y_1 | X_1, Z_1) = 0 \]

Note, \( Z_i = Y_i - X_i \) which means that

So, max is achieved when \( X_1: n \) is also Gaussian, and in fact can be jointly Gaussian (exercise: show this).
Knowing previous $X$ value does not help to code any better than knowing previous $Y$ values. I.e.,

\[ X_i = f(w, Y_{1:i-1}) = f(w, X_{1:i-1}, Y_{1:i-1}) = f(Z_{1:i-1}) \]  

which follows since everything is jointly Gaussian, and $Z = Y - X$, and since what is required to deduce $X_i$ from $X_{1:i-1}, Y_{1:i-1}$ when $Y = X + Z$ is this difference $Y - X$. 
Knowing previous $X$ value does not help to code any better than knowing previous $Y$ values. I.e.,

$$X_i = f(w, Y_{1:i-1}) = f(w, X_{1:i-1}, Y_{1:i-1}) = f(Z_{1:i-1}) \quad (68)$$

which follows since everything is jointly Gaussian, and $Z = Y - X$, and since what is required to deduce $X_i$ from $X_{1:i-1}, Y_{1:i-1}$ when $Y = X + Z$ is this difference $Y - X$.

I.e., all $X_i$’s can be written w.l.o.g. as

$$X_i = \sum_{j=1}^{i-1} b_{ij} Z_j + V_i \text{ with } V_i \perp Z_j \quad \forall i, j \quad (69)$$
Knowing previous $X$ value does not help to code any better than knowing previous $Y$ values. I.e.,

$$X_i = f(w, Y_{1:i-1}) = f(w, X_{1:i-1}, Y_{1:i-1}) = f(Z_{1:i-1}) \tag{68}$$

which follows since everything is jointly Gaussian, and $Z = Y - X$, and since what is required to deduce $X_i$ from $X_{1:i-1}, Y_{1:i-1}$ when $Y = X + Z$ is this difference $Y - X$.

I.e., all $X_i$’s can be written w.l.o.g. as

$$X_i = \sum_{j=1}^{i-1} b_{ij} Z_j + V_i \text{ with } V_i \perp Z_j \quad \forall i, j \tag{69}$$

Or $X_{1:n} = BZ_{1:n} + V_{1:n}$ with $B$ a strictly lower diagonal matrix with $B_{ii} = 0$ for all $i$. 

\---

**Prof. Jeff Bilmes**

feedback and correlated noise

- Note that $B$ controls the feedback — if $B = 0$ then no feedback.
feedback and correlated noise

- Note that $B$ controls the feedback — if $B = 0$ then no feedback.
- This is a general form of conditional Gaussian, where $V$ is also Gaussian and independent of $Z$. 

\[
\text{pred}(B) = \text{tr}(BKZB^\top + KV) \leq nP.
\]
feedback and correlated noise

- Note that $B$ controls the feedback — if $B = 0$ then no feedback.
- This is a general form of conditional Gaussian, where $V$ is also Gaussian and independent of $Z$.
- So we can write this capacity as the following optimization:

$$C_{n,FB} = \max_{\frac{1}{n} \text{tr}(K_X) \leq P} \frac{1}{2n} \log \frac{|K_{X+Z}|}{|K_Z|}$$

(70)

where $K_{X+Z} = E[(X + Z)(X + Z)^\top]$

$$= \max_{K_V \succeq 0} \frac{1}{2n} \log \frac{|(B + I)K_Z(B + I)^\top + K_V|}{|K_Z|}$$

(71)

where $\text{pred}(B) = \text{true}$

where $\text{pred}(B) = \text{tr}(BK_ZB^\top + K_V) \leq nP$. Justification on next slide.
feedback and correlated noise

This follows since when $X = BZ + V$ with $Z \perp V$ we have

$$K_X = EXX^\top = E[(BZ + V)(BZ + V)^\top] = BK_ZB^\top + K_V$$  \hspace{1cm} (72)

and

$$E[(X + Z)(X + Z)^\top] \hspace{1cm} (73)$$

$$\quad = E[(BZ + V + Z)(BZ + V + Z)^\top] \hspace{1cm} (74)$$

$$\quad = E[((B + I)Z + V)((B + I)Z + V)^\top] \hspace{1cm} (75)$$

$$\quad = E[(B + I)ZZ^\top(B + I)^\top + (B + I)ZV^\top$$

$$\quad \quad \quad + V[(B + I)Z]^\top + VV^\top] \hspace{1cm} (76)$$

$$\quad = (B + I)K_Z(B + I)^\top + K_V \hspace{1cm} (78)$$
feedback and correlated noise

**Theorem 5.2**

*For a Gaussian channel with feedback, the rate $R_n$ for any sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \rightarrow 0$ satisfies $R_n \leq C_{n,FB} + \epsilon_n$ when $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.*

**Proof.**

- Assume $W$ is uniform over $2^{nR}$.
- We have Markov chain
  \[ W \rightarrow X_1:n \rightarrow Y_1:n \rightarrow \hat{W} \]  
  (79)
- Thus, Fano’s inequality still holds, i.e.
  \[ H(W|\hat{W}) \leq 1 + nR_n P_e^{(n)} = n \left( \frac{1}{n} + R_n P_e^{(n)} \right) = n\epsilon_n \]  
  (80)

  where $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$ (which means $P_e^{(n)} \rightarrow 0$).
feedback and correlated noise

... proof continued.

\[ nR_n \]
feedback and correlated noise

... proof continued.

\[ n R_n = H(W) \]
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \]
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\varepsilon_n \]  (81)
feedback and correlated noise

... proof continued.

\( nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\epsilon_n \) (81)

\( \leq I(W; Y^n) + n\epsilon_n \)
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\epsilon_n \]  \hspace{1cm} (81)

\[ \leq I(W; Y^n) + n\epsilon_n = \sum_i I(W; Y_i|Y_{1:i-1}) + n\epsilon_n \]  \hspace{1cm} (82)
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\epsilon_n \]  \hspace{566pt} (81)

\[ \leq I(W; Y^n) + n\epsilon_n = \sum_i I(W; Y_i|Y_{1:i-1}) + n\epsilon_n \]  \hspace{108pt} (82)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1})] + n\epsilon_n \]  \hspace{8pt} (83)

\[ \geq h(Y_{1:n}) - h(Z_{1:n}) + n\epsilon_n \]  \hspace{893pt} (88)
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\epsilon_n \]  \hspace{1cm} (81)

\[ \leq I(W; Y^n) + n\epsilon_n = \sum_i I(W; Y_i|Y_{1:i-1}) + n\epsilon_n \]  \hspace{1cm} (82)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1})] + n\epsilon_n \]  \hspace{1cm} (83)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1})] + n\epsilon_n \]  \hspace{1cm} (84)
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\epsilon_n \quad (81) \]

\[ \leq I(W; Y^n) + n\epsilon_n = \sum_i I(W; Y_i|Y_{1:i-1}) + n\epsilon_n \quad (82) \]

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1})] + n\epsilon_n \quad (83) \]

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1})] + n\epsilon_n \quad (84) \]

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1}, Z_{1:i-1})] + n\epsilon_n \quad (85) \]

\[ \leq I(W; Y^n) + n\epsilon_n = \sum_i I(W; Y_i|Y_{1:i-1}) + n\epsilon_n \quad (82) \]

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}) + n\epsilon_n] \quad (86) \]

\[ = h(Y_1|Y_1^n) - h(Z_1|Z_1^n) + n\epsilon_n \quad (88) \]
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\epsilon_n \] (81)

\[ \leq I(W; Y^n) + n\epsilon_n = \sum_i I(W; Y_i|Y_{1:i-1}) + n\epsilon_n \] (82)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1})] + n\epsilon_n \] (83)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1})] + n\epsilon_n \] (84)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1}, Z_{1:i-1})] + n\epsilon_n \] (85)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Z_i|W, Y_{1:i-1}, X_{1:i-1}, Z_{1:i-1})] + n\epsilon_n \] (86)
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\epsilon_n \] (81)

\[ \leq I(W; Y^n) + n\epsilon_n = \sum_i I(W; Y_i|Y_{1:i-1}) + n\epsilon_n \] (82)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1})] + n\epsilon_n \] (83)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1})] + n\epsilon_n \] (84)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1}, Z_{1:i-1})] + n\epsilon_n \] (85)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Z_i|W, Y_{1:i-1}, X_{1:i-1}, Z_{1:i-1})] + n\epsilon_n \] (86)

\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Z_i|Z_{1:i-1})] + n\epsilon_n \] (87)

\[ = h(Y_1:n) - h(Z_{1:n}) + n\epsilon_n \] (88)
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\epsilon_n \]  
\[ \leq I(W; Y^n) + n\epsilon_n = \sum_i I(W; Y_i|Y_{1:i-1}) + n\epsilon_n \]  
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1})] + n\epsilon_n \]  
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1})] + n\epsilon_n \]  
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1}, Z_{1:i-1})] + n\epsilon_n \]  
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Z_i|W, Y_{1:i-1}, X_{1:i-1}, Z_{1:i-1})] + n\epsilon_n \]  
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Z_i|Z_{1:i-1})] + n\epsilon_n \]  
\[ = h(Y_{1:n}) - h(Z_{1:n}) + n\epsilon_n \]
...proof continued.

So

\[
R_n \leq \frac{1}{n} [h(Y_{1:n}) - h(Z_{1:n})] + \epsilon_n \tag{89}
\]

\[
\leq \frac{1}{2n} \log \frac{|K_Y|}{|K_Z|} + \epsilon_n \tag{90}
\]

since \( h(Y_{1:n}) \) is maximized when \( Y_{1:n} \) is jointly Gaussian.
feedback and correlated noise

... proof continued.

- So

\[
R_n \leq \frac{1}{n} [h(Y_1:n) - h(Z_1:n)] + \epsilon_n \tag{89}
\]

\[
\leq \frac{1}{2n} \log \left| \frac{|K_Y|}{|K_Z|} \right| + \epsilon_n \tag{90}
\]

since \(h(Y_1:n)\) is maximized when \(Y_1:n\) is jointly Gaussian.

- Next, we are going to prove a number of very useful matrix inequality theorems that are important in their own right, but that are also going to be useful to compare the capacity with and without feedback and we’ll find that feedback doesn’t help that much.
feedback and correlated noise

... proof continued.

So

\[ R_n \leq \frac{1}{n} [h(Y_{1:n}) - h(Z_{1:n})] + \epsilon_n \]  \hspace{1cm} (89)
\[ \leq \frac{1}{2n} \log \left| \frac{|K_Y|}{|K_Z|} \right| + \epsilon_n \]  \hspace{1cm} (90)

since \( h(Y_{1:n}) \) is maximized when \( Y_{1:n} \) is jointly Gaussian.

Next, we are going to prove a number of very useful matrix inequality theorems that are important in their own right, but that are also going to be useful to compare the capacity with and without feedback and we’ll find that feedback doesn’t help that much.

See also chapter 17 in the book (2nd edition).
Matrix Inequalities

**Theorem 5.3**

∀ \(X, Z\) \(n\)-D random variables (not necessarily Gaussian), we have

\[
K_{X+Z} + K_{X-Z} = 2K_X + 2K_Z
\]  

(91)

**Proof.**

\[
K_{X+Z} = E(X + Z)(X + Z)^T = EXX^T + EXZ^T + EZX^T + EZZ^T
\]

\[= K_X + K_{XZ} + K_{ZX} + K_Z\]  

(92)

\[
K_{X-Z} = E(X - Z)(X - Z)^T = EXX^T - EXZ^T - EZX^T + EZZ^T
\]

\[= K_X - K_{XZ} - K_{ZX} + K_Z\]  

(93)

sum up both to get the result.
Matrix Inequalities

Theorem 5.4

If $A \succeq 0$, $B \succeq 0$, and $A - B \succeq 0$, then $|A| \geq |B|$

Proof.

- Let $C = A - B$, $X_1 \sim \mathcal{N}(0, B)$, $X_2 \sim \mathcal{N}(0, C)$, $X_1 \perp \perp X_2$. 
Matrix Inequalities

Theorem 5.4

If \( A \succeq 0, \ B \succeq 0, \text{ and } A - B \succeq 0 \), then \(|A| \geq |B|\).

Proof.

- Let \( C = A - B \), \( X_1 \sim \mathcal{N}(0, B) \), \( X_2 \sim \mathcal{N}(0, C) \), \( X_1 \perp \perp X_2 \).
- If \( Y = X_1 + X_2 \) then \( Y \sim \mathcal{N}(0, B + C) = \mathcal{N}(0, A) \).
Matrix Inequalities

Theorem 5.4

If \( A \succeq 0, \ B \succeq 0, \) and \( A - B \succeq 0, \) then \(|A| \geq |B|\)

Proof.

Let \( C = A - B, \ X_1 \sim \mathcal{N}(0, B), \ X_2 \sim \mathcal{N}(0, C), \ X_1 \perp \perp X_2. \)

If \( Y = X_1 + X_2 \) then \( Y \sim \mathcal{N}(0, B + C) = \mathcal{N}(0, A). \)

Then we get

\[
h(Y) \geq h(Y|X_2) = h(X_1|X_2) = h(X_1)
\]

(94)
Theorem 5.4

If $A \succeq 0$, $B \succeq 0$, and $A - B \succeq 0$, then $|A| \geq |B|

Proof.

- Let $C = A - B$, $X_1 \sim \mathcal{N}(0, B)$, $X_2 \sim \mathcal{N}(0, C)$, $X_1 \perp \perp X_2$.
- If $Y = X_1 + X_2$ then $Y \sim \mathcal{N}(0, B + C) = \mathcal{N}(0, A)$.
- Then we get

$$h(Y) \geq h(Y|X_2) = h(X_1|X_2) = h(X_1) \quad (94)$$

- Then, using the entropy of Gaussian formula,

$$\frac{1}{2} \log((2\pi e)^n |A|) \geq \frac{1}{2} \log((2\pi e)^n |B|) \quad (95)$$

and the result follows
Matrix Inequalities

Theorem 5.5

Let $X, Z$ be arbitrary $n$-D random variables. Then

$$|K_{X+Z}| \leq 2^n |K_X + K_Z|$$  \hspace{1cm} (96)

Proof.

$$2(K_X + K_Z) = K_{X+Z} + K_{X-Z}$$  \hspace{1cm} (97)

$$|K_{X+Z}| \leq |2(K_X + K_Z)| = 2^n |K_X + K_Z|$$  \hspace{1cm} (100)
Matrix Inequalities

**Theorem 5.5**

Let $X, Z$ be arbitrary $n$-D random variables. Then

$$|K_{X+Z}| \leq 2^n |K_X + K_Z| \quad (96)$$

**Proof.**

$$2(K_X + K_Z) = K_{X+Z} + K_{X-Z} \quad (97)$$

$$\Rightarrow 2(K_X + K_Z) - K_{X+Z} = K_{X-Z} \geq 0 \quad (98)$$

$$\Rightarrow 2^n |K_X + K_Z| \quad (100)$$
Matrix Inequalities

**Theorem 5.5**

*Let $X, Z$ be arbitrary $n$-D random variables. Then*

$$|K_{X+Z}| \leq 2^n |K_X + K_Z|$$  \hspace{1cm} (96)

**Proof.**

\begin{align*}
2(K_X + K_Z) &= K_{X+Z} + K_{X-Z} \\
\Rightarrow 2(K_X + K_Z) - K_{X+Z} &= K_{X-Z} \geq 0 \\
\Rightarrow K_{X+Z} &\leq 2(K_X + K_Z)
\end{align*}

\hspace{1cm} (97) \hspace{1cm} (98) \hspace{1cm} (99) \hspace{1cm} (100)
Matrix Inequalities

Theorem 5.5

Let $X, Z$ be arbitrary $n$-D random variables. Then

$$|K_{X+Z}| \leq 2^n |K_X + K_Z| \tag{96}$$

Proof.

$$2(K_X + K_Z) = K_{X+Z} + K_{X-Z} \tag{97}$$

$$\Rightarrow 2(K_X + K_Z) - K_{X+Z} = K_{X-Z} \geq 0 \tag{98}$$

$$\Rightarrow K_{X+Z} \leq 2(K_X + K_Z) \tag{99}$$

$$\Rightarrow |K_{X+Z}| \leq |2(K_X + K_Z)| = 2^n |K_X + K_Z| \tag{100}$$
Theorem 5.6

Let $A \succeq 0$, $B \succeq 0$, and $0 \leq \lambda \leq 1$, then

$$|\lambda A + (1 - \lambda)B| \geq |A|^{\lambda}|B|^{1-\lambda} \quad (101)$$

or in other words, $\log(|\lambda A + (1 - \lambda)B|) \geq \lambda \log |A| + (1 - \lambda) \log |B|$ or log-determinant is concave for non-negative definite matrices.

Proof.

- Let $X \sim \mathcal{N}(0, A)$, $Y \sim \mathcal{N}(0, B)$

...
Matrix Inequalities

Theorem 5.6

Let $A \succeq 0$, $B \succeq 0$, and $0 \leq \lambda \leq 1$, then

$$|\lambda A + (1 - \lambda)B| \geq |A|^\lambda |B|^{1-\lambda}$$

or in other words, $\log(|\lambda A + (1 - \lambda)B|) \geq \lambda \log |A| + (1 - \lambda) \log |B|$ or log-determinant is concave for non-negative definite matrices.

Proof.

- Let $X \sim \mathcal{N}(0, A)$, $Y \sim \mathcal{N}(0, B)$

- Let $Z = X1(\theta = 1) + Y1(\theta = 0)$ with $X \perp \perp Y$ and where $P(\theta = 1) = 1 - P(\theta = 0) = \lambda$. 

...
Matrix Inequalities

Theorem 5.6

Let $A \succeq 0$, $B \succeq 0$, and $0 \leq \lambda \leq 1$, then

$$|\lambda A + (1 - \lambda)B| \geq |A|^\lambda |B|^{1-\lambda}$$

(101)

or in other words, $\log(|\lambda A + (1 - \lambda)B|) \geq \lambda \log |A| + (1 - \lambda) \log |B|$ or log-determinant is concave for non-negative definite matrices.

Proof.

- Let $X \sim \mathcal{N}(0, A)$, $Y \sim \mathcal{N}(0, B)$
- Let $Z = X \mathbf{1}(\theta = 1) + Y \mathbf{1}(\theta = 0)$ with $X \perp \perp Y$ and where $P(\theta = 1) = 1 - P(\theta = 0) = \lambda$. Note $Z$ is not necessarily Gaussian.
Matrix Inequalities

Theorem 5.6

Let $A \succeq 0$, $B \succeq 0$, and $0 \leq \lambda \leq 1$, then

$$|\lambda A + (1 - \lambda)B| \geq |A|^\lambda |B|^{1-\lambda}$$

(101)

or in other words, $\log( |\lambda A + (1 - \lambda)B| ) \geq \lambda \log |A| + (1 - \lambda) \log |B|$ or log-determinant is concave for non-negative definite matrices.

Proof.

- Let $X \sim \mathcal{N}(0, A)$, $Y \sim \mathcal{N}(0, B)$
- Let $Z = X \mathbf{1}(\theta = 1) + Y \mathbf{1}(\theta = 0)$ with $X \perp \!\!\!\!\perp Y$ and where $P(\theta = 1) = 1 - P(\theta = 0) = \lambda$. Note $Z$ is not necessarily Gaussian.
- foo
Matrix Inequalities

... proof continued.

Then

Thus, we get

\[ (104) \]

\[ (107) \]
Then

\[ K_Z \]

Thus, we get

\[ \lambda h(X) + (1 - \lambda) h(Y) = \frac{1}{2} \log \left( \frac{2 \pi e}{n |\lambda A| (1 - \lambda) B} \right) \]
Matrix Inequalities

... proof continued.

- Then

\[ K_Z = EZZ^\top \]

Thus, we get

\[ \frac{1}{2} \log \left( \frac{2\pi e}{n} \right) \left| \lambda A + (1 - \lambda) B \right| \geq h(Z) \geq h(Z | \theta) \]

(104)

(107)
Then
\[ K_Z = EZZ^\top = E[(X1(\theta = 1) + Y1(\theta = 0))^2] \] (102)

Thus, we get
\[ \frac{1}{2} \log((2\pi e)^n |\lambda A + (1-\lambda)B|) \geq h(Z) \geq h(Z|\theta) \] (106)
Then

\[ K_Z = EZZ^\top = E[(X1(\theta = 1) + Y1(\theta = 0))``2`] \]  \label{eq:102}

\[ = E[XX^\top1(\theta = 1)1(\theta = 1)] + E[XY^\top1(\theta = 1)1(\theta = 0)] \]  \label{eq:103}

\[ + E[YX^\top1(\theta = 0)1(\theta = 1)] + E[YY^\top1(\theta = 1)1(\theta = 1)] \]  \label{eq:104}

Thus, we get

\[ \frac{1}{2} \log((2\pi e)^n|\lambda A + (1-\lambda)B|) \geq h(Z) \geq h(Z|\theta) \]  \label{eq:105}

\[ = \frac{1}{2} \log((2\pi e)^n|A|\lambda|B|1-\lambda) \]  \label{eq:106}
... proof continued.

Then

\[
K_Z = EZZ^\top = E[(X1(\theta = 1) + Y1(\theta = 0))^{\prime 2}]
\]

\[
= E[XX^\top 1(\theta = 1) 1(\theta = 1)] + E[XY^\top 1(\theta = 1) 1(\theta = 0)]
\]

\[
+ E[YX^\top 1(\theta = 0) 1(\theta = 1)] + E[YY^\top 1(\theta = 1) 1(\theta = 1)]
\]

\[
= \lambda A + (1 - \lambda)B
\]

Thus, we get

\[
\text{(107)}
\]
Matrix Inequalities

... proof continued.

Then

\[
K_Z = EZZ^\top = E[(X1(\theta = 1) + Y1(\theta = 0))^{\cdot 2}]
\]

\[
= E[XX^\top 1(\theta = 1)1(\theta = 1)] + E[XY^\top 1(\theta = 1)1(\theta = 0)]
\]

\[
+ E[YY^\top 1(\theta = 0)1(\theta = 1)] + E[YY^\top 1(\theta = 1)1(\theta = 1)]
\]

\[
= \lambda A + (1 - \lambda)B
\]

Thus, we get

\[
\frac{1}{2} \log((2\pi e)^n|\lambda A + (1 - \lambda)B|)
\]

(105)

(107)
proof continued.

Then

\[ K_Z = EZZ^\top = E[(X \mathbf{1}(\theta = 1) + Y \mathbf{1}(\theta = 0))^2] \]

\[ = E[XX^\top \mathbf{1}(\theta = 1) \mathbf{1}(\theta = 1)] + E[XY^\top \mathbf{1}(\theta = 1) \mathbf{1}(\theta = 0)] \]

\[ + E[YY^\top \mathbf{1}(\theta = 0) \mathbf{1}(\theta = 1)] + E[YY^\top \mathbf{1}(\theta = 1) \mathbf{1}(\theta = 1)] \]

\[ = \lambda A + (1 - \lambda)B \]

Thus, we get

\[ \frac{1}{2} \log((2\pi e)^n |\lambda A + (1 - \lambda)B|) \]

\[ \geq h(Z) \]

(102) (103) (104) (105)
... proof continued.

Then

\[
K_Z = EZ Z^\top = E[(X1(\theta = 1) + Y1(\theta = 0))^{\text{"2"}}] \\
= E[XX^\top 1(\theta = 1)1(\theta = 1)] + E[XY^\top 1(\theta = 1)1(\theta = 0)] \\
+ E[YX^\top 1(\theta = 0)1(\theta = 1)] + E[YY^\top 1(\theta = 1)1(\theta = 1)] \\
= \lambda A + (1 - \lambda) B
\]  

(102)

(103)

(104)

Thus, we get

\[
\frac{1}{2} \log((2\pi e)^n|\lambda A + (1 - \lambda) B|) \\
\geq h(Z) \geq h(Z|\theta)
\]

(105)

(106)

(107)
... proof continued.

Then

\[ K_Z = EZZ^\top = E[(X\mathbf{1}(\theta = 1) + Y\mathbf{1}(\theta = 0))^2] \]
\[ = E[XX^\top\mathbf{1}(\theta = 1)\mathbf{1}(\theta = 1)] + E[XY^\top\mathbf{1}(\theta = 1)\mathbf{1}(\theta = 0)] \]
\[ + E[YX^\top\mathbf{1}(\theta = 0)\mathbf{1}(\theta = 1)] + E[YY^\top\mathbf{1}(\theta = 1)\mathbf{1}(\theta = 1)] \]
\[ = \lambda A + (1 - \lambda)B \]  

Thus, we get

\[ \frac{1}{2} \log((2\pi e)^n |\lambda A + (1 - \lambda)B|) \geq h(Z) \geq h(Z|\theta) \]
\[ = \lambda h(X) + (1 - \lambda)h(Y) \]
Then

\[ K_Z = EZZ^\top = E[(X\mathbf{1}(\theta = 1) + Y\mathbf{1}(\theta = 0))^{\text{"2"}}] \]  
\[ = E[XX^\top\mathbf{1}(\theta = 1)\mathbf{1}(\theta = 1)] + E[XY^\top\mathbf{1}(\theta = 1)\mathbf{1}(\theta = 0)] \]
\[ + E[YX^\top\mathbf{1}(\theta = 0)\mathbf{1}(\theta = 1)] + E[YY^\top\mathbf{1}(\theta = 1)\mathbf{1}(\theta = 1)] \]
\[ = \lambda A + (1 - \lambda)B \]  

Thus, we get

\[ \frac{1}{2} \log((2\pi e)^n|\lambda A + (1 - \lambda)B|) \geq h(Z) \geq h(Z|\theta) \]
\[ = \lambda h(X) + (1 - \lambda)h(Y) = \frac{1}{2} \log((2\pi e)^n|A|^\lambda|B|^{1-\lambda}) \]