# An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

Aug. 29, 2016

## Lecture #12B: Parallel Gaussian Channel

## Outline of the lecture

- Parallel Gaussian channel



## Outline of the lecture

- Parallel Gaussian channel
- Channels with colored Gaussian noise

## Parallel Gaussian channels

- Consider k independent Gaussian channels in parallel with a common power constraint, $E \sum_{j=1}^{k} X_j^2 \leq P$



## Parallel Gaussian channels

- Consider k independent Gaussian channels in parallel with a common power constraint, $E \sum_{j=1}^{k} X_j^2 \leq P$
- The output of jth channel is given by

$$Y_j = X_j + Z_j, \quad j = 1, 2, \cdots, k,$$

with $Z_j \sim N(0, N_j)$ and independent from channel to channel.

## Parallel Gaussian channels

- Consider k independent Gaussian channels in parallel with a common power constraint, $E\sum_{j=1}^{k} X_j^2 \leq P$
- The output of jth channel is given by

$$Y_j = X_j + Z_j, \ \ j = 1, 2, \cdots, k,$$

with $Z_j \sim N(0, N_j)$ and independent from channel to channel.

- Interested in distributing the power among various channels so as to maximize the total channel capacity.



## Parallel Gaussian channels

- Consider k independent Gaussian channels in parallel with a common power constraint, $E\sum_{j=1}^{k} X_j^2 \leq P$
- The output of jth channel is given by

$$Y_j = X_j + Z_j, \ \ j = 1, 2, \cdots, k,$$

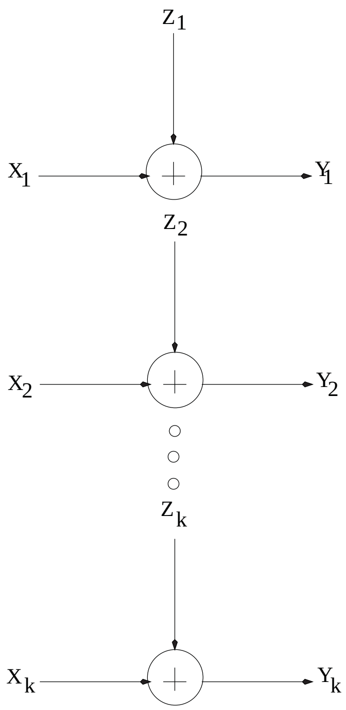
with $Z_j \sim N(0, N_j)$ and independent from channel to channel.

- Interested in distributing the power among various channels so as to maximize the total channel capacity.

- The capacity is given by

$$C = \max_{f(x_1, x_2, \cdots, x_k): \sum EX_i^2 \leq P} I(X_1, X_2, \cdots, X_k; Y_1, Y_2, \cdots, Y_k)$$

# Parallel Gaussian channels




Parallel Gaussian Channels



# Parallel Gaussian channels

- Since $Z_i's$ are independent

$$
\begin{aligned}
& I(X_1, X_2, \cdots, X_k; Y_1, Y_2, \cdots, Y_k) \\
= \; & h(Y_1, Y_2, \cdots, Y_k) - h(Y_1, Y_2, \cdots, Y_k | X_1, X_2, \cdots, X_k) \\
= \; & h(Y_1, Y_2, \cdots, Y_k) - h(Z_1, Z_2, \cdots, Z_k | X_1, X_2, \cdots, X_k) \\
= \; & h(Y_1, Y_2, \cdots, Y_k) - h(Z_1, Z_2, \cdots, Z_k) \\
= \; & h(Y_1, Y_2, \cdots, Y_k) - \sum_i h(Z_i) \\
\leq \; & \sum_i (h(Y_i) - h(Z_i)) \\
\leq \; & \sum_i \frac{1}{2} \log \left(1 + \frac{P_i}{N_i}\right)
\end{aligned}
$$

where $P_i = EX_i^2$ and $\sum P_i = P$.

# Parallel Gaussian channels

- Equality is achieved by

$$(X_1, X_2, \cdots, X_k) \sim N\left(0, \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_k \end{bmatrix}\right)$$



# Parallel Gaussian channels

- Equality is achieved by

$$(X_1, X_2, \cdots, X_k) \sim N\left(0, \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_k \end{bmatrix}\right)$$

- The problem of optimum power allocation reduces to finding the power allocation that maximizes capacity subject to the constraint that $\sum P_i = P$.

## Parallel Gaussian channels

- Equality is achieved by

$$(X_1, X_2, \cdots, X_k) \sim N\left(0, \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_k \end{bmatrix}\right)$$

- The problem of optimum power allocation reduces to finding the power allocation that maximizes capacity subject to the constraint that $\sum P_i = P$.
- This can be solved using Lagrange multiplier method. The Lagrangian can be written as

$$J(P_1, \cdots, P_k) = \sum \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right) + \lambda\left(\sum P_i\right)$$



## Parallel Gaussian channels

- Differentiating with respect to $P_i$, we get

$$\frac{1}{2}\frac{1}{P_i + N_i} + \lambda = 0$$

or

$$P_i = \nu - N_i$$

# Parallel Gaussian channels

- Differentiating with respect to $P_i$, we get

$$\frac{1}{2}\frac{1}{P_i + N_i} + \lambda = 0$$

or

$$P_i = \nu - N_i$$

- Since $P_i$ must be non-negative, we have the optimal power allocation as

$$P_i = (\nu - N_i)^+$$

where $\nu$ is chosen such that

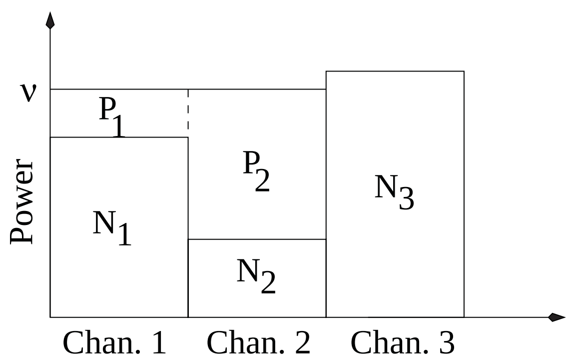
$$\sum (\nu - N_i)^+ = P$$

where

$$(x)^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

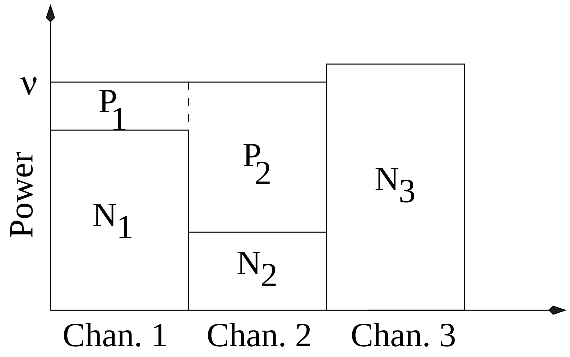


# Parallel Gaussian channels




- Vertical level indicates noise levels in various parallel channels.

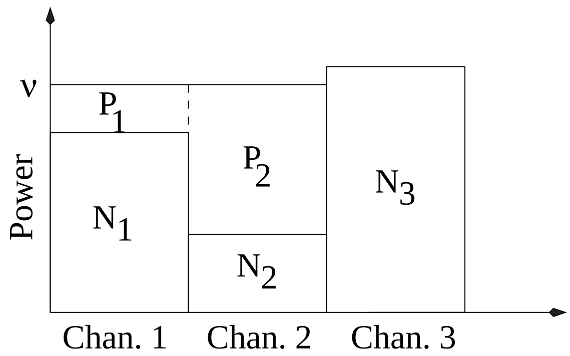
# Parallel Gaussian channels




- Vertical level indicates noise levels in various parallel channels.
- As power is increased from zero, we start allocating power to the channels with lowest noise.



# Parallel Gaussian channels




- Vertical level indicates noise levels in various parallel channels.
- As power is increased from zero, we start allocating power to the channels with lowest noise.
- This process of power allocation among various channels is identical to water distributing itself in a vessel, hence this is known as “waterfilling” method.

## Channels With Colored Gaussian Noise

- Let $K_Z$ be the covariance matrix of the noise, and let $K_X$ be the input covariance matrix. The power constraint on the input can then be written as

$$\frac{1}{n}\sum_i EX_i^2 \leq P$$

or equivalently,

$$\frac{1}{n}\mathrm{tr}(K_X) \leq P$$



## Channels With Colored Gaussian Noise

- Since the power constraint depends on $n$; the capacity will have to be calculated for each $n$.

## Channels With Colored Gaussian Noise

- Since the power constraint depends on $n$; the capacity will have to be calculated for each $n$.
- Just as in the case of independent channels, we can write

$$I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) = h(Y_1, Y_2, \ldots, Y_n) - h(Z_1, Z_2, \ldots, Z_n)$$



## Channels With Colored Gaussian Noise

- Since the power constraint depends on $n$; the capacity will have to be calculated for each $n$.
- Just as in the case of independent channels, we can write

$$I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) = h(Y_1, Y_2, \ldots, Y_n) - h(Z_1, Z_2, \ldots, Z_n)$$

- Here $h(Z_1, Z_2, \ldots, Z_n)$ is determined only by the distribution of the noise and is not dependent on the choice of input distribution.

## Channels With Colored Gaussian Noise

- So finding the capacity amounts to maximizing $h(Y_1, Y_2, \ldots, Y_n)$.



## Channels With Colored Gaussian Noise

- So finding the capacity amounts to maximizing $h(Y_1, Y_2, \ldots, Y_n)$.
- The entropy of the output is maximized when $Y$ is normal, which is achieved when the input is normal.

# Channels With Colored Gaussian Noise

- So finding the capacity amounts to maximizing $h(Y_1, Y_2, \ldots, Y_n)$.
- The entropy of the output is maximized when $Y$ is normal, which is achieved when the input is normal.
- Since the input and the noise are independent, the covariance of the output $Y$ is $K_Y = K_X + K_Z$ and the entropy is

$$h(Y_1, Y_2, \ldots, Y_n) = \frac{1}{2}\log((2\pi e)^n|K_X + K_Z|)$$



# Channels With Colored Gaussian Noise

- So finding the capacity amounts to maximizing $h(Y_1, Y_2, \ldots, Y_n)$.
- The entropy of the output is maximized when $Y$ is normal, which is achieved when the input is normal.
- Since the input and the noise are independent, the covariance of the output $Y$ is $K_Y = K_X + K_Z$ and the entropy is

$$h(Y_1, Y_2, \ldots, Y_n) = \frac{1}{2}\log((2\pi e)^n|K_X + K_Z|)$$

- Now the problem is reduced to choosing $K_X$ so as to maximize $|K_X + K_Z|$, subject to a trace constraint on $K_X$.

# Channels With Colored Gaussian Noise

- To do this, we decompose $K_Z$ into its diagonal form,

$$K_Z = Q\Lambda Q^t, \quad \text{where } QQ^t = I$$



# Channels With Colored Gaussian Noise

- To do this, we decompose $K_Z$ into its diagonal form,

$$K_Z = Q\Lambda Q^t, \quad \text{where } QQ^t = I$$

- Then

$$\begin{aligned} |K_X + K_Z| &= |K_X + Q\Lambda Q^t| \\ &= |Q||Q^t K_X Q + \Lambda||Q^t| \\ &= |Q^t K_X Q + \Lambda| \\ &= |A + \Lambda| \end{aligned}$$

where $A = Q^t K_X Q$

# Channels With Colored Gaussian Noise

- Since for any matrices $B$ and $C$,

$$\text{tr}(BC) = \text{tr}(CB)$$

we have

$$\begin{aligned} \text{tr}(A) &= \text{tr}(Q^t K_X Q) \\ &= \text{tr}(QQ^t K_X) \\ &= \text{tr}(K_X) \end{aligned}$$



# Channels With Colored Gaussian Noise

- Since for any matrices $B$ and $C$,

$$\text{tr}(BC) = \text{tr}(CB)$$

we have

$$\begin{aligned} \text{tr}(A) &= \text{tr}(Q^t K_X Q) \\ &= \text{tr}(QQ^t K_X) \\ &= \text{tr}(K_X) \end{aligned}$$

- Now the problem is reduced to maximizing $|A + \Lambda|$ subject to a trace constraint $\text{tr}(A) \leq nP$.

## Channels With Colored Gaussian Noise

- We apply Hadamards inequality which states that the determinant of any positive definite matrix $K$ is less than the product of its diagonal elements, that is,

$$|K| \leq \prod_i K_{ii}$$

with equality iff the matrix is diagonal.



## Channels With Colored Gaussian Noise

- We apply Hadamards inequality which states that the determinant of any positive definite matrix $K$ is less than the product of its diagonal elements, that is,

$$|K| \leq \prod_i K_{ii}$$

with equality iff the matrix is diagonal.

- Thus,

$$|A + \Lambda| \leq \prod_i (A_{ii} + \lambda_i)$$

with equality iff $A$ is diagonal.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

# Channels With Colored Gaussian Noise

- Since $A$ is subject to a trace constraint,

$$\frac{1}{n}\sum_i A_{ii} \leq P$$

and $A_{ii} \geq 0$, the maximum value of $\prod_i (A_{ii} + {}_i)$ is attained when

$$A_{ii} + \lambda_i = \nu$$



# Channels With Colored Gaussian Noise

- Since $A$ is subject to a trace constraint,

$$\frac{1}{n}\sum_i A_{ii} \leq P$$

and $A_{ii} \geq 0$, the maximum value of $\prod_i (A_{ii} + {}_i)$ is attained when

$$A_{ii} + \lambda_i = \nu$$

- However, given the constraints, it may not always be possible to satisfy this equation with positive $A_{ii}$. In such cases, we can show by the standard KuhnTucker conditions that the optimum solution corresponds to setting

$$A_{ii}(\nu - \lambda_i)^+,$$

where the water level $\nu$ is chosen so that $\sum A_{ii} = nP$.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

# Channels With Colored Gaussian Noise

- Since $A$ is subject to a trace constraint,

$$\frac{1}{n}\sum_i A_{ii} \leq P$$

and $A_{ii} \geq 0$, the maximum value of $\prod_i (A_{ii} + {}_i)$ is attained when

$$A_{ii} + \lambda_i = \nu$$

- However, given the constraints, it may not always be possible to satisfy this equation with positive $A_{ii}$. In such cases, we can show by the standard KuhnTucker conditions that the optimum solution corresponds to setting

$$A_{ii}(\nu - \lambda_i)^+,$$

where the water level $\nu$ is chosen so that $\sum A_{ii} = nP$.

- This value of $A$ maximizes the entropy of $Y$ and hence the mutual information.