

**Applied Optimization for Wireless, Machine Learning, Big Data**  
**Prof. Aditya K. Jagannatham**  
**Department of Electrical Engineering**  
**Indian Institute of Technology, Kanpur**

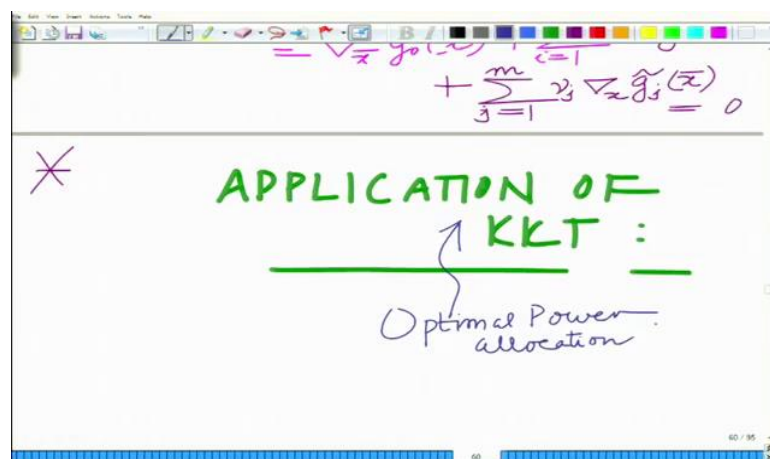
**Lecture - 67**

**Application of KKT conditions : Optimal MIMO Power allocation (Waterfilling)**

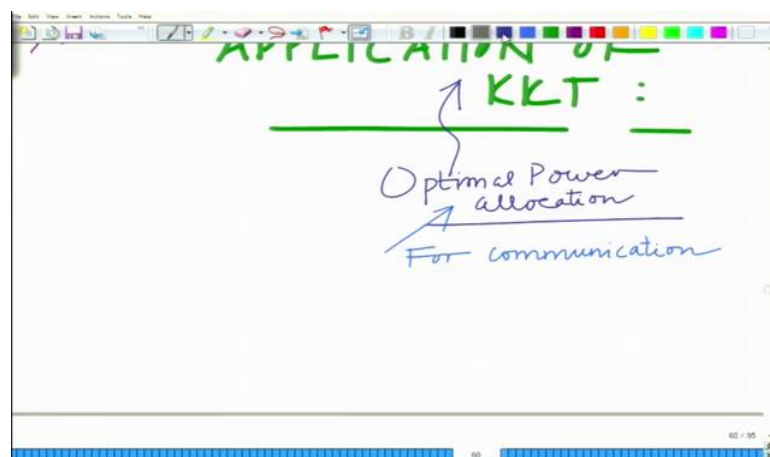
**Keywords:** Karush-Kuhn-Tucker (KKT) conditions, Optimal MIMO Power allocation Waterfilling Algorithm

Hello, welcome to another module in this massive open online course. So we have looked at the KKT conditions to solve an optimization problem. Let us look at an application to better understand how one can use the KKT conditions to solve an optimization problem.

(Refer Slide Time: 00:34)

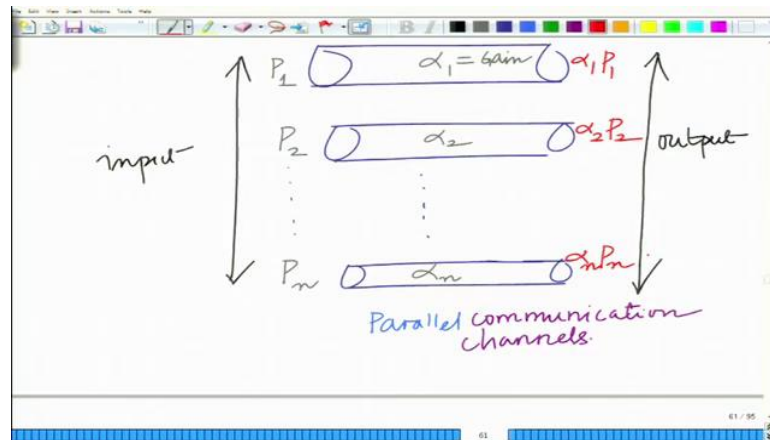


(Refer Slide Time: 01:14)



So let us say you have a set of parallel channels.

(Refer Slide Time: 01:35)



So let us assume that these are arranged in decreasing order of gains. So you can transmit at a certain bit rate over each of these communication channels and bit rate depends on the power that is allocated to that particular channel. So let us say the power allocated for first channel is  $P_1$ , second channel is  $P_2$  and so on and for  $n^{\text{th}}$  channel is  $P_n$ . Let us say the gain of first channel is  $\alpha_1$ , gain of channel 2 is  $\alpha_2$  and so on gain of channel  $n$  is  $\alpha_n$ . So as shown in the slide, this is the input and this is the output or you can think of it as a transmitter and the receiver. The received power across channel 1 will be  $\alpha_1 P_1$ , similarly across channel 2 will be  $\alpha_2 P_2$  and so on across channel  $n$  will be  $\alpha_n P_n$ . Now in addition for every communication at the receiver we will have thermal noise or Gaussian noise which is typically modelled as additive Gaussian noise.

(Refer Slide Time: 03:57)

The slide shows the equation for the received signal in channel  $i$ :  $y_i = \sqrt{\alpha_i} x_i + n_i$ . An arrow points from the noise term  $n_i$  to a note: 'Additive white Gaussian noise mean  $\Rightarrow 0$  var  $\Rightarrow \sigma^2$ '. Below this, the Signal to Noise Ratio (SNR) for channel  $i$  is defined as the 'Signal to Noise power ratio' and is given by the equation:  $\text{SNR} = \frac{\alpha_i P_i}{\sigma^2}$ .

So we have  $y_i = \sqrt{\alpha_i} x_i + n_i$  where this quantity  $n_i$  is the additive white Gaussian noise, with mean 0 and variance  $\sigma^2$ . So noise power is  $\sigma^2$  for each channel. The SNR for channel  $i$  is  $\frac{\alpha_i P_i}{\sigma^2}$ .

(Refer Slide Time: 05:29)

Handwritten notes on a whiteboard:

- At the top right, it says: "Additive white Gaussian noise mean = 0 var =  $\sigma^2$ ".
- Below that, "SNR: for channel  $i$ " is circled in purple.
- Next to it, "Signal to Noise power ratio" is written.
- The formula  $= \frac{\alpha_i P_i}{\sigma^2}$  is circled in green.
- Below the formula, it says "Maximum rate = Shannon channel capacity".
- Then,  $= \log_2(1 + \text{SNR})$  is written.
- Finally,  $= \log_2\left(1 + \frac{\alpha_i P_i}{\sigma^2}\right)$  is written.

And now the maximum information rate is given by the Shannon's formula for the capacity of the channel. So this is given as  $\log_2\left(1 + \frac{\alpha_i P_i}{\sigma^2}\right)$ .

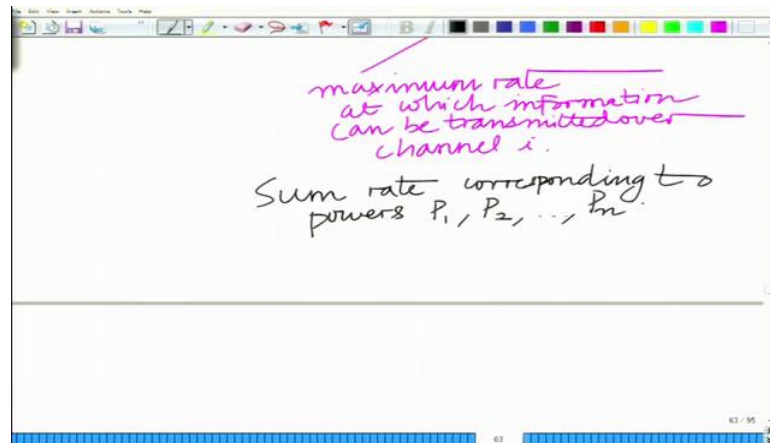
(Refer Slide Time: 06:37)

Handwritten notes on a whiteboard:

- At the top, "Maximum rate = Shannon channel capacity" is written.
- Below that,  $= \log_2(1 + \text{SNR})$  is written.
- Then,  $= \log_2\left(1 + \frac{\alpha_i P_i}{\sigma^2}\right)$  is written.
- An arrow points from this formula to the text: "maximum rate at which information can be transmitted over channel  $i$ :".

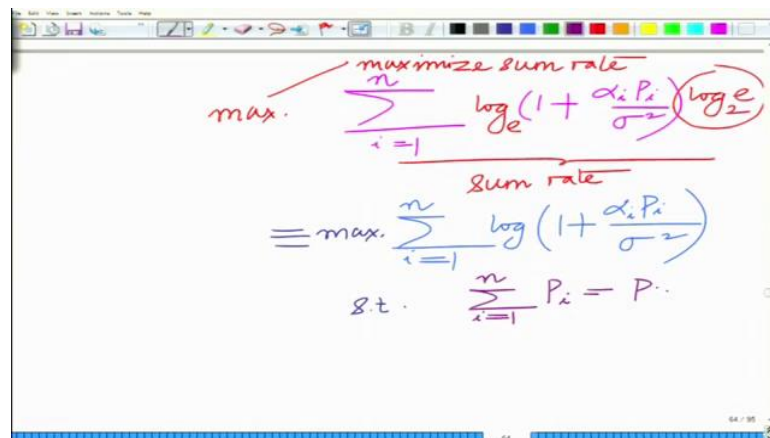
So this is the maximum rate at which information can be transmitted over the channel  $i$ . And therefore the maximum sum rate of information transmitted across all these  $n$  parallel channels will be given by the sum of the individual rates across each of these  $n$  parallel channels.

(Refer Slide Time: 07:43)



So the maximum sum rate corresponding to powers  $P_1, P_2, \dots, P_n$  is to be calculated.

(Refer Slide Time: 08:10)



We want to maximize this sum rate, so we have  $\max \sum_{i=1}^n \log_2 \left( 1 + \frac{\alpha_i P_i}{\sigma^2} \right)$ . Now we are making a minor modification here  $\max \sum_{i=1}^n \log_2 \left( 1 + \frac{\alpha_i P_i}{\sigma^2} \right) \log_2 e$ . So this then becomes simply the natural logarithm, so we have  $\max \sum_{i=1}^n \log \left( 1 + \frac{\alpha_i P_i}{\sigma^2} \right)$ , so instead of maximizing the objective function times a constant, we can simply ignore the constant factor. Now the constraint is that the total transmit power is a fixed quantity.

(Refer Slide Time: 10:27)

Handwritten notes on a whiteboard:

$$\equiv \max. \sum_{i=1}^n \log \left( 1 + \frac{\alpha_i P_i}{\sigma^2} \right)$$

s.t.  $\sum_{i=1}^n P_i = P$

maximum Tx Power of Transmitter

$$P_i \geq 0$$

$$\Rightarrow -P_i \leq 0$$

$$\bar{P} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix} \succeq 0$$

$$\max \sum_{i=1}^n \log \left( 1 + \frac{\alpha_i P_i}{\sigma^2} \right)$$

So the optimization problem is s.t.  $\sum_{i=1}^n P_i = P$  . This log is a concave function and

$$P_i \geq 0$$

the sum of log is also a concave function. So this is the maximization of concave objective function.

(Refer Slide Time: 12:01)

Handwritten notes on a whiteboard:

$$\equiv \min. - \sum_{i=1}^n \log \left( 1 + \frac{P_i \alpha_i}{\sigma^2} \right)$$

s.t.  $\sum_{i=1}^n P_i = P$

$$-P \leq 0$$

convex optimization problem for "optimal Power Allocation"

Powers are being allocated optimally to n channels

$$\mathcal{L}(\bar{P}, \bar{\lambda}, \bar{\nu})$$

$$\sum_{i=1}^n \alpha_i P_i$$

$$\max - \sum_{i=1}^n \log \left( 1 + \frac{\alpha_i P_i}{\sigma^2} \right)$$

So this can equivalently be written as  $s.t. \sum_{i=1}^n P_i = P$  . So this is the convex  
 $- P_i \leq 0$

optimization problem for optimal power allocation. You are allocating the powers optimally and hence it is termed as optimal power allocation.

(Refer Slide Time: 13:11)

$$\begin{aligned} \mathcal{L}(P, \lambda, \nu) &= - \sum_{i=1}^n \log \left( 1 + \frac{\alpha_i P_i}{\sigma^2} \right) \\ &+ \nu \left( \sum_{i=1}^n P_i - P \right) \\ &- \bar{\lambda}^T \bar{P} \\ \bar{\lambda} &= \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} \end{aligned}$$

Now we will use the KKT conditions to solve this, let us start with the Lagrangian. So we have  $L(\bar{P}, \bar{\lambda}, \nu) = \log \left( 1 + \frac{\alpha_i P_i}{\sigma^2} \right) + \nu \left( \sum_{i=1}^n P_i - P \right) - \bar{\lambda}^T \bar{P}$ . So you have one Lagrange multiplier for each inequality constraint.

(Refer Slide Time: 15:43)

$$\begin{aligned} \frac{\partial \mathcal{L}(P, \lambda, \nu)}{\partial P_i} &= 0 \\ \Rightarrow - \frac{1}{1 + \frac{\alpha_i P_i}{\sigma^2}} + \nu &= 0 \end{aligned}$$

Now  $\nabla_{\bar{P}} L(\bar{P}, \bar{\lambda}, \bar{\nu}) = 0$  is one of the KKT conditions. So on solving this we get

$$\nu = \frac{\frac{\alpha_i}{\sigma^2}}{1 + \frac{\alpha_i P_i}{\sigma^2}} + \lambda_i.$$

(Refer Slide Time: 16:35)

Handwritten derivation on a digital whiteboard showing the KKT condition for the dual variable  $\nu$ . The equations are:

$$1 + \frac{\alpha_i}{\sigma^2} - \lambda_i = 0$$

$$\nabla_{\bar{P}} L(\bar{P}, \bar{\lambda}, \bar{\nu}) = 0$$

$$\Rightarrow \frac{\alpha_i / \sigma^2}{1 + \frac{\alpha_i P_i}{\sigma^2}} + \lambda_i = \nu$$

(Refer Slide Time: 17:52)

Handwritten statement of Complementary Slackness on a digital whiteboard:

Complementary Slackness.

$$\lambda_i P_i = 0$$

$\Rightarrow$  if  $P_i > 0$  then  $\lambda_i = 0$  (non-negative power allocated to a channel).

And now from the complementary slackness we have  $\lambda_i P_i = 0$  that is either the constraint is slack or the Lagrange multiplier is slack, but not both. So let us consider these two conditions. If  $P_i > 0$ , that is power allocated to a channel is non-negative, then  $\lambda_i = 0$ .

(Refer Slide Time: 19:09)

Handwritten derivation on a whiteboard:

$$\lambda_i P_i = 0$$

non-negative power allocated to a channel.

$\Rightarrow$  if  $P_i > 0$  Then  $\lambda_i = 0$

$$\Rightarrow \frac{\alpha_i / \sigma^2}{1 + \alpha_i P_i / \sigma^2} = \nu$$

$$\Rightarrow \frac{1}{\nu} = \frac{\alpha_i}{1 + \frac{\alpha_i P_i}{\sigma^2}}$$

$$= \frac{\sigma^2}{\alpha_i} + P_i$$

So this implies that  $\frac{1}{\nu} = \frac{\sigma^2}{\alpha_i} + P_i$ .

(Refer Slide Time: 19:51)

Handwritten derivation on a whiteboard:

$$= \frac{\sigma^2}{\alpha_i} + P_i$$

$$\Rightarrow \boxed{P_i = \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i}}$$

Optimal power to  $i^{\text{th}}$  channel.

$$\Rightarrow \frac{1}{\nu} \geq \frac{\sigma^2}{\alpha_i}$$

$$\lambda_i = 0$$

Case 2: if  $\lambda_i > 0$

$$\Rightarrow P_i = 0$$

$$\Rightarrow \frac{\alpha_i / \sigma^2}{1 + \frac{\alpha_i P_i}{\sigma^2}} + \lambda_i = \nu$$

This implies that  $P_i = \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i} \Rightarrow \frac{1}{\nu} \geq \frac{\sigma^2}{\alpha_i}$  and the corresponding eigen value is 0. So this is the optimal power allocated to the  $i^{\text{th}}$  channel.



(Refer Slide Time: 20:58)

Case 2: if  $\lambda_i > 0$   
 $\Rightarrow P_i = 0$   
 $\Rightarrow \frac{\alpha_i/\sigma^2}{1+0} + \lambda_i = \nu$   
 $\Rightarrow \frac{\alpha_i}{\sigma^2} + \lambda_i = \nu$   
 $\Rightarrow \lambda_i = \nu - \frac{\alpha_i}{\sigma^2}$   
 $\lambda_i > 0$   
 $\Rightarrow \nu > \frac{\alpha_i}{\sigma^2}$

On the other hand, if you consider the case 2, if  $\lambda_i > 0$ , that is Lagrange multiplier is slack which implies that  $P_i = 0$ . Now this implies  $\frac{1}{\nu} < \frac{\sigma^2}{\alpha_i}$  as shown in the slide.

(Refer Slide Time: 22:14)

$\lambda_i > 0$   
 $\Rightarrow \nu > \frac{\alpha_i}{\sigma^2}$   
 $\Rightarrow \left[ \frac{1}{\nu} < \frac{\sigma^2}{\alpha_i} \right]$   
 Therefore,  
 $P_i = \begin{cases} \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i} & \text{if } \frac{1}{\nu} \geq \frac{\sigma^2}{\alpha_i} \\ 0 & \text{if } \frac{1}{\nu} < \frac{\sigma^2}{\alpha_i} \end{cases}$

So there are two cases and therefore if you summarize it we have

$$P_i = \begin{cases} \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i} & \text{if } \frac{1}{\nu} \geq \frac{\sigma^2}{\alpha_i} \\ 0 & \text{if } \frac{1}{\nu} < \frac{\sigma^2}{\alpha_i} \end{cases}.$$

(Refer Slide Time: 23:57)

$$P_i = \max \left\{ \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i}, 0 \right\}$$

$$P_i = \left( \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i} \right)^+$$

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

And therefore, you can write this  $P_i$  as  $P_i = \left( \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i} \right)^+$ . So  $P_i = \max \left\{ \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i}, 0 \right\}$ .

(Refer Slide Time: 25:28)

To find  $\nu$ , solve

$$\sum_{i=1}^n \left( \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i} \right)^+ = P.$$

Total Power Constraint

So to find  $\nu$  solve  $\sum_{i=1}^n \left( \frac{1}{\nu} - \frac{\sigma^2}{\alpha_i} \right)^+ = P$  that is the total power constraint.

(Refer Slide Time: 26:05)

Handwritten slide showing the total power constraint equation and channel ordering. The equation is  $\sum_{i=1}^n \left( \frac{1}{\gamma} - \frac{\sigma^2}{\alpha_i} \right) = P$ . A red arrow points from the equation to the text "Total Power Constraint". Below the equation, it says "Let  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$ " and "channel 1 = strongest" and "channel n = weakest".

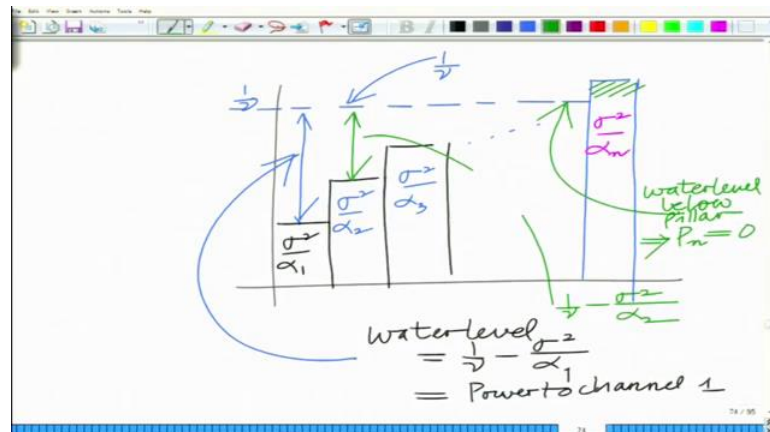
Let us let us assume that these are ordered as the first channel is the strongest and the last one is the weakest.

(Refer Slide Time: 26:54)

Handwritten slide showing the relationship between  $\alpha_i$  and power allocation. It starts with the term  $\frac{1}{\gamma} - \frac{\sigma^2}{\alpha_i}$  circled. Below it, it says "As  $\alpha_i$  increases,  $\frac{\sigma^2}{\alpha_i}$  decreases." and " $\Rightarrow \frac{1}{\gamma} - \frac{\sigma^2}{\alpha_i}$  increases." and " $\Rightarrow P_i$  increases." and " $\Rightarrow$  more power allocated to stronger channel."

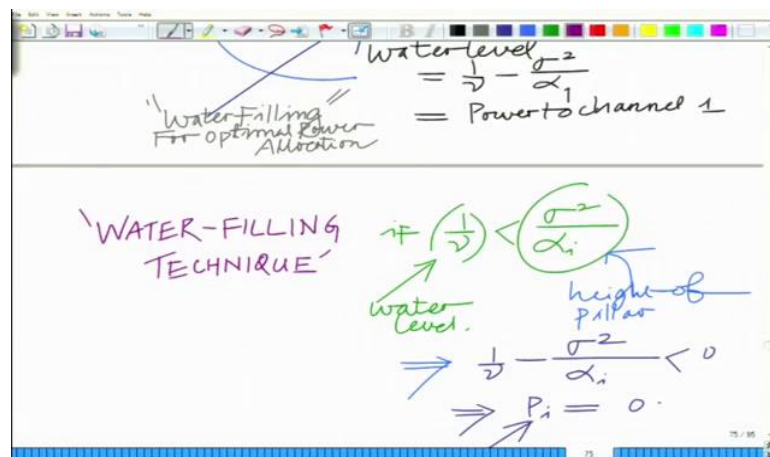
Now what this implies is that more power is allocated to the stronger channel as shown in slides.

(Refer Slide Time: 28:24)



Now, let us look at this representation, for instance a sort of bowl or you can call it an area with this kind of pillars. So the first pillar is corresponding to first channel and then it decreases. Now if you draw here the level  $\frac{1}{2}$ , you can think of this as a water level, now the power allocated to the first channel is basically the amount of water.

(Refer Slide Time: 30:40)



This is as shown in slide.

(Refer Slide Time: 31:27)

$$\frac{1}{\gamma} - \frac{\sigma^2}{\alpha_i} < 0$$
$$\Rightarrow P_i = 0$$

Power allocated to  $i$ th channel = 0

Annotations:  
-  $\frac{1}{\gamma}$  is labeled "water level."  
-  $\alpha_i$  is labeled "height of pillar."

Therefore, this scheme is known as the optimal water filling algorithm. So you can think of this as a water level. So it is a solution of a convex optimization problem derived or obtained using the KKT conditions and the complementary slackness plays a very key role. So this is a nice scheme or this is the optimal scheme to allocate power across the parallel channels that maximizes the sum rate of communication between the transmitter and the receiver. So we will stop here and continue in the subsequent module. Thank you very much.