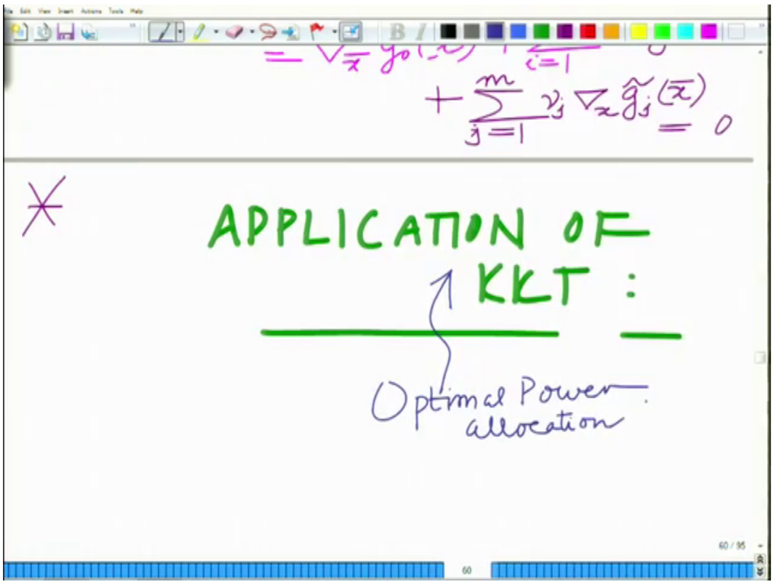
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture - 67
Application of KKT conditions : Optimal MIMO Power allocation (Waterfilling)

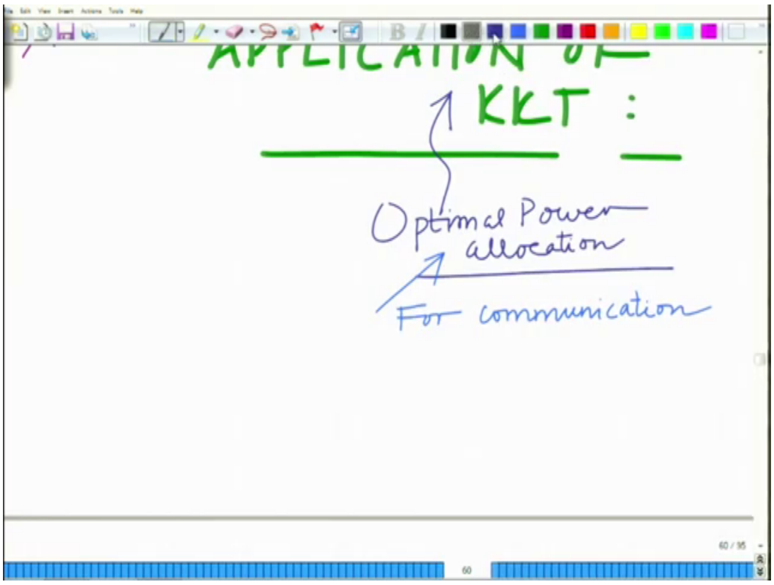
Hello, welcome to another module in this massive open online course. So, we have looked at the KKT conditions; so, this that is the Karush Kuhn Tucker conditions to solve an optimization problem. Let us looked at it and apply it; let us look at an application to better understand that better understand how can one how one can use the KKT conditions to solve an optimized to solve a practical optimization problem.

(Refer Slide Time: 00:34)



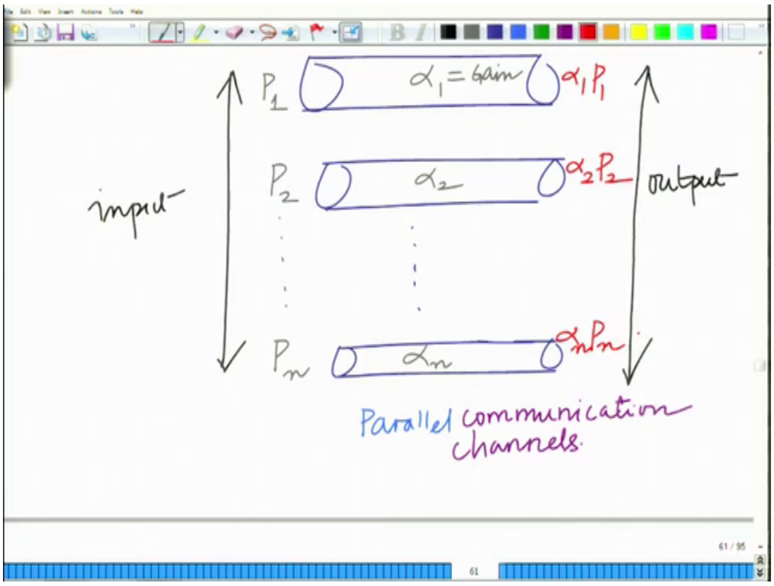
So, we have seen the KKT conditions; let us now look at an application of the KKT conditions and this we will the particular application that we will consider is for optimal power allocation optimal; there is a typically a problem that arises very frequently in communication systems ok.

(Refer Slide Time: 01:14)



So, this is optimal power allocation you can say for communication. Now what happens in communication is that let us say you have a set of frequently what happens is you have a set of parallel channels.

(Refer Slide Time: 01:35)

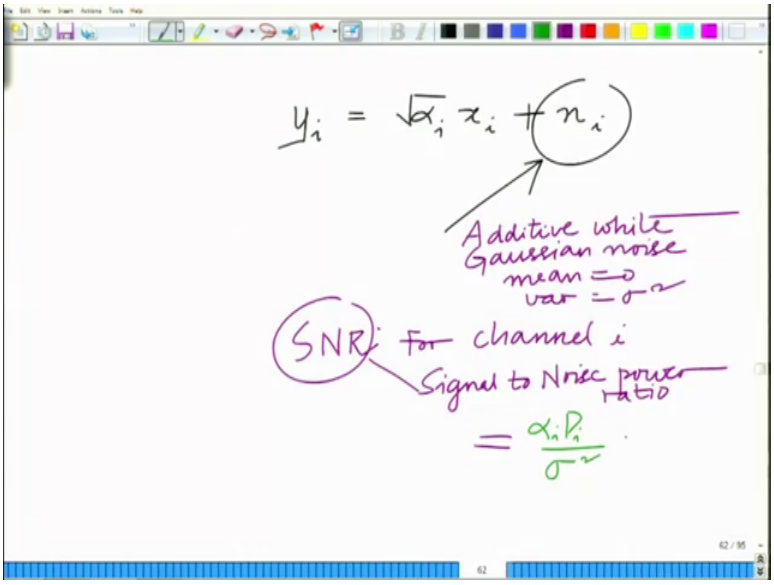


So, you have set of parallel pipes one can think of this and so you have channel 1, you have channel 2 and so on and so forth; you have channel let assume that these are arranged in decreasing order of gains. So, what these are? These are a set of parallel channels or let us think of parallel communication channels minute when I say a channel, it goes without saying that I am talking about a communication channel; so you have parallel communication channels.

And of course, you can transmit at certain bit rate over each of these communication channels right and bit rate depends on the power that is allocated to that particular channel. So, let us assume that let us say the power allocated for first channel is P 1 second channel is P 2 so on, your set of n parallel communication channels power allocated a spear. Let us say the gain of first one is alpha 1; this is the gain of channel 1, gain of channel 2 is alpha 2, gain of channel n is alpha l.

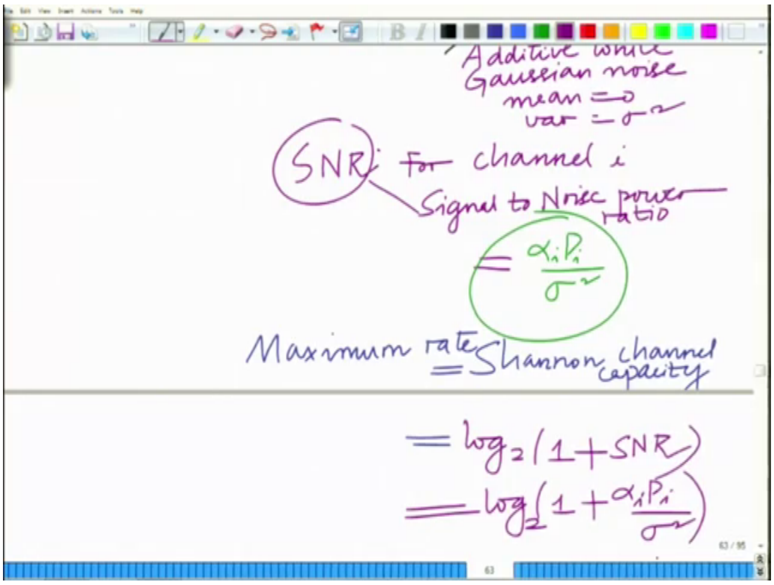
So, necessarily the output person this is the input correct; so, this side is basically information is flowing from left to right. So, this is the input and this is the output or you can think of it as left as a transmitter, the right as receiver. The received power will be of course, gain is a alpha 1, power will be alpha times P 1 transmitter. Similarly across channel 2 gain alpha 2 times power P 2 gain alpha n times power P n ok. So, what this means is your a set of n parallel channels; it cross each channel i; your the input power is p i; the gain is alpha i; so the output power is alpha i and P i. Now in addition for every communication at the receiver we will have thermal noise aright or Gaussian noise which typical modeled as additive Gaussian noise ok.

(Refer Slide Time: 03:57)



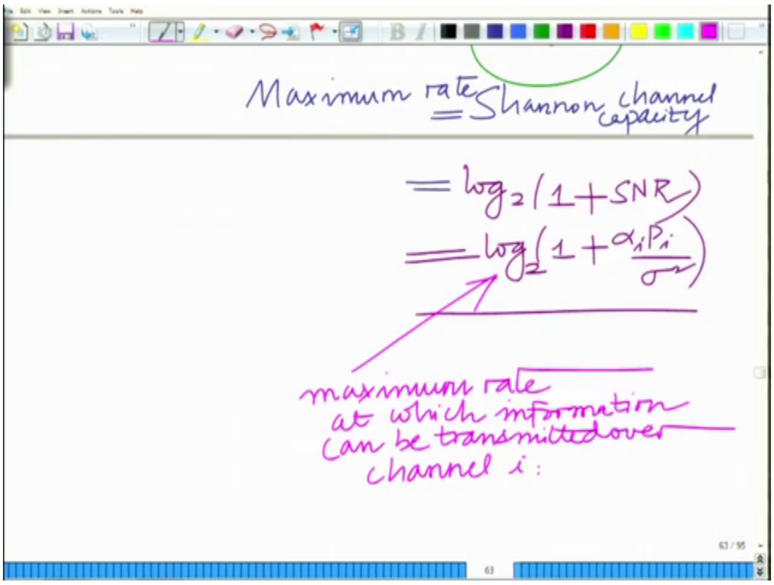
So, we have this channel that is y i equals you can think of it as square root of alpha i because alpha i is the gain in power square root of alpha i x i plus n i; where this quantity n i is the additive white Gaussian noise mean equals 0 variance equals sigma square ok. So, noise power is sigma square for each channel and so what we have is basically; if you look at the SNR for channel i SNR is nothing, but this is the signal 2 noise power ratio; this is equal to well the received power alpha P i divided by sigma square.

(Refer Slide Time: 05:29)



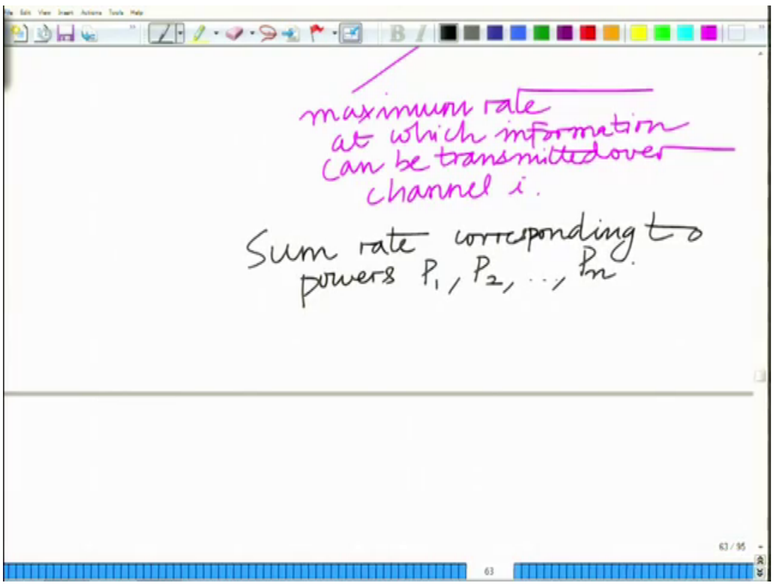
And now one can ask corresponded to this SNR; what is the maximum rate information rate or let the bit rate at which one can transmit over this channel and we have a convenient formula for that that is given by the Shannon; the Shannon's formula for the capacity of the channel or the maximum information maximum bit rate of the channel; so, from the Shannon capacity formula. So, the maximum is from the Shannon and that is basically given as log to the base 2 1 plus SNR; which for this is log to the base 2 alpha i.

(Refer Slide Time: 06:37)



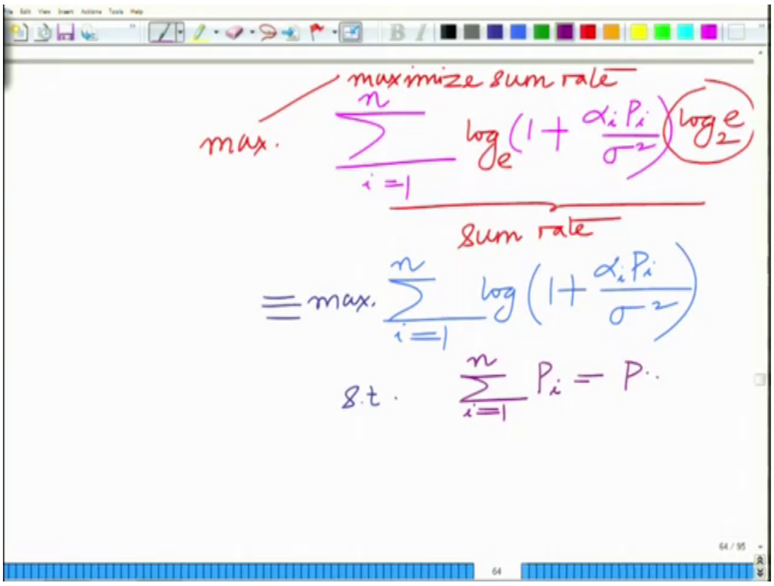
So, this is the maximum rate at which information can be transmitted over the; so this is the maximum rate; transmitted maximum rate at which information can be transmitted over channel i. And therefore, now what is the; so, this is the maximum rate at which information can be transmitted across channel i. Now, therefore, what is the sum rate; the maximum sum rate of information transmitted across all these n parallel channels? That will be given by the sum of the individual rate or the bit rate, some of the individual maximum rates across each of these n parallel channels.

(Refer Slide Time: 07:43)



So, the maximum sum rate or the sum rate corresponding to this power allocation corresponding to powers P 1, P 2, P n.

(Refer Slide Time: 08:10)



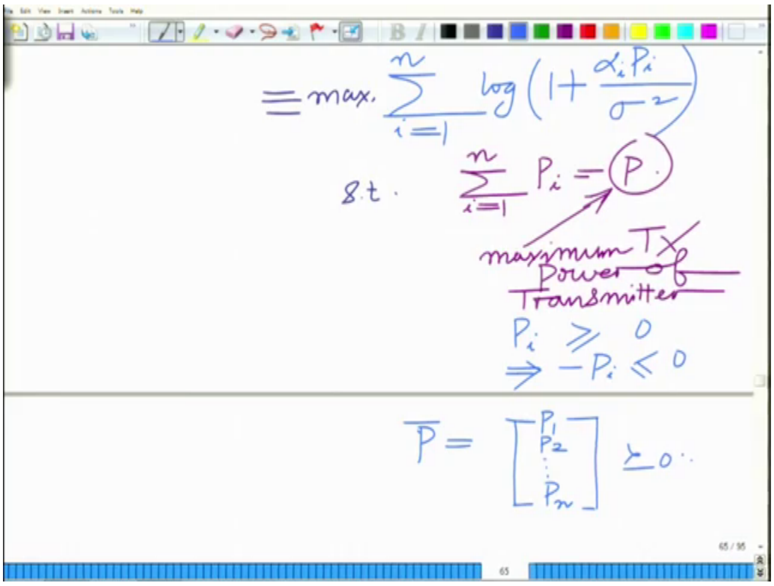
Well that will be given as the sum of the individual rates sum i equals 1 to n; log to the base 2, 1 plus alpha i; P i divided by sigma square. And naturally what we want to do is we want to maximize this sum rate. So, we want to maximize; maximize the sum rate this is your sum rate we want to maximize the sum rate. Now I am going to do a minor modification here; instead of choosing log to the base 2, I am going to choose log to the base e into log e to the base 2; so that is a natural logarithm.

So, this then becomes simply the natural logarithm I am simply going to write it as log or log to the base e. And this thing is as you can see log e to the base 2 is simply a constant factor. So, instead of multiplying; so instead of maximizing the objective function times a constant I can simply ignore the constant factor.

So, I can write this because it is going to be easier for me later; I am just going to write this equivalently the symbol denotes equivalently as maximizing the sum of the natural logarithm ok. So, when there is no base it automatically means that this is a natural logarithm ok. And now, so this is maximized; so equivalent to maximization of this. Now in any practical system, naturally there is going to be constraint what is the constraint?

The constraint is we do not have unlimited power, the total transmit power is a fixed quantity; so the sum of the powers P i across all these channels has to be fixed and that is given by the total transmit power of the transmitter the maximum possible transmit power ok. So, now we want to maximize the sum rate subject to the constraint that summation i equals 1 to n of P i equals P.

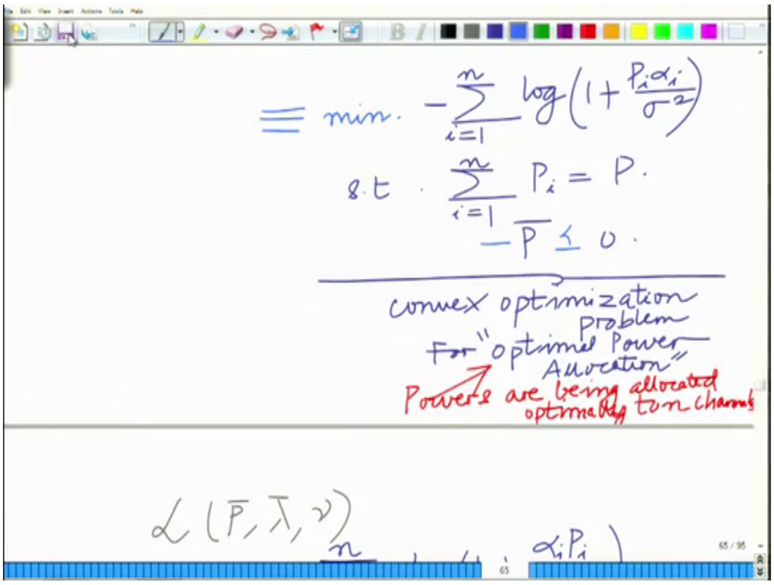
(Refer Slide Time: 10:27)



So, this is you can think of this as the maximum transmit; the maximum transmit power of the transmitter subject to that constraint. Now of course, now the other; so, this is that, so the sum of the powers of all the channels has to be P, there is a maximum transfer and in addition it is power; remember it is a power. So, power has to be nonnegative which means each P i has to be greater than equal to 0. So, you have each P i greater than or equal to 0 which implies writing it in standard form each minus P i is less than equal to 0 or if you look at this vector P bar as P 1, P 2, P n; then this has to be component wise greater than or equal to 0 ok.

So, this is our optimization problem; maximize the sum the sum rate, subject to the total power constraint and the non negativity constraint of each power that each power has to be each power P has to be greater than or equal to 0. And you can look at this now this is well log is a concave function the sum of log is a concave function. So, I can this; a this is a concave of maximization of concave objective function which can write as a convex optimization problem.

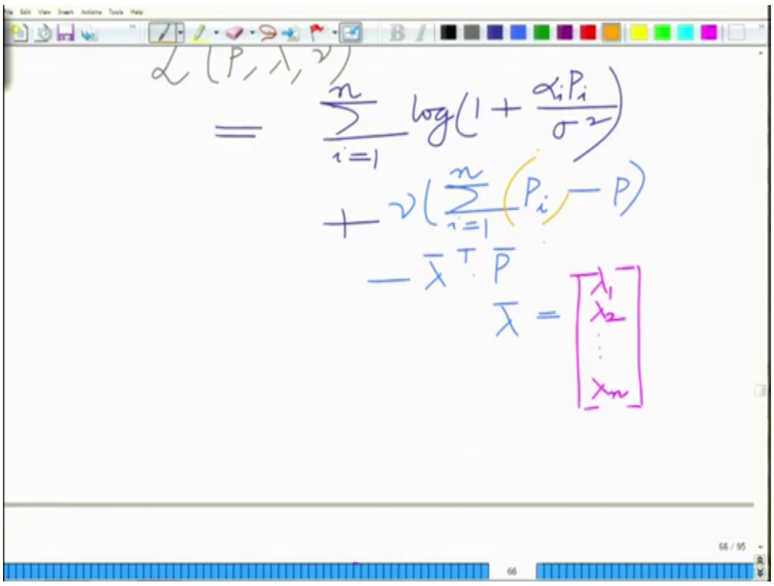
(Refer Slide Time: 12:01)



So, again I can write this as again equivalently I can write this as minimize i equals 1 to n log to the base 1 plus P i alpha i by sigma square; subject to the constraint I equals 1 to n this is the equality constraint and inequality constraint that is each and P bar is component was greater than equal to each element P i is greater that or equal to 0.

So, this is your convex optimization problem for optimal power allocation alright. You are allocating the powers optimally typically termed as optimal power allocation powers are being allocated optimally to the n channels.

(Refer Slide Time: 13:11)

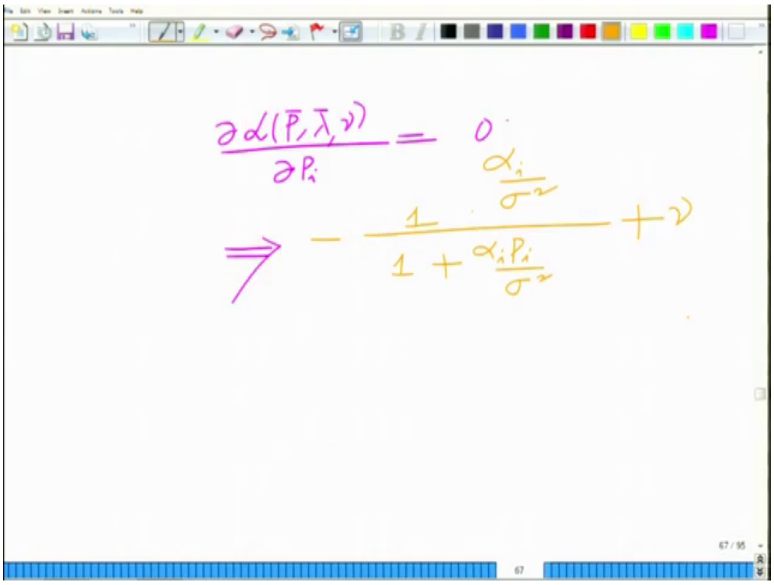


Optimally powers by being allocated optimally to the n channels and now we will use the KKT conditions to solve this right; how to come up with optimal power relation; what are the optimal powers? To maximize the sum rate what are the optimal power P 1 P 2 P n subject to the total power constraint the non negativity power constraint, which maximizes the sum rate across this set of parallel n channels. To do that let us start with again the Lagrangian already going to use the KKT conditions to find remember what are the KKT conditions; is the Lagrangian alright the derivative of the Lagrangian has to vanish at the optimal.

So, you form the Lagrangian; the Lagrangian of P bar, lambda bar since there is only one equality constraint; there is only going to be one equality the one Lagrange multiplier nu associated with the equality constraint. So, this will be the objective function i equals 1 to n log to the base one alpha i; P i by sigma square plus nu summation of i equal to 1 to n; P i minus P or P minus alright. This is the equality constraint minus or plus lambda bar transpose into minus P bar or you can write this simply as minus lambda bar transpose P bar alright because this is simply P bar greater than or equal to 0 component wise or minus P bar less than equal to 0 ok.

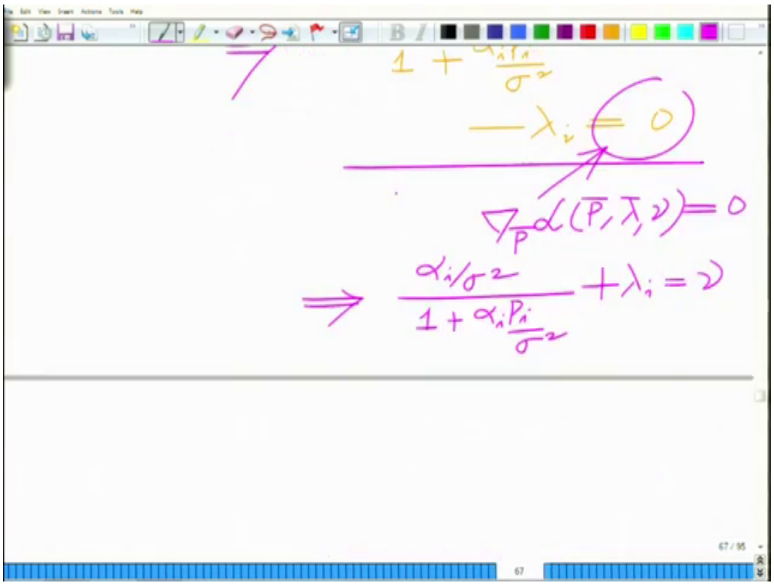
So, this is minus lambda bar; will lambda bar you have one Lagrange multiplier for each inequality constraint when is basically one Lagrange multiplier for each power; so lambda 1, lambda 2 upto lambda n ok; so this is the Lagrange multiplier.

(Refer Slide Time: 15:43)



Now, differentiate this with respect to P i; now if you differentiate this with respect to P i, what you obtain is basically minus 1 over that is the derivative log plus alpha P i by sigma square that is 1 over 1 plus alpha i P i by sigma square; times alpha i by sigma square plus along when you differentiate P i; you will have nu the Lagrange multiplier you simply get nu and the derivative of this minus lambda bar transpose P bar with respect to P i; the partial derivative is simply minus lambda i which is equal to 0.

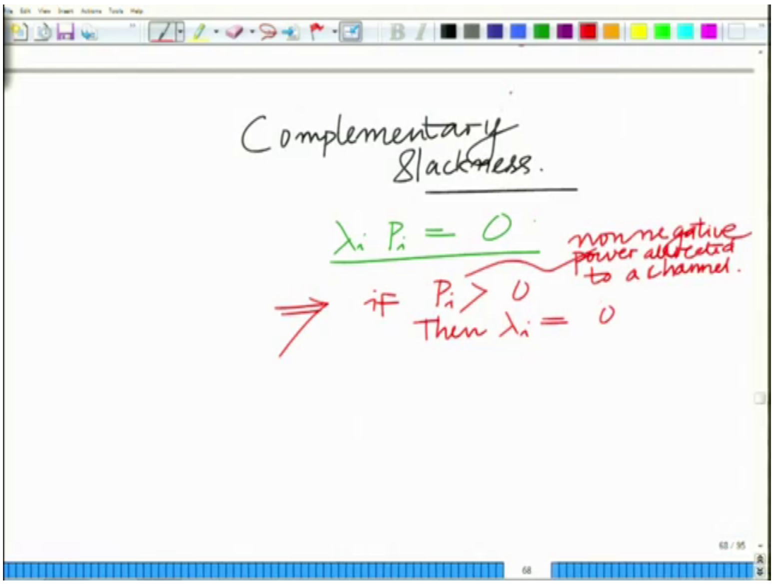
(Refer Slide Time: 16:35)



And now this is the key equation ok; this setting equal to this is a KKT condition because the gradient with respect to P bar of lambda P bar, lambda gradient with respect to power of the Lagrangian P bar lambda bar nu has to be equal to 0 ok.

So, this is one of the KKT conditions and this implies that alpha i divided by sigma square by 1 plus alpha i; P i divided by sigma square plus lambda i equals nu ok. The objective function is because of it is there has to be a minus here because we are minimizing the negative of the convex of the; negative of the objective function alright minimizing the negative of the logarithm which is a convex of objective function ok.

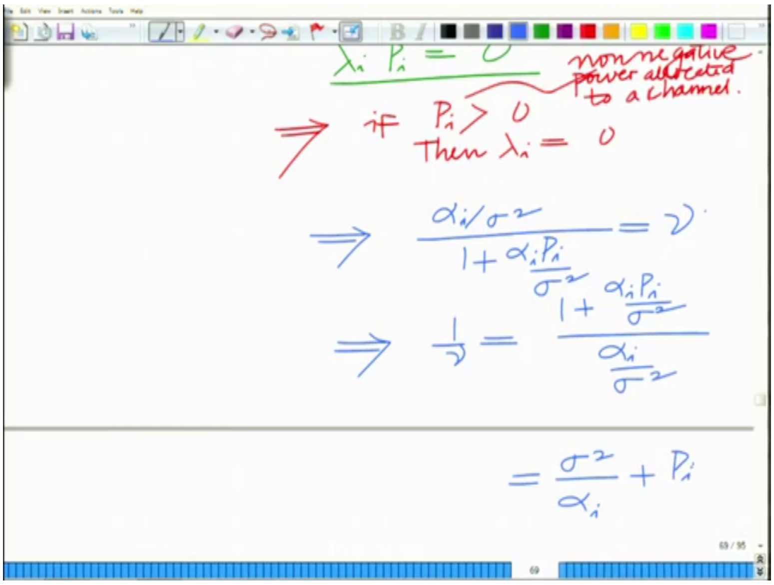
(Refer Slide Time: 17:52)



And now therefore if you look at this what you have is now remember from the complementary slackness let us go back to know the complementary slackness. Complementary slackness will be we have lambda i; P i equal to 0 alright; this is that is either the constraint is slack or the Lagrange multiplier is slack, but not both ok.

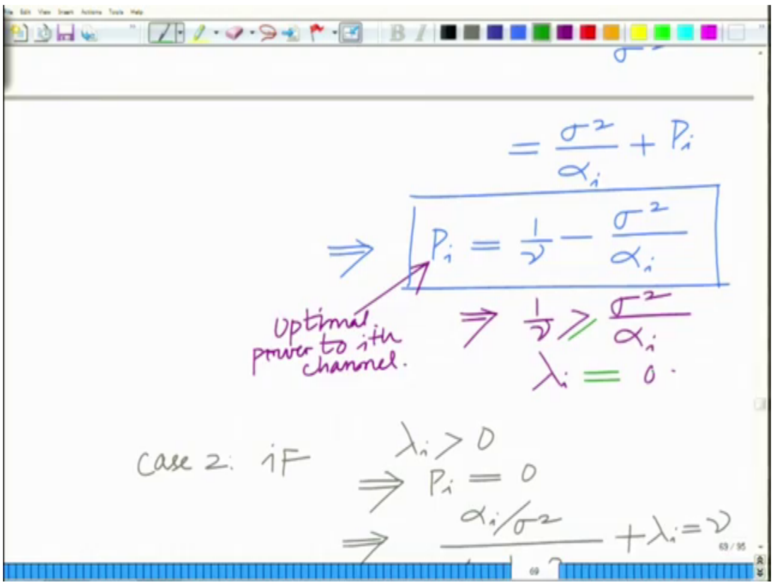
So, lambda by lambda into P i equals 0; now this implies something very interesting. This implies; now let us consider 2 conditions if P i is greater than 0 that is power allocated that is non-negative power; power allocated to a; to a channel nonnegative power is allocated to a channel, then lambda i equals because lambda i P i pi has to be 0.

(Refer Slide Time: 19:09)



So, this implies that alpha i divided by sigma square by 1 plus alpha i P i by sigma square equals nu because lambda i is 0; you simply set lambda i in above equal to 0 which implies that 1 over nu equals 1 plus alpha i P i by sigma square by alpha i by sigma square which is simply sigma square by alpha i plus P i.

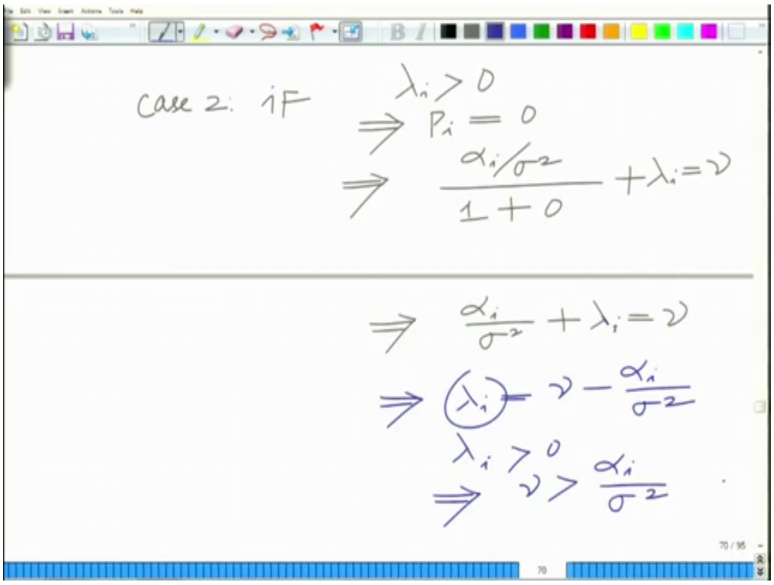
(Refer Slide Time: 19:51)



Which implies that P i equals 1 over nu minus sigma square by alpha i ok; which also implies by the way so this is the optimal power P i which is greater than or equal to which is greater than which implies also 1 over nu is greater than sigma square by alpha i ok.

So, since 1 over nu is greater than sigma square square over alpha i P i is over nu minus sigma square by alpha i and lambda i corresponding eigen value is 0. So, this is the optimal power allocated to ith channel ok. So, that is 1 over nu minus sigma square divided by alpha which implies that 1 over nu is greater than sigma square over alpha i because the power has to be nonnegative ok.

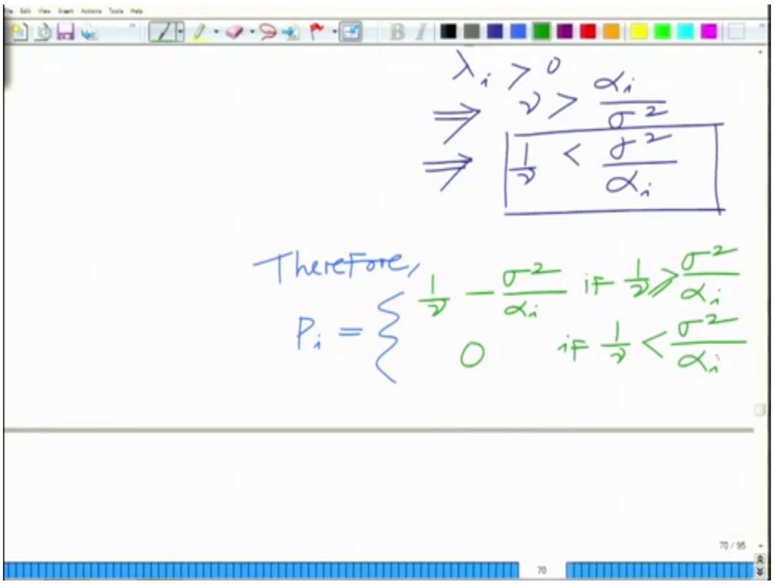
(Refer Slide Time: 20:58)



On the other hand, now if this is your case 2; so the above is your case 1, case 1 is P i grade. Now if lambda i greater than 0; that is Lagrange multiplier is slack this implies that P i equals 0. Now this implies now alpha i by sigma square by 1 plus alpha i; P i by sigma square P i 0.

So, 1 plus 0 plus lambda i equals nu; so this implies alpha i by sigma square alpha i by sigma square plus lambda i equals nu which implies that which implies that well lambda i equals nu minus alpha i by sigma square and the lambda i is greater than 0 which implies that nu greater than alpha i by sigma square.

(Refer Slide Time: 22:14)

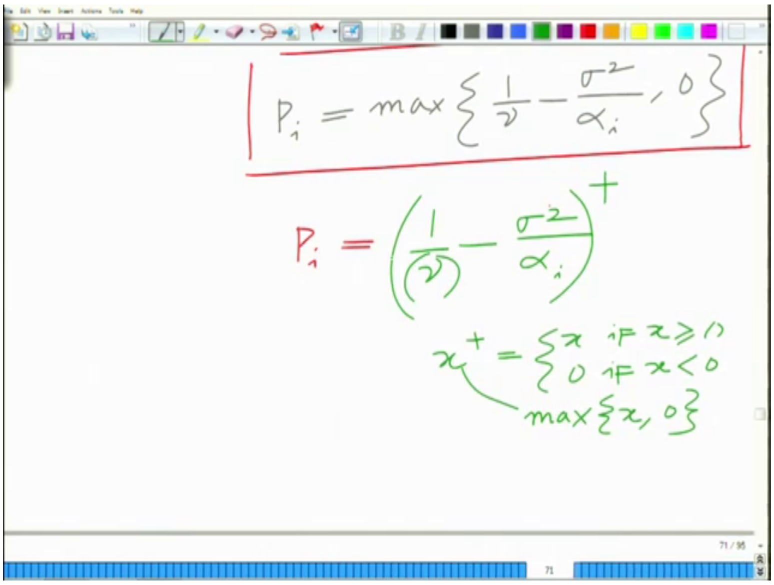


Which implies also that 1 over nu is less than alpha i or which implies that 1 over nu is less than sigma square divided by alpha ok; so, there are 2 cases 1 over nu less. And therefore, now if you summarize it what happens what is happening is if you observe it therefore, what is happening is P i; if you look at P i, P i equals well P i equals if you look at the condition above there is 1 over nu minus sigma square by alpha i.

So, that is 1 over nu minus sigma square by alpha i if 1 over nu greater than sigma square by alpha i; else P i is simply 0 if or you can make this as greater than equal to actually 1 over; so alpha i lambda i, in fact, lambda is greater than or equal to 0. So, let me just correct this a little bit; so lambda i 1 over nu greater than equal to 0.

So, lambda i lambda i is equal to lambda i is equal to 0 ok; so P i is greater than equal to 0 ok; now it does not really matter. So, basically what it means is if P i; P i is equal to 1 over nu minus sigma square by alpha i; if 1 over nu is greater than equal to sigma square by alpha i and it is 0 if 1 over nu less than sigma square by alpha ok.

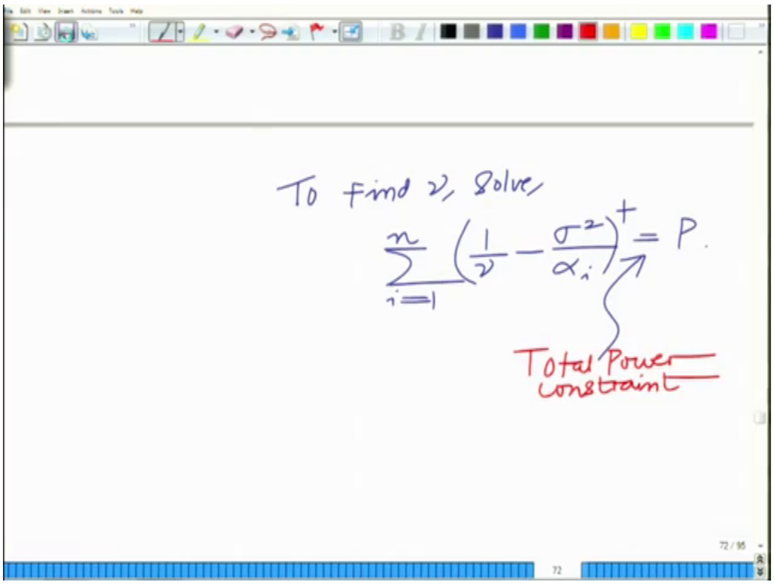
(Refer Slide Time: 23:57)



And therefore, you can write this P i succinctly as P i equals the maximum of 1 over nu minus sigma square by alpha i comma 0. Because if 1 over nu is greater than sigma square by alpha i, then the maximum is 1 over nu minus sigma square by alpha i which is the power P i. If 1 over nu is less than sigma square by alpha i; then this quantity is negative and therefore, the maximum will be and the power is 0.

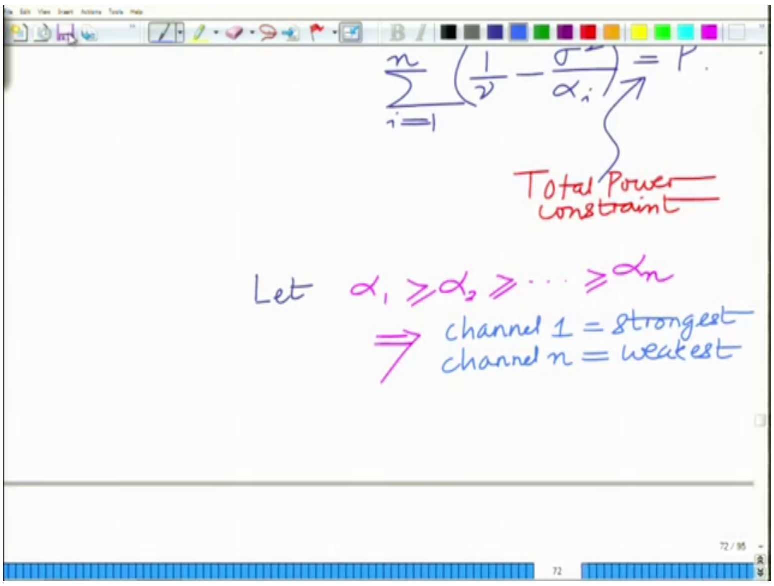
So, one power P i equals maximum of 1 over nu minus sigma square by alpha i and 0 and typically this is represented as P i equals the optimum power which is 1 over nu minus sigma square by alpha i plus where x plus equals x if x is greater than equal to 0; 0 otherwise or basically if x is less than 0. Basically again saying the same thing this is the maximum of x coma 0; that is the optimal power equation that is 1 over u minus sigma square by alpha i plus. Now how do you find this up till because you need still find lambda nu? That is from the total power constraint ok.

(Refer Slide Time: 25:28)



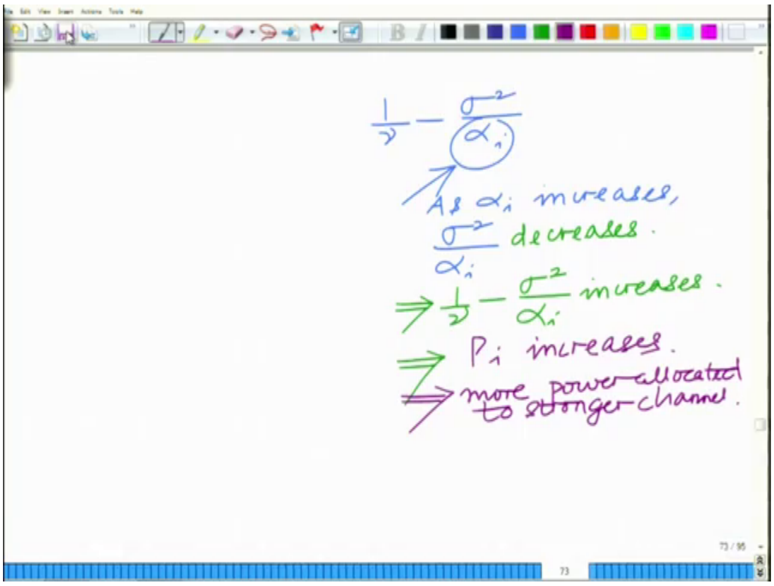
So, now how do you find nu? To find nu you solve summation i equals 1 to n; 1 over nu minus sigma square by alpha i plus equals P, that is from the total power constraint. You find this from the total power constraint; now you observe something interesting.

(Refer Slide Time: 26:05)



Let us go back to what we said before, let us assume without loss of generality that these are ordered that is alpha 1 gains of alpha 1, the first channel is strongest greater than second channel; greater than equal to alpha n ok. This implies that channel 1 has the strongest gain highest gain which implies channel 1 equals strongest and channel communication channel n equals the weakest.

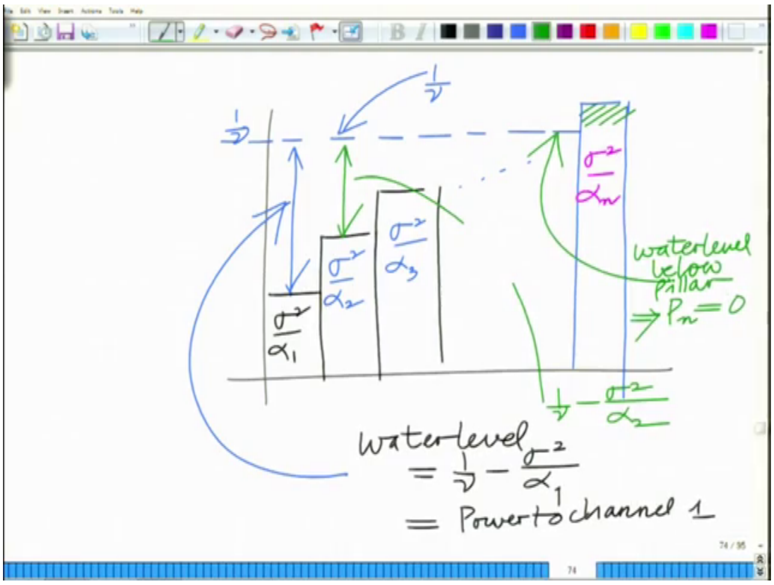
(Refer Slide Time: 26:54)



Now what this implies is if you look at this; look at carefully look at this 1 over all nu minus alpha i divided or minus 1 over nu minus sigma square by alpha i. So, as sigma i increases; observe that as alpha i increases sigma square by alpha i this quantity decreases, this implies 1 over nu minus sigma square by alpha i increases implies that as alpha increases; the power P i increases alright.

So, more power is allocated to the stronger channel which implies P i increases; which implies more power is allocated to the stronger channel which implies that more power is allocated to the stronger channel ok. Now there is an interesting representation for this if you look at.

(Refer Slide Time: 28:24)

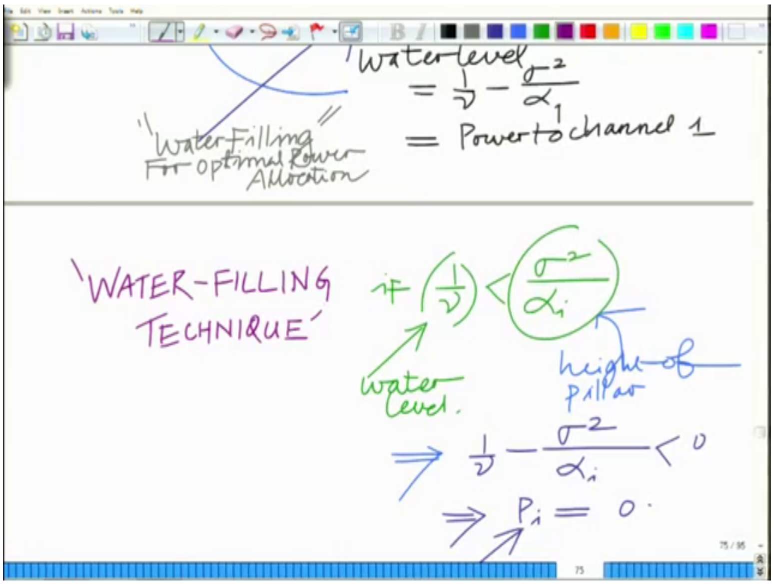


Now, let us look at this representation if you look at for instance a sort of bowel you can call it or a area with this kind of pillars ok. So, the first pillar is sigma square by alpha 1 corresponding to first channel because alpha 1 is larger sigma square by alpha 1 is smallest. So, this is sigma square by alpha 2 sigma square by alpha 3 so on and so forth; remember alpha n alpha is decreasing towards alpha n.

So, sigma square by alpha n will be the largest ok; so this is your sigma square by alpha n. Now if you draw here the level 1 over nu; if you call this as your level 1 over nu. So this is your level 1 over nu you can think of this as a water level; now the power allocated to the first channel is basically the amount of water that is a level of water 1 over nu. So, if you look at the water level here alright 1 over nu minus sigma square equals by alpha 1 which is the power allocated to channel 1.

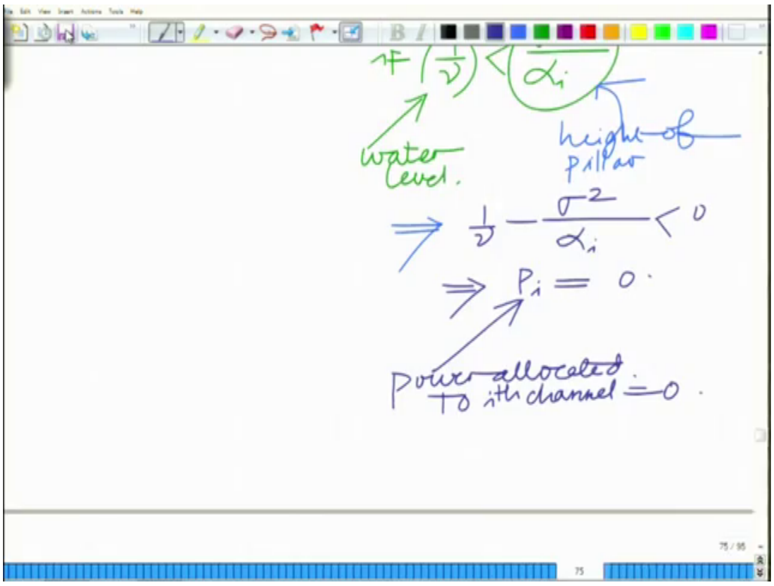
Similarly, if you look at this level; this is basically your 1 over nu minus sigma square by alpha 2 at; you see something interesting over here water level is below the pillar. So, here the water level below the pillar this implies that P n power allocated to the channel is 0.

(Refer Slide Time: 30:40)



That is if 1 over nu is less than sigma square by alpha i for any particular which means this is your water level, this is the height of the pillar, this implies that 1 over nu minus sigma square by alpha i is less than 0 implies P i equals 0. So, power allocated to ith channel is 0; power allocated to ith channel is 0.

(Refer Slide Time: 31:27)



Therefore, this scheme which you can think of it as a water level poured into a bowel with several pillars; the height of each pillar is sigma square by alpha i corresponding to ith channel; power is allocated only if there is a finite non zero water level about that ith pillar that is 1 over nu is greater than sigma square over alpha i; if the pillar is higher than the water level then no power is allocated to that particular channel therefore, this is known as the optimal water filling algorithm.

So, this is known as the optimal water filling algorithm. So, you can think of this as a water level. So, this is known as a water filling algorithm; this is known as the water filling technique the water filling scheme for optimal power allocation ok. So, if the level of; so this is known as water filling for optimal power let me just write it here; the water filling scheme or water filling technique or water filling procedure or water filling scheme for optimal power (Refer Time: 33:06). That is basically even think of it is a convex; 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.

You can see that if 1 over nu; so if P i is greater than 0 then you have lambda i has to be 0; if P i equals 0 if lambda is greater than 0 if lambda is greater than then P has to be 0. And the optimal power allocated the ith channel is the maximum of 1 over nu minus sigma square by alpha i comma 0. Meaning that if 1 over nu is greater than sigma square by alpha i, finite power is allocated to the channel as the power is allocated is 0 alright.

So, this is a nice scheme or this is the optimal scheme; so this is the optimal scheme for to maximize to allocate power across the parallel across parallel channels; that maximizes the sum rate of communication between the transmitter and tracy. So, we will stop here and continue in the subsequent.

Thank you very much.