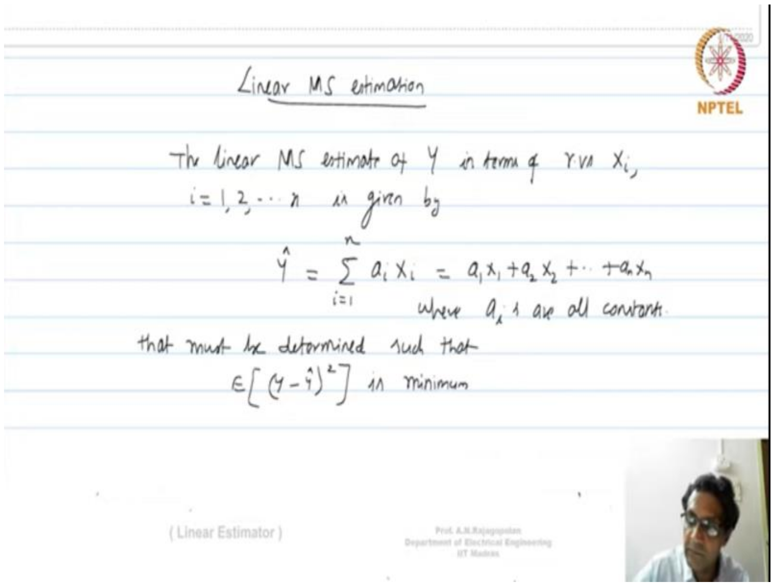
Image Signal Processing
Professor A.N. Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology, Madras
Lecture 79
Linear Estimator

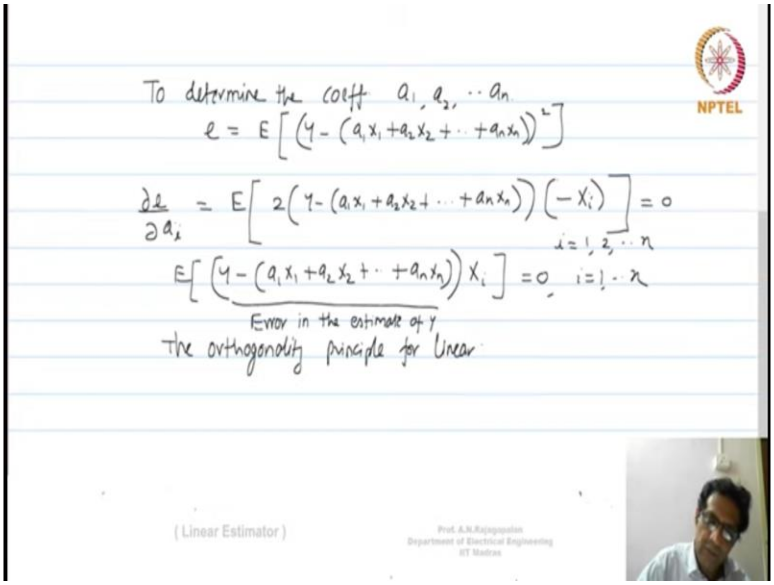
(Refer Slide Time: 00:18)



We are looking at linear mean square estimator or estimation of a random variable y from a number of random variables x. N number of variables let us say. The linear mean square estimates of y in terms of random variables. Let us say x i, i equal to 1 to n is given by, now in this case we are, this is a constrained situation. We are constraining the estimator to be linear is given by, let us say y hat is equal to summation a i x i. In general, of course, you can also have a plus b in order to make it an assigned estimator, but then I will take a simpler case where we simply have, anyway b is also very straightforward to solve.

I equal to 1 to n or in other words, this a1 x1 plus a2 x2 plus xn where ai's are all constants which is why this is a linear estimator and the idea is that we would like to solve for a1 a2 right up to an such that expectations so idea is constants that must be determined such that, we would like to find out these constants, such that the expectation of y minus y hat where y hat is this linear estimator, y is as small as possible, is minimum, under the constraint that y hat is linear. Correct? This is why we are looking at these constants.

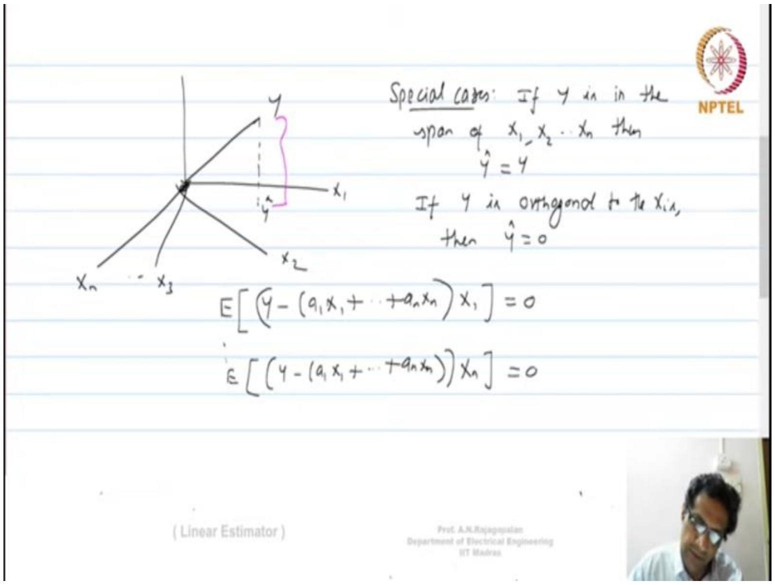
(Refer Slide Time: 02:22)



What this means is that now, let us again so to determine the coefficients ai’s a1 to an, let us define the error to be e is equal to expectation y minus y hat is a1 x1 plus a2 x2 plus plus plus a n xn. The whole square. This we want to be minimum. Now suppose, we do a partial differentiation of e with respect to some a, let us say aj or ai let us say. Let us put it as ai. Then what we find is, this will be equal to expectation.

Then we will have 2 times y minus a1 x1 plus a2 x2 plus plus plus all the way up to a n xn and the whole thing. If you were to examine this equation more carefully, this simplifies to expectation y minus a1 x1 plus a2 x2 plus plus plus a n xn into xi is equal to 0. I equal to 1 to n. Now, what this means is, this is the error in the estimator y. Correct? The error in the estimator y because this is y hat. So, y minus y hat is the error, the estimator and what it is saying is this error is orthogonal to e to the observe random variables xi x1 to xn and this is called the orthogonality principle for linear MS, mean square estimator.

(Refer Slide Time: 04:49)

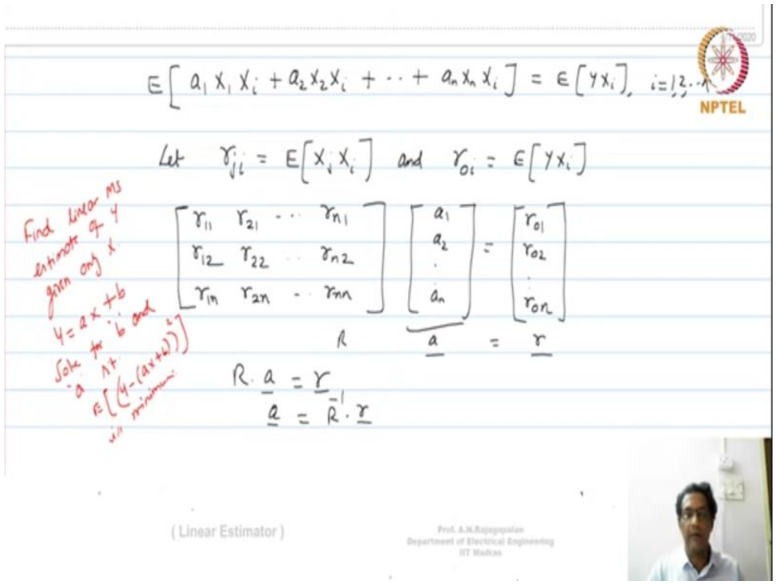


So, in general, so this orthogonality principle, in a way, if you want to look at the geometric interpretation, what it would mean is that, if you are trying to, if you have your x1 to xn like that exceed it all the way up to xn. You have a y, that is sitting somewhere up there and what you are trying to do is, you are going to drop this orthogonally onto the space spanned by x1 to xn and this is your y hat and this error which you have here, this error is actually orthogonal to each of the xi's and some special cases, what will be the special cases that you might have here.

Special cases will be like if y is a vector in the span of x1 to xn, x1, x2 up to xn. Then your y hat will be equal to y itself. You cannot make any error because of the fact that it lies in the span of the axis. Now if y is orthogonal on the other hand the other extreme, the y is orthogonal to the xi's then your y hat is 0, if it is orthogonal. Now, in this case what you find is that through this principle, orthogonality principle says the error is orthogonal to all the xi's, each of the xi's and therefore this itself leads you to n number of equations if you observe.

This equation is valid for every i and therefor you can write down a set of equations. So we can go ahead and write this down as from the earlier page. What we have is expectation y minus a1 x1 plus plus plus a n, xn into x1. It is equal to 0 then all the way up to expectation y minus a1 x1 plus plus plus a n xn, Xn is equal to 0 because the other equation is valid for i equal to 1 to n. Therefore, if you write down a bunch of equations, you will get any equations like that.

(Refer Slide Time: 07:12)



Therefore, for each of these equations, it will get something like expectation a1 x1 xi plus a2 x2 xi plus plus plus a n xn xi is equal to expectation y times xi where i goes from 1 to n. Now let us suppose r ji is, suppose this indicates expectation xj xi and r 0i indicates expectation of y xi. Then what we have is, we can write this entire thing in matrix vector form as where the unknowns are your a1 a2 all the way up to an.

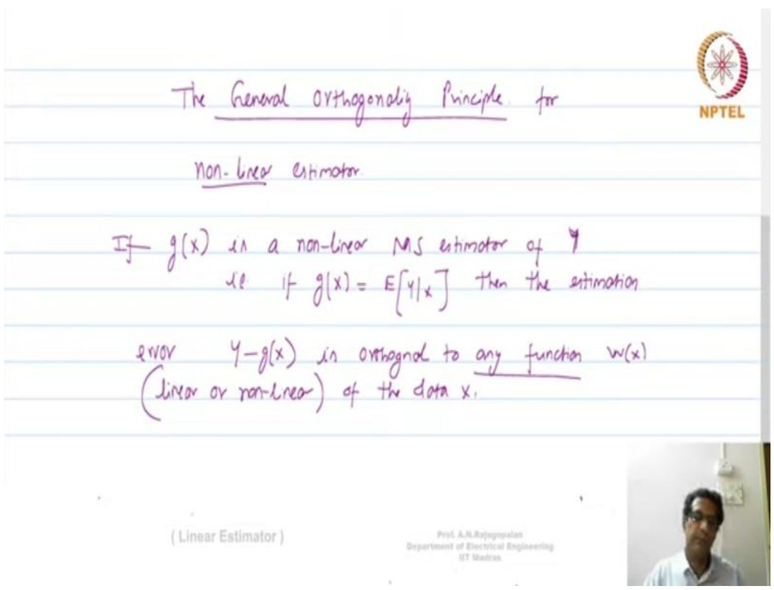
These are your unknowns and we can write this as a matrix multiplying a vector is giving you a vector and this will be like, so we start with i equal 1. So what you will have is r11 and then you have x2. So now for i equal 1, you have got the what r11 which is expectation x1 square. Then you have got r21 which would be the expectation x2 x1 and then all the way up to expectation x and x1. Therefore, you have got like r11 r21 all the way up to rn1.

Then similarly, you can have r12 because when you put i equal to 2 then you will get x1 x2, x1 x2 is, so when you put x1 x2 that is r12. X1 x2 is r12. Then r22, you got like rn2 and then all the way down to r1n because you going to put xn here, if you are going to put x1 xn and you know x1 xn is like r1n and then you have got r1n and then you got r2n all the way up to rnn. This is equal to r01 because expectation y times xi. So yx1 is 01, yx 2s are 02 and then all the way up to r0n or in other words, if you call this as a and if you call this is as r matrix and then this we call as r, then we got like r times a is equal to r or in other words a is nothing but r inverse times r.

So, your a is which is your a1 a2 right up to an, those can be found out from this kind of a statistical information. Try to kind of solve this and leave it as a small exercise. Show or find linear MS, mean square estimates of y, given only x. Now, what we have solved is a more general case, solved for the general case x1 to xn, to solve it for the case, when let us say you have y and then you observe only another random variable which is x and so you need to be able to show us to what should he then your a.

So, in this case your a will be, so in general try to solve for y is equal to ax plus b. So, solve for a and b. Now in our case we chose b to be 0 but then let me just give you a simple example. Try to estimate y from x as ax plus b, which is like an affine estimator, purely linear due to y equal to x. Now, we are looking at some kind of an affine estimator so solve for b and a and a such that b na such that expectation y minus ax plus b the whole square is minimum. So, it is a simpler example of the one which we have already seen. Now, this linear mean square estimator is actually nice to have.

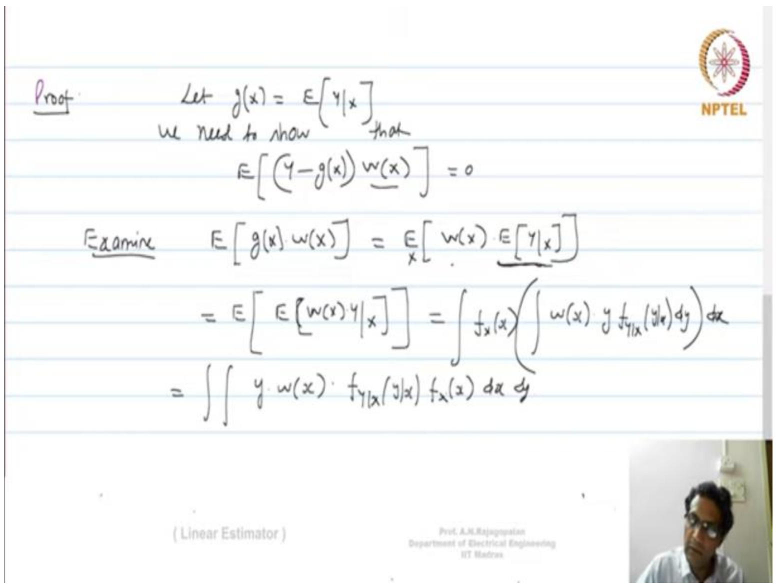
(Refer Slide Time: 12:00)



Now, we saw that it obeys the, what is called, orthogonality principle. We said that error is orthogonal to the observed random variables x1 to xn. Moving forward, we are going to look at something called the general orthography principle and this is with respect to the conditional mean. The general orthogonality principle. See, as far as linear estimator is concerned, we said that the error that you make in the estimate of y is orthogonal through xi’s. Now the general orthogonality principle goes further beyond and states the following. This is for non-linear estimator that means the conditional mean.

What it says is if g of x is a non-linear mean square estimator of y that is if g of x equal to the conditional mean of y given x then, this estimation error which is y minus y hat or y minus g of x in this case, is orthogonal to any function. Note this, to any function. There we only had that it is orthogonal to xi's. We are saying to any function wx linear or non-linear. Now, let me just remove the comma of the data x. Hence, so goes further beyond.

(Refer Slide Time: 14:20)

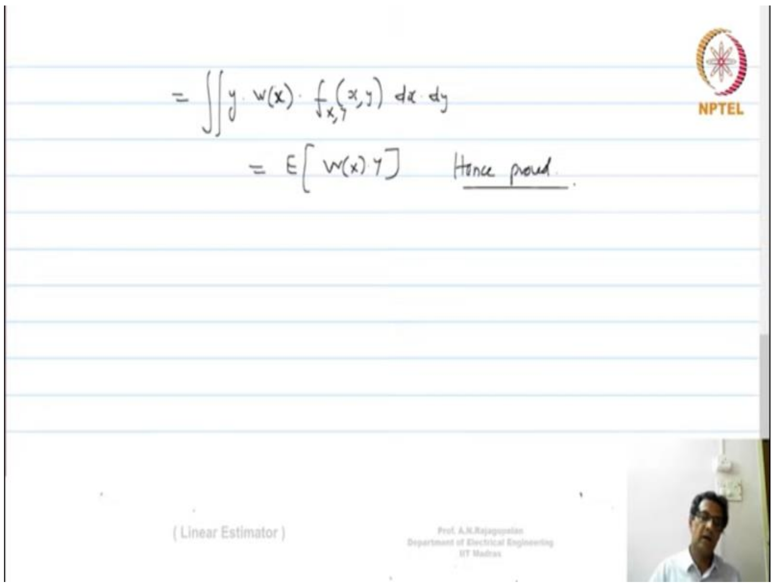


So, we will try to prove this. What do we mean by a general orthogonality principle? So the idea is that so we know that proof would go as follows. The proof goes as follows. Let g of x, as has been already stated, is the conditional mean of y given x. Let us put it like that. So, we need to show, what do we have to show, we need to show that expectation y minus g of x into any guy, any this one w of x, any function of x is equal to 0, for any this one w of x. We do not want to really constrain w of x.

It could be any function of x linear or non-linear. Now examine expectation g of x into w of x. This we know is nothing but expectation of w of x into g of x is conditional mean of y given x and this we know is a function x. This one, the conditional mean of y, let me just write this, and this we know is simply a function of x and therefore this outer expectation is with respect to x.

So, this equal to, if not exclusively say that, we should automatically infer or we should automatically know that the outer expectation is with respect to x. This we can write because of the fact that the inner integral here is with respect to y, we can write this as expectation of expectation of w of x into y given x and this furthermore is nothing but integral fx of x and then followed by an integral which is w of x into y fy given x, dy and outside you will have dx or this in other words is, double integral y into w of x into f y given x f x of x dx dy.

(Refer Slide Time: 16:56)



This is equal to, so if you see observe here, this we can write as y double integral y into w of x into joint because you are multiplying, if y given x with fx xy dx dy or this in turn is expectation of wx into y and hence we have the proof which is equal to this y into g w of x. So, we wanted to show that this is equal to 0 that is equal to showing that expectation g of x into w of x is equal to expectation y into w of x, which is what we have proved here. So hence proved.

Now, at this point, what we would like to do next really is what is called a normality theorem, what is called the normality theorem and this theorem is the one that actually, you know, reveals the equivalence between a linear estimator and a non-linear estimator, the conditions under which the conditional mean becomes linear. Typically, we know that it is a nonlinear function of x but then under certain condition, it becomes a linear function of x, which means that the linear estimators is indeed the optimal one. It is no longer sub-optimal. So, we so we will see next, what is the normality theorem.