Estimation of Signals and Systems

Prof. S. Mukhopadhyay
Department of Electrical Engineering
Indian Institute of Technology, Kharagpur

Lecture No - 11
Linear Mean Sq. Error Estimation

Linear mean squared error estimation. It immediately shows two things; if you see this title, that in any estimation problem there has to be a signal model, actually what you are trying to say is that, you are you are trying to estimate some quantity which you cannot probably measure or I mean you you are trying to estimate it in term of other quantities or you are trying to predict something whose value is not a variable or something is very noisy. So you want to estimate the actual correct thing and you always have the noisy measurement. So in general what happens is that; you have some signals at hand which are called measurements and you want to get something else, which you cannot measure directly, that is that is the that is the estimation problem.

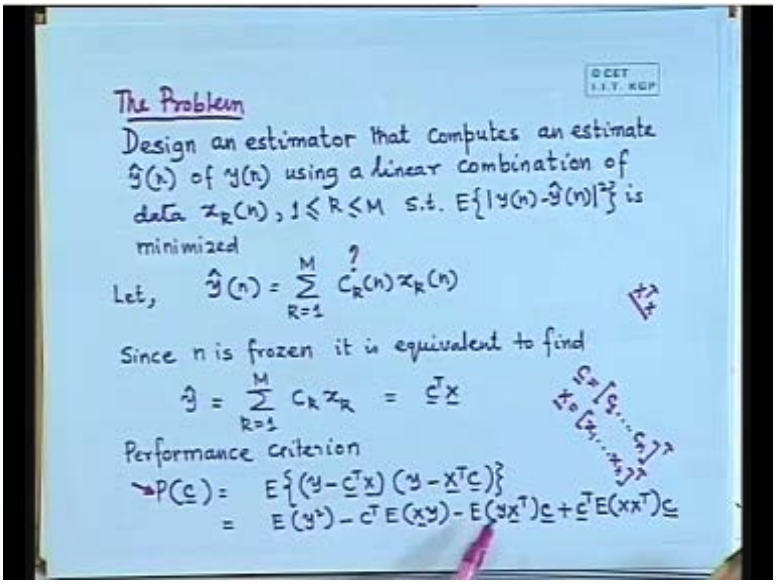
Now the, now the thing is that first of all there are two things that will characterize; it the first thing that will characterize, it is called as is called a model set. That is in general; the set of you will have to have some rule of computing that quantity producing the estimate based on the measurement, in fact that is the estimator. So you so you cannot in general I mean there are they are there are so many possible rules, I mean all possible linear and non-linear dynamic functions, etcetera etcetera. That it is not feasible to leave, that question open, so people will generally focus their scope then, okay if I am if I am looking for a looking for a solution to the problem, I am looking for this kind of a signal model.

Let us say, in this class we will you will see that we are we are we will be looking for a solution, using particular kind of signal models. So we will try to find out, some estimator which is of that structure, so that is why this term linear has come and now there are so many possible linear systems, so the so the question is which one of them will you pick up, as a good estimator? So you need to define, a a a performance criterion by which you can access, whether the estimator is good or not, right. And there could be several possible performance indexes and mean square error is is the performance index that, we are going to consider

today. So you see that with every estimation problem, there will have to be a signal model and there will has to and there will have to be a performance criterion. Then then these two things are essential quantities which characterize any estimation problem, okay.

So what do we do in this particular estimation problem? So this the problem, initially we consider the problem in a kind of a general setting and then we will show that, this problem can be applied in several practical cases.

(Refer Slide Time: 3:39)



So the problem is to design an estimator that computes an estimate y hat n of y n, so y n is the signal that we want to estimate. We cannot estimate that value, you cannot measure it, so we so we produce an estimate of y n, which is called the y hat n and we produce this estimate using a linear combination of data. So so these are my my my measurements. So I have k measurements, k signals where k rise between 1 to M. Each signal I have at different sampling instants and based on this k signals at a given sampling instant, I want to generate an estimate y hat n of y n.

Now what should be their estimate property? It should be generated such that, y n minus y hat n magnitude whole square, but you do not need to put magnitude; you can put magnitude if this y is a vector, this y could be a vector in general. But let us think of scalar only now, so y

n minus y hat n whole square is minimized. That is if you use any other y hat n, other than using that particular rule which is the which is the MMSE estimator, if you use any other linear estimator, then that will definitely give you a more error. You want to construct any such that, among the linear estimators; it will give you the mean; it will give you the minimum mean squared error. Among the nonlinear estimators, you can still do better. So maybe you could construct a construct a neural network, which will give you a still better still lower estimation, error. But among the linear ones; no one will it will not be possible to give a better estimate, so we want to construct such an estimator.

So when we say linear linear combination of data we mean that, I I want to generate y hat n using a rule such as this C k n, x n. So I want to compute; this this is what I want to find out, what are these going to be, such that if I feed it to a data and generate an estimate this condition, will be satisfied. This is the problem, okay. Now let us first, so so this means that for every time instant, I am going to choose a different coefficient, right. At every n for for estimating y y hat one, I am going to get a set of m coefficients, for estimating y hat two I am going to get a set of set of coefficients, etcetera. This is the most general problem, okay.

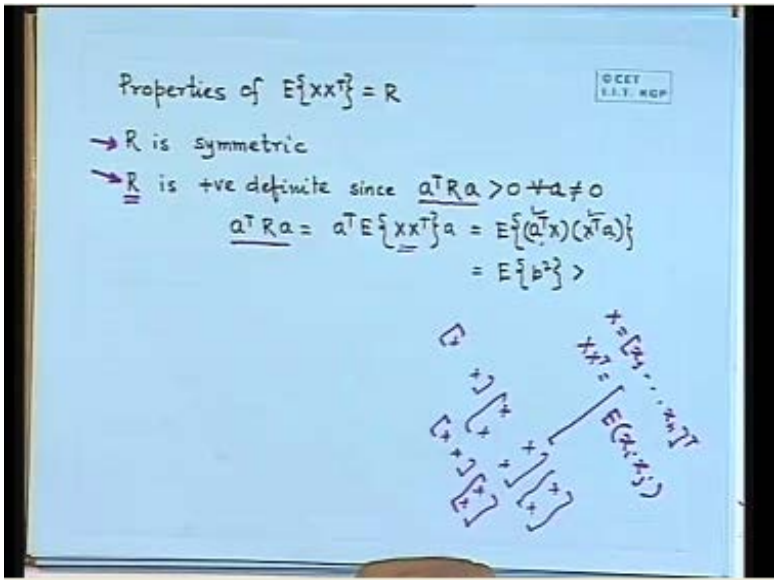
So let us first see, what is what is the what is the elementary problem, which I have to solve if I have to get this. That is if I want this, that solve it for some particular n so if I freeze n, choose n is equal to something. Then what happens is that, these these stochastic processes will now become random variables. So now you have to find a problem such that, y hat is equal to c k, x k, find out this c k. So I am just freezing n, so I am eliminating removing it from the equation. So y hat is now a random variable; x k are other random variables, which are measured and I want to actually, y is a random variable and I want to estimate y hat based on these random variables. Using these coefficient now what should be the coefficient? That is the question.

So now what is y minus y hat, it is y minus. See this equation if I have k equal to one to m, I can always write it like a vector matrix. Where this c hat, this is my problem, I I cannot write here properly, so I will write here so, what is c hat is a vector whenever; I give an under bar, it is a vector because I cannot show bold or anything here. So it is c1, c n .And this X hat is equal to x1 x n; these are the two vectors actually they transposes, it is the column vectors. So naturally this can be written as C transpose x, okay. So now what is, what do I want to

minimize? So I have a performance criterion; which is a function of my weight and I want to minimize it, I want to choose such a C such that, the value of this performance criterion will be minimum. So what is the performance criterion? It is expectation of y minus y hat, so y hat is C transpose X, have just defined. So it is y minus C transpose X into y minus X transpose C. These are written in this fashion; if it is a scalar it does not matter because, y minus C transpose X is a scalar, so I could have written even whole square, but but in general it will be a matrix, so I have written it as if which will satisfy even a matrix case. So the a matrixes non-squared means, that X transpose x. So if you say non, that if you say X one square plus, X two square plus, X three square plus, X four square etcetera, then you can write it in matrix rotation, as X transpose X.

So I am writing the error vector transpose error, vector actually this should have been y transpose, in that case anyway. So now, if you multiply what will happen? You will get expectation of y square, just multiply this and this then you distribute the linear operator and the the expectation operator. So you have to expectation of y square minus c transpose; c is a constant, so it has it is not a random variable. So therefore the C transpose expectation of the x y minus, expectation of y x transpose. This is if you multiply this to this, you get this term. If you multiply this with this, you get this term just just term, by term multiplication C transpose you get this term. So this is how your performance criterion depends on C. Now you have to choose a C, which will minimize this, that is the problem, okay. So that is very simple actually, but but before we do that, we need to find need to know the properties of this matrix, expectation of X X transpose or the autocorrelation matrix.

(Refer Slide Time: 10:26)



First, we have obviously R is symmetric, what is the expectation of X X transpose? If you have a vector, If you have a vector, I have to write like this, I hope you can read this. So if you have a vector X is equal to x1 to x n, it is a column vector actually. Then what is X X transpose? X X the general ijth element of this matrix, this will be your matrix. So column into row, to the ijth element will be given by expectation of x I, x j. So obviously the jth element will be expectation of X j X i, which is equal to ijth element; so this matrix is symmetric. If the jth element of a of an matrix is equal to its ijth element symmetry.

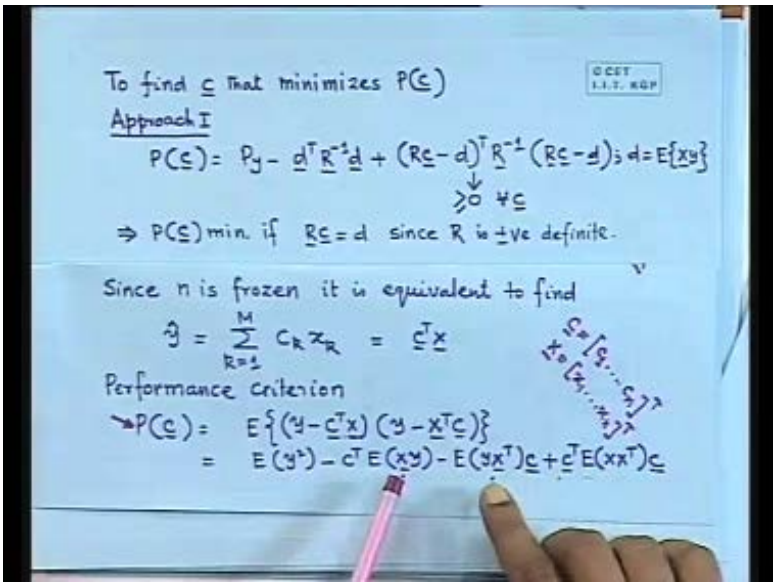
It is also positive definite, now what is what is meant by positive definite? For a matrix positive definite means; that if you take any vector a, and evaluate this quantity, a transpose Ra this will be a scalar. See a matrix, you are say two into two we are multiplying it by a transpose and multiplying this by a so you will get a scalar, this into this plus this into this this will be this going to be, a row multiplied by a column. This will generally give, what is known as a quadratic form.

So therefore this if the matrix R is such that, for all vectors a which only if the vector is identically zero, that is zero zero zero zero elements; then this has to be zero. Otherwise you choose any other vector a, it should be nonzero. The R is it should be positive, the R is such then then then the matrix R is called positive definite. Now this R is positive definite, why?

because; if you choose any a then, what is a transpose Ra, it is a transpose into, I am just substituting this. Now you take this expectation out, then you get expectation of a transpose x, into x transpose a. Now this is a scalar this is the scalar, they are they are actually the same scalars. So so suppose a transpose X is b, then you have expectation of b squared, which cannot be negative. So therefore for any other for all a R is this is going to be positive, that is why the matrix the the correlation matrix is positive, definite symmetric. This is a property of R, which is fairly obvious to see.

Now let us minimize that, there are there are two approaches; you can this is the this is the I mean a more this is an approach where what what I have done is see, I have just because I because I know the result I I I have written it, you can see that this this performance criterion, this one.

(Refer Slide Time: 13:30)



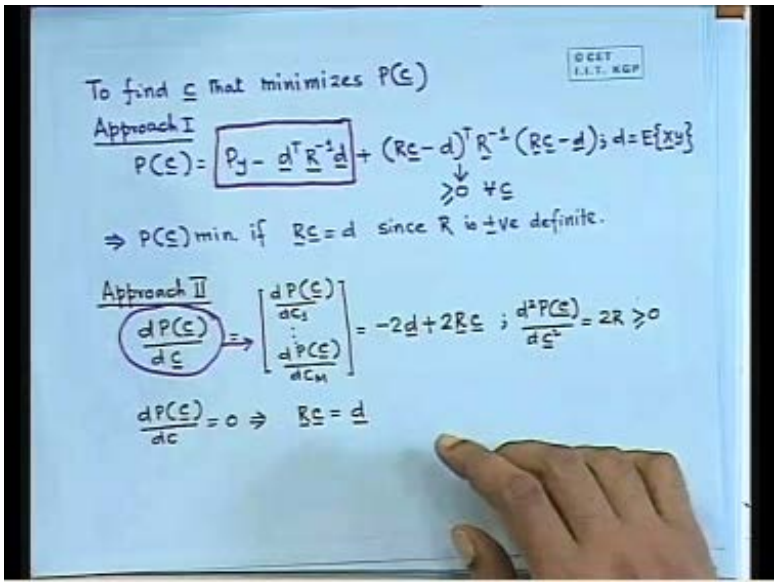
Let's put it here, okay. I can put it here. See this performance criterion can be cast like this, where d is expectation of x y, see this is this is d this if if this is d this is this is d transpose. So you have now, what I have simply, what I have done is now, if you if you just multiply this, what is going to happen? You are going to get terms like by the way a plus b, whole transpose is equal to a transpose plus b transpose. So and a b transpose equal to b transpose a transpose, I think you know these things. So suppose you multiply this, this term is going to give after this transpose is taken out.

It will be C transpose R, R is symmetric so therefore R transpose is R. This R multiplied by this R inverse will give you will give you, identity rather the this R multiplied by this R itself. Actually if you if you just break it up, you will get that this term basically. This term can be cast like this that is very simple do to do, if you just break it up. I do not need to explain it too much, this is just a this is just I mean this this this has been done in one step, because the result is known. So now, once you can break it up now you see; that obviously this part does not depend on C, this part only depends on x and y, okay

So they are in no way depended on the choice of c. So by choosing c you cannot affect this part, by choosing c you can only affect this part. Now if R is a positive definite matrix, obviously R inverse is also positive definite. If if if x is positive one, by x is also positive. So therefore, this term can only be greater than equal to zero. This this is now the vector; a so a transpose R inverse a is greater than zero, because this is positive definite. Which means that, the this performance criterion can can only be greater than this. It can only increase by by this amount and when is it going to be least, when this vector is identically zero or when so this is minimum of R c is equal to d. If that condition is satisfied; then you get the weight which will make this performance criterion minimum, right.

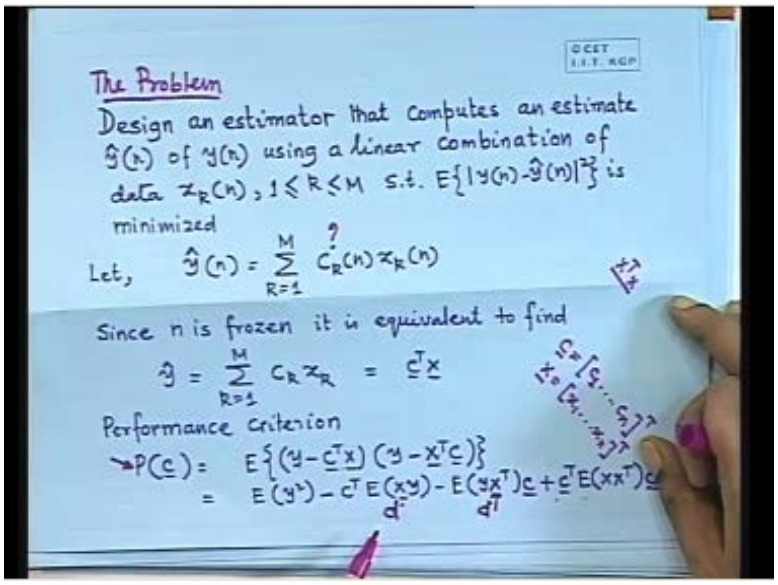
So so so this is how you get c, this is one way. The other way is which we all know, that is simply differentiate only; thing is that here we have vectors, so you have to differentiate with respect to vectors, but that simple.

(Refer Slide Time: 16:26)



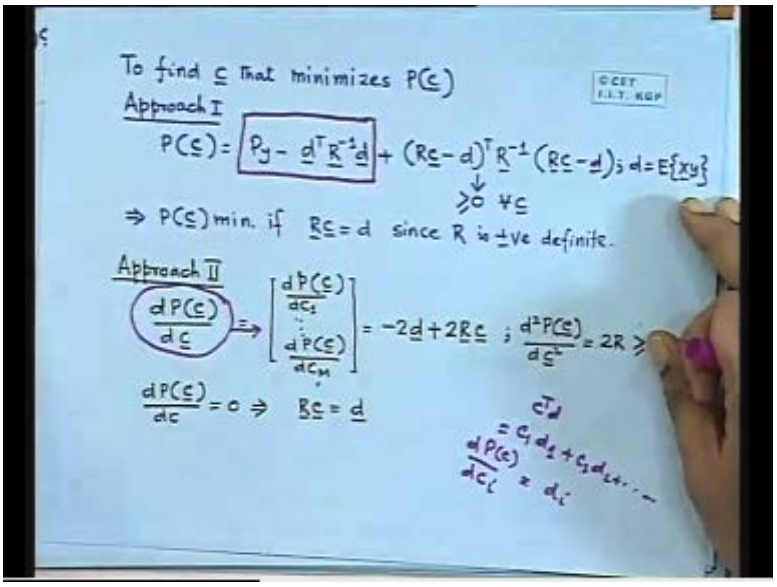
So the other approach is simply take d P c of dc, now remember that this P c is a scalar. It is a performance index so it is a scalar and this c is a vector. So what do we mean by by differentiating of vector and differentiating a scalar with a vector, this is this is a standard notation. This actually has is a symbol which means these, that it means that d P c by dc1, d P c by d cm, it is a it is a vector. First differentiate P c with respect to c one, then with c two, then with c three and put them in a vector there, this is a notation which means, this okay. So now it will if we do d P c dc1, look at this this part, if if just we differentiate it, what you will get? This term will go away, because it does not have to depend on c. And in this term, okay it is it it is easier to see it from here; you see it from here, because they are same.

(Refer Slide Time: 17:38)



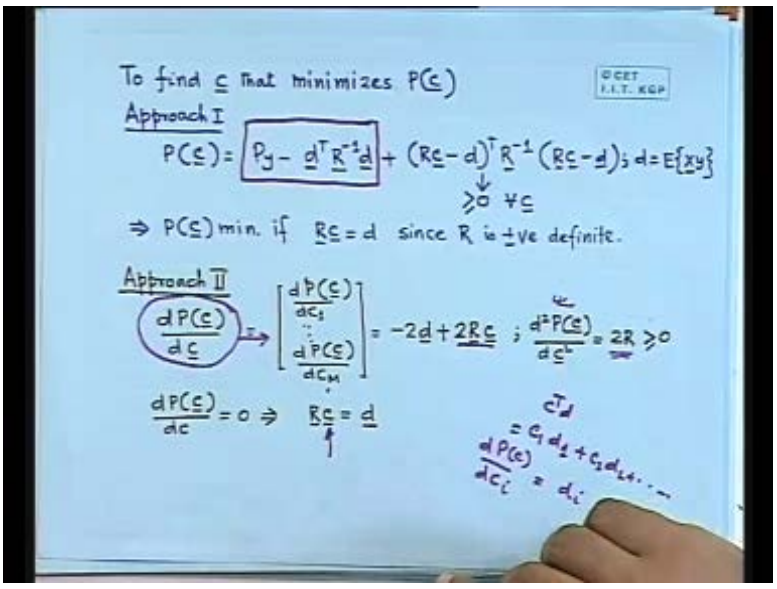
See if you differentiate it with respect to c, this term will be zero, this term will be, what? It will give minus c rather it will give minus. This is d this is d this is d transpose. So if you take minus c transpose d and then you differentiate it with respect to c, what you get? You will get d rather minus d, why?

(Refer Slide Time: 18:09)



Because because, what is c transpose d; c transpose d is equal to c1 d1 plus c2 d2, plus dot dot dot. See if you differentiate it with respect to any c i, so d P c by d c i any particular one, you will get d i. So when, you so here you will get d1, here you will get d2 and then at d c m you will get dm. So you will get the vector d itself, that okay.

(Refer Slide Time: 18:44)

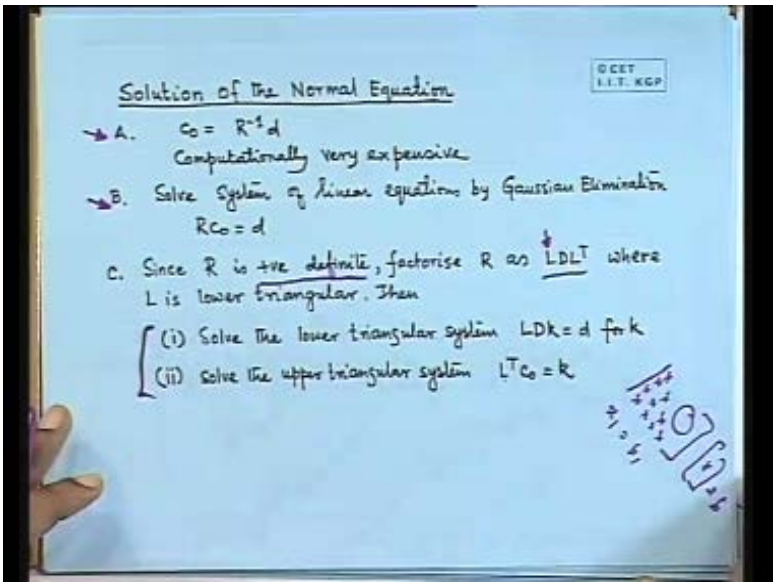


So so here you get minus minus d, here also you get minus d; because c transpose d and b transpose c are same thing. So you are here, you get minus d, here you get minus d and here. What do you get? It is like x square, I mean you can you can actually term by term do it and see; but because it is it is s square x square, if you differentiate it is like c square, c transpose c. So it is like a x square. Say if you get that, then you will get twice R c. If you if you take a x square and if you differentiate with respect to a x, you get twice a x. You can do it term by term just like; I did it here, you can again multiply c transpose, actually take c one, c two cm, and then take R one one, R one two, R one three and do it, there is no problem.

So if you differentiate you actually get this; now just setting this equal to zero, does not ensure that it is a minimum. We have to also see that, the second derivative is positive only, then it is a minimum otherwise it will be a maximum. So if you do d2, P c d c square, that is you you are again differentiate it with d dc then; you will twice R and we know that R is positive definite. So it is greater than equal to zero, this is a matrix, therefore it is a minimum, okay. So now if we set it to zero, then we get the same equation back, that is R c equal to d. So this so the optimal rate must satisfy this relationships, that R c going to d, okay.

Fairly simple problem, discovered long back in nineteen forty's, now the question is how do you solve this equation? R c equal to d, We have to we have we have to compute c. So we have to solve this equation, okay. Obviously now as long as we do not go to the computer, do not work by hand, the the notational we are always right. Obviously c naught equal to R R inverse d. That is very easy to write a where it is much more difficult to compute. Nobody does R inverse d, because I mean computing R inverse is computationally very heavy. So therefore obvious ways, I mean the notational way to say that, you compute c naught equal to R inverse d, you actually cannot do that; I mean actually should not do that. What is the other way? The other way is to solve the equation R c equal to d, just like you solve a x equal to b by doing Gaussian elimination. So you you will find x, so like that you can find c; that is a that it is a better way than inverting it.

(Refer Slide Time: 21:50)

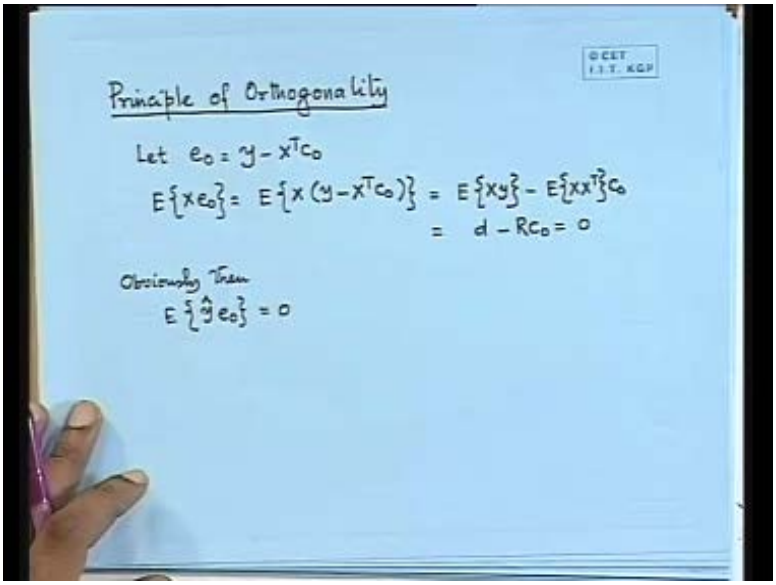


You can also inversion is about inversion will be about, three times a more expensive than this. I mean there are exist a I mean computational complexity result, so if will show that if you want to invert; you will spend about three times effort than this. And sometimes, what you do is you will, these are you know advanced methods of numerical linear algebra. That is you any if R is a positive definite matrix, then it can be shown that R is all can be factored like this; where L is a lower triangular matrix, you know, what is a lower triangular matrix? That their elements are like this, here it is zero, this is what triangular matrix. So R can always, if R is positive definite symmetry then, R can always be factored as L D L transpose where this L is a is a lower triangular matrix, D is a diagonal matrix and U is an basically L transpose is an upper triangular matrix, okay.

Sometimes, it is also called L D U factorization. So it can be factorize like this, what is the advantage? Why we are going to factorize like this, because solving triangular set of equations is trivial. If you have this sort of an equation into x is equal to B, it is easier to solve it. Just you will have a x 1 equal to v 1, then you substitute that, you will get x 2 is to substitute those two. It is extremely simple, if you triangularize the equations. So if you have LDL transpose, then then you can solve it in two simple steps. First is first you assume that that, L transpose c is equal to k some vector. So there you solve L D k is equal to d. See d is diagonal, so the for L D is also triangular. So first you solve for L D k equal to d, which is a

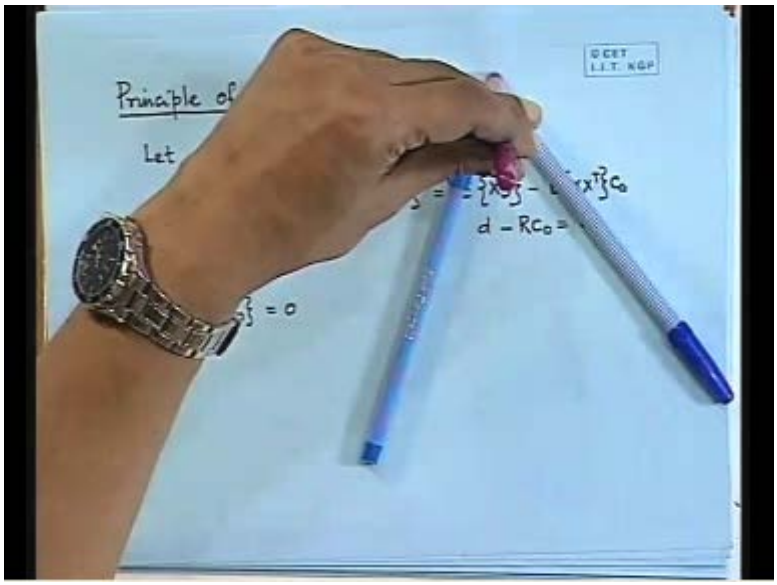
lower triangular system. Then once you get k, then you solve L transpose c equal to k; that is an upper triangular system, so so both are extremely easy to solve. So this is a you know sort of a modern way to solve these equations, okay, which are also numerically stable.

(Refer Slide Time: 24:12)



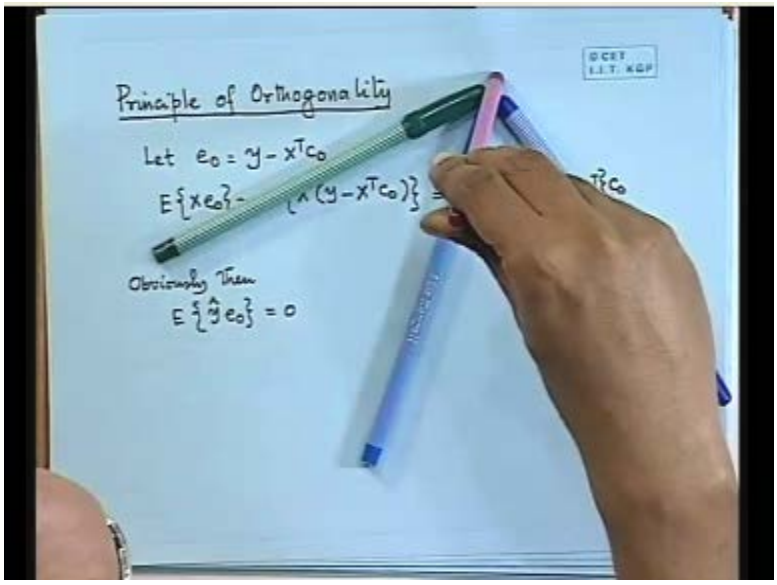
Now there is a beautiful principle, which is very profound in in the real estimation, and will come back again and again. That is the Principle of Orthogonality; which says that the error, that is basic basic idea is that, if you are see you are estimating y, okay. Suppose y is a vector in some space and you are estimating y, let us say with other vector x. So you are you have these vectors which are your measurements and you want to estimate this vector.

(Refer Slide Time: 24:50)



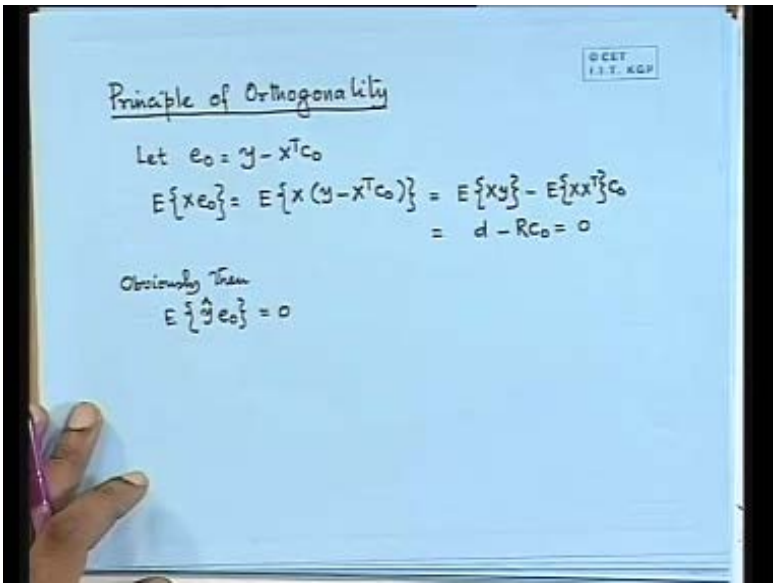
And in this case you are you are generating the, you are generating the estimate by linearly combining axis, which means that whatever if you estimate is going to be on this plane. If you take a linear combination of two vectors, it is going to be on the same plane, right. So that means all your estimates are actually, are in this plane. Now now suppose suppose; this is an estimate this is an estimate, which you have generated using these two vectors on this plane.

(Refer Slide Time: 25:23)



So so what is the error? Error is this vector. Norm of that is length of this this minus this. This is vector. Now the question is when is this, this norm going to be minimum? That is what we want to find. When if you draw a perpendicular on on this plane; that is where it is going to minimum, so which means that, the error vector is going to be perpendicular to all these vectors, necessarily, right. This is like this is the principle, which is which is satisfied in in in many many cases and which has a very nice geometrical interpretation.

(Refer Slide Time: 26:20)



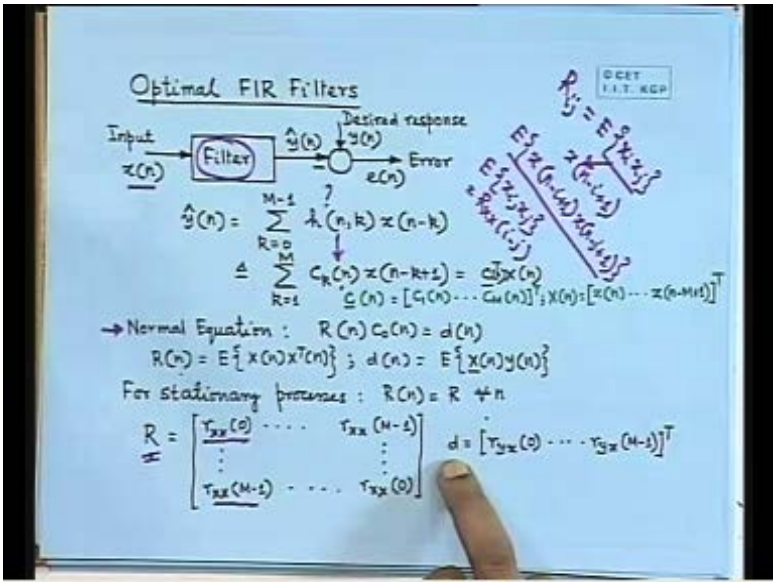
So so what it means, so now the the principle of Orthogonality says that; this error e which is defined as not only e, they are not just e, e for the optimal estimate, not for any estimates e. So now I have I have substituted the optimal estimate c naught, if you put the optimal estimate c naught, still you will get some error. Let that error be e naught, so y minus x transpose c naught is going to give an e naught, this is the error, right. This is your y hat, so it says that the expectation of x e naught which means that, what is the where is now this expectation of x and e naught, that is if you take a product if you take a product of x and e naught, they are going to be and if you take expectation, over all possible random variables is going to be zero. Why? That you can prove prove because, you have the way you have chosen c naught.

So expectation of X into e naught is y e naught, I have just just substituted. Then I have broken up, so you again get the equation d minus R c naught, E d minus R c naught. So

basically d minus R c naught, the way I have chosen a d R c, because it is c naught, so it is zero. So which means that, the error is going to be Orthogonality to all measurements, that you are using fact number one and and obviously, since this y hat is generated linearly using all the elements of x; so so it is also on the same plane, so it is going to be Orthogonality to to y hat. This is this principle will come back and back and back in in in in most mean square estimators. That is the moment you want to estimate, you know want to minimize the length of a vector; which is a which is a quadratic norm, you will always get that the that the error is going to be perpendicular, to the to the measurement which you use. That is the way to make it minimum length, okay. So this is a very profound thing, Which we will get back.

Now let's specialize this case. Previously we did not know, what is x k, what is R. We took a fairly general case; that is x care some measurements of some other quantities and y is some other. How they are related we did not bother, right. So now we will consider several special cases. For example, one one standard problem is to define that, if you have if that you have a single input, non-specializing, you have various cases.

(Refer Slide Time: 29:03)



That is now you do not have previously, I had taken k measurements x k's and each one, I was assuming a vector. Often what we do is we choose, suppose we are we have a we have some desired response, okay and we want and we have got some inputs now. What I what I want to know is that, what is that filter through which if I pass this input, I will get this output

very close to this output? This is a this is a valid problem, okay. So then you write, so now we are talking about linear filters, okay. So when we are talking about linear filters, obviously so now my x one, x two, x m are are the samples of the same signal. So now I have see I am I am now specializing, the the the first problem that I have done. This is this is a case we will get maybe we can have many other cases; which are of great practical importance, which we can solve by exactly the same method. So here now I am defining y hat n as this, so I am now trying to generate the estimate of y hat n, using the past estimates of x. Now rather the using the past measurements of x and I want to compute these, okay.

So because why you have written it like this, because this will turn as linear and not only linear, it is FIR means it is finite impulse response. See normally the impulse is for, if it is an arbitrarily linear filter; then it should be h, zero, M basically it should be zero to infinity, if it is a causal filter. If it is a it is a non-causal filter, it should be minus infinity to plus infinity. But so here we are talking about causal and FIR filter say; FIR filter means that after sometime, the the the impulse response sequence goes to zero. Only there are finitely many h one, h two, h three, h four up to h M. After m greater than zero, h M is zero, so that is why this summation, gets truncated okay, so if that is why its optimal FIR filter.

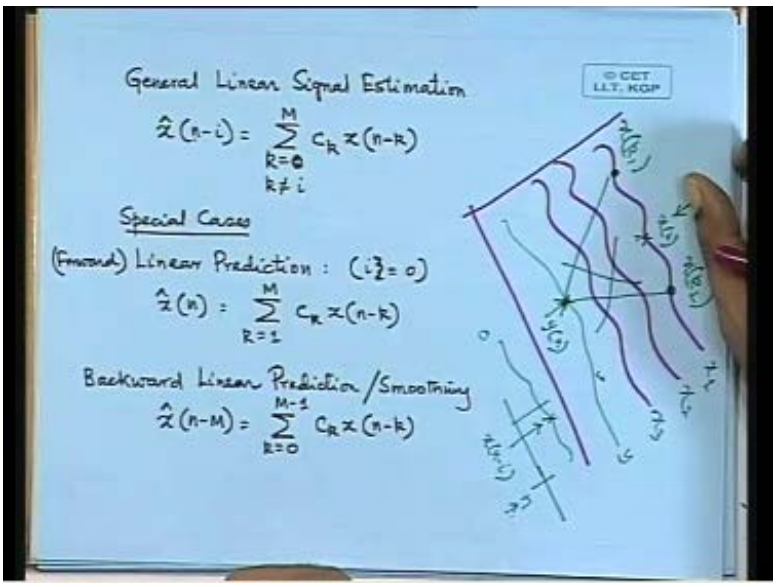
And naturally so if so you see that, that estimating this filter is the same problem that we have considered previously, it is our old problem. So h n k we can call them c k n and put a put zero to M minus 1 equal to one is equal to M. Say so this is I have just changed the indecent error there; and I am trying to show that this is the same problem that, you have solved previously. Again and trying to estimate c k, only now my x k's are x n minus k plus one, it is the samples of the same signal. So it is a same problem, so the problem of estimating this filter; such that if you give x n input, you will get y n output close to y n output. It is the same same problem, that you have already solved. So we will get this equation again, only we will have to put appropriate vectors in this case.

So in this case, what is so again we will get a it is the same problem, so again c zero n will give you this R n. I mean R inverse d, the same thing you will get. Now in this case, what is a what does the R look like? What was R?, R was x i, x j, previously what was R? The ijth element of R is expectation of xi, x j. Now what is xi here? See see this is x n. So it is what is x i, it is x n minus i plus 1. You put i is equal to one, you get x n. You put i is equal to two,

you will get x n minus one. You put i is equal to M, you get you get x n minus m plus one. So this is x i, so what is x n minus i plus, so what is it is x n minus j plus one, expectation this is R i j in this case, and just put x i x j. Now what is this? this is nothing but the if the if the R process is stationary; then then expectation of x i x j is equal to R x x i minus j, it does not depend on what is i what is j. It just depends on i and j difference, remember stationary process?

White sense stationary, so the mean is constant and the autocorrelation is depends only on the difference between two points, right. So, x t x t plus tau will give you R x x tau. So so so if you do that, you will get here you will get R x x zero, up to this so so you will get these terms. So it is so it is again symmetric; you can see and this so this is 'what R looks like' in this particular problem. It is a just just a special case, okay. And this is what d looks like, so again you have to compute, you will have to solve for R c zero equal to d equation, right.

(Refer Slide Time: 34:51)



See this is a this is a general problem, okay. General Linear Signal Estimation, that is what is what we are doing is that, we have the signal x case; these are my measurements and you have this signal y here, so this is x1 now that was our original problem, okay. So in this case; we don not have that, in this case we got we have is we have only one of them and we have several samples, okay. So suppose in general I can define a problem in which I can to generate a particular suppose, I want to generate this one that is y n. Now question is what

data I am going to use, so I can in general use data from some k 1 to k 2. If I have the whole strings suppose, I have acquired data using some data acquisition card; I I have acquired data and I have what the whole strings x stored in my computer. Then then there is no problem all this is not a online case, that that there are signal is coming. I want to do something. Signal is already there, from let us say from zero to infinity; so then for for estimating this y n, I can in general use any times some span of x, and then see, what will be best. Maybe if we value this, we will we will get better result actually, right.

So in general I can define a problem in which, I am it is also of interest sometimes that that we want to generate the. That is using the using these measurements; we can also try to generate x n, that is we are going to use data from x k 1 to x k 2. Which we will not use x, the value of x n; and we want to estimate, what is going to what is x n? This is also possible This problem, if of importance because of two things because of the fact; for example, we will just see special cases, so we can define a define an estimator where I want to I want to see I am standing at n suppose. I am standing at time instant n okay I am here, this is n. So if I am here even if data is coming, streaming in then, I have all data from zero to n I have now.

In general I can try to I can try to generate, an estimate of x n minus i using data from n to n minus m. This is possible, that is I use the data from this to this, to generate x n minus i while I now using x n minus i is contain here. So I do not use x n minus I, I use other things. So that is why you are trying to generate a linear estimator using data from x n put k equal to zero, you will get x n, put k is equal to capital N you will get n minus m and you are not using x i. And you want to generate an x n, you are not using x n minus i and you want to generate an estimate of x n minus i this is a problem.

There is a general problem. Now what are the special cases? A simple special case is putting; put i is equal to zero you will get a prediction problem. That is you now have data from n minus one to n minus m, and you want to generate n, right. So you are using past data to obtain future, that is prediction, which is a very very important problem used in many many cases. So if you if you specialize the case for i is equal to zero, you will get a predictor; that is why I sort of you know if you just solve a general problem, you you get everything.

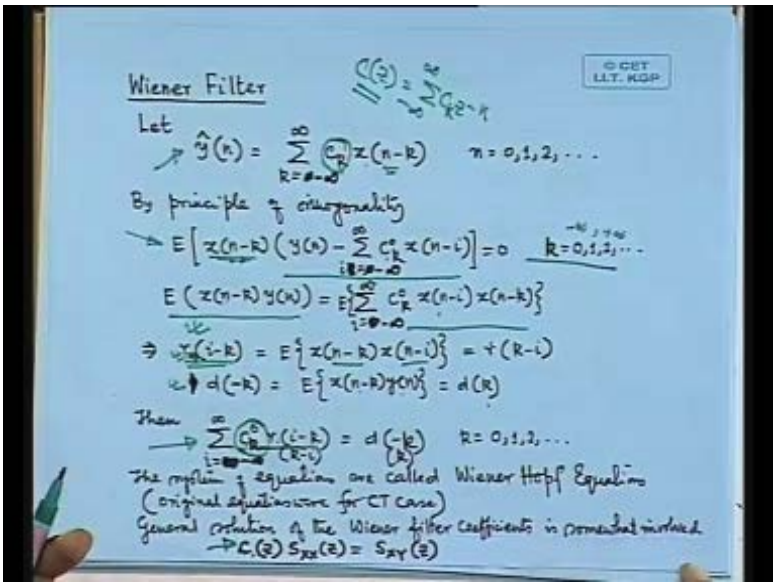
Similarly, there is there is something called a backward linear prediction. If you see D S P books, they will say forward linear prediction and backward linear prediction. It in a backward linear prediction, as a bit of a I mean I am not comfortable with the term; because because prediction in English usually means, going forward but so so basically that what backward linear prediction means is; that you use now, you see. You use from x n to x m minus one, so you have you have got future data and after you got some more data you want to estimate a past value. So, you are there you are not going forward in time but you are going backward in time.

So you have all your future samples, you are using to estimate a past value that is also sometimes is called smoothie. This thing is called smoothie; that is after you have, you had some measurement now. But after you have many other measurements; you can make a better guess of that, especially when the signal is noisy. So the in this way, you can define several special problems; so all of which will form in the same way and all of which can be solved by exactly in the same way, by solving the normal equation. That is what I am trying to show, okay.

Now we have, what we call as this is the this is this is the origin of this; you know, there there is a great mathematician called Norbert Wiener, who actually formulated this problem and solved it in nineteen forties. And actually this Weiner Kolmogorov, whereas these are the I mean fathers of modern statistical signal processing. So in fact Wiener and Kolmogorov both both solved it, I I mean independently around the same time this problem, okay. We have solved it in continuous domain, Kolmogorov solved it in discrete domain and as usual, it is it is known more by the name of Wiener than by Kolmogorov, because Wiener comes from western world.

So so here we are defining y hat n like this. What is this this? This is a a non-causal infinite impulse response, filter. So it is the most general case. I am not saying that; it is going to be causal also I have taken minus infinity to plus infinity, and since I have taken up infinity. It means that it is infinite impulse response, so that is this is the most general case, okay.

(Refer Slide Time: 41:19)



So again the same thing, only thing is that only here there is a problem. What is the problem? The problem is that, the the the number of terms are infinite. So all the matrix size is everything is going to be infinite. So you cannot directly write a matrix equation here because; the matrix has number rows number of columns, everything is infinite. That is why you write it in a different way; but it is a same equation, so you see again if you if you if you apply principle of Orthogonality. What do you get? You get each element x n minus k is going to be Orthogonality error; that is this c k should be chosen in such a manner, that this is your error. This is your model, so y n minus y hat n is an error. So the error is going to be Orthogonality to all the data, that you are used to generate it.

Principle, Orthogonality says that, the error is going to be Orthogonality; to all the data that you are used it to generate the estimate, which you have that error. So the so so the data which you have used to to generate, the estimate at these; so if you take anyone of them, the error is going to be Orthogonality, correct. And and this will hold for all data, now k varies for minus so, you can take k everywhere. Actually actually it need not be zero one two, it can be minus infinity and plus infinity. So initially I put it for a causal filter then, I have changed the notation, that is why I did not change it here, it should be minus anything need to plus infinity.

So if you do that now, this equation you can you can write in a different way because because; you do not simply because, you cannot write it in terms of a matrix, because the matrix is of infinite size. So now you have to write it in terms of in a different way, it is the same matrix an. So now I would like to again multiply and I will put these terms on the left hand side, the put these terms on the right hand side. Again now it is a infinite summation. So you so you get this. This is your correlation matrix r x y between y. And this is no no no if I if I define this as the autocorrelation matrix, x x n minus k n minus i will will be r? x x i minus k or k minus , all the same. And d of minus k is equal to d k n; and that is x n minus k y n, this is what I am defining, this clear to you?

So correlation matrix, so the correlation matrix, depend on the difference between the indexes. So what is the in..? What is the index here? n minus k, the index here is n minus y. So what is the difference? i minus k, so as simple as that. So if you put that, then then if you put these notations; these are just notations, I means just n say n place of this cumbersome sums, I am giving the names. Then I can write this, then now you put it here. So you will here, you will get c k zero into R R k minus i or i minus k all the same, and d k. So this equation is very famous; it is called the Weiner Hoff's equation. So for this general problem, you have to solve c k; by by by solving what is known as, the Weiner Hoff's equation for all k.

So you have a infinite set of equations, from which you have to solve these elements. These these infinite sets of elements, now now what does, how do you do that? How do you solve an infinite set of unknowns, from an infinite set of equations? So for doing that, what you can do is actually what people do is, this this is why frequency domain is so useful. Actually there are some cases, where you can get nice close form solutions of this problem. It is not that it is not that, this problem is not solvable just because, you have an infinite number of terms. So for doing that, what you can do is you just observe that, what is this? This looks like a convolution integral now. This this is c k into R k minus i, in the previous case we had h k, x n minus k and and you are summing over i, right.

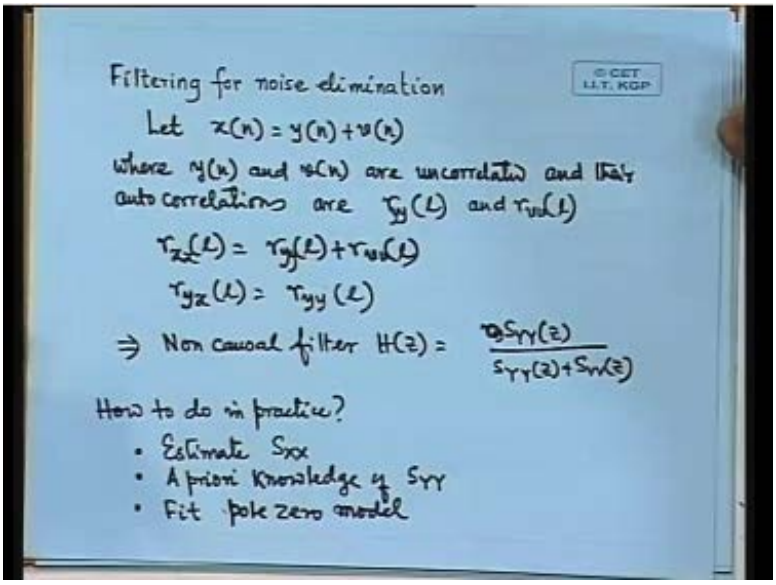
So you see that, you have this is this is like convolution integral. So if you take a z transform of this, convolution would become product. So then you will get, if you take z transform on the left hand side, you will get c z; which is your filter transfer function, after all what is c z,

these are the impulse response coefficient. So c z is sigma minus infinity, what is c z? c z is sigma minus infinity to plus infinity, c k, z to the power minus k. So this term will will this is the convolution of two, convolution of the impulse response sequence with the, with the input so called input; which is the correlation sequence. So if you take great transform, you will here you will get, the filter transfer function c z. And here you will get S x x z, which is the power spectrum of x. Autocorrelation function, if you take, if you take, if you take z transform, what will you get? You will get the power spectrum.

And this here, you will get cross power spectrum of a S x x. And y so the so so the what is the, why did we do this? I mean what is the beauty of this now? The beauty of this is that, if you can estimate these and if you can estimate this, by some other means then by simply dividing you will get the transfer function c z. Which may or may not be rational in the rational form, remember that all transfer functions are not rational; in the sense that that every transfer function cannot be expressed, as a ratio of two polynomials v z by a z. For example, there are there are various cases; for example e to the power z, e to the power z is not expressed as as b z by a z, it is expressible only if you take an infinite number of terms. So it is important as an irrational in transfer function, it is still a transfer function, right.

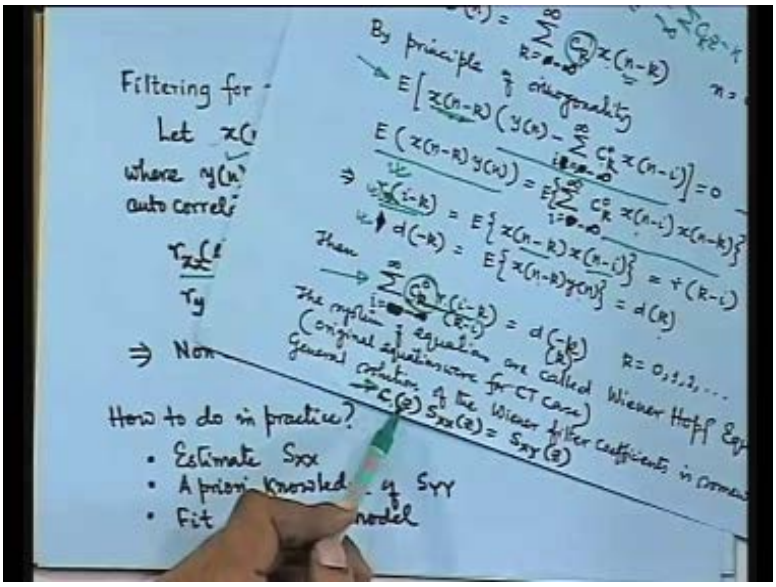
So, here also, you can at least if you just numerically, if you compute S x x and from signal and numerically if you compute S x y, then simply by dividing these two power spectrums, you cannot need c z. At least you are going to get the form then you can approximate c z in some way, but what I am trying to say is that, in this practically possible solve for c z, in some cases, it is not that it is not possible, if you go to the frequency domain. So I mean this is how the Weiner Hoff's equation is generally solved, okay. And and this is why frequency domain, is so so so interesting. See, then even one by one minus z; which is a nice transfer function, it also has an infinite number of coefficients. If you see it as one plus z, plus z inverse and it has infinite number of components, but so what does not mean that; it cannot be represented very compactly using a transfer function one by one minus z, it can be.

(Refer Slide Time: 00:49:25)



So for example, this this problem we can now apply, to to this case. See this is a very very common case; that is if we we you have a some actual signal y, which is corrupted with some noise v, and this is your measurement x. Now from x you want to estimate, what was.. what was the actual time. This is this is one of the central problems in instrumentation, there is some actual process variable; you want to sense it, when you sense it you get noise mixed with it. So from the measurement can you get the actual process variable? Standard problem, very very common. So so what will happen? Now we apply odd theory to it. If we apply odd theory to it, then we will get that or non-causal filter H z, so now what what is r x x? r x x is r y y plus r g g, just take r x x because because; y and g are we are assumed un-correlated, so all x y term should get all y v terms will go to zero. So you get only y squared and v squared terms; so you get r y y plus r v v, there is no cross term here because, I have assumed that y n here, I am correlating. And now what is r y x? r y x is nothing but, r y x if you multiply y here; you will get r y y and you get r y v, r y v zero because again, they are un-correlated. So r y x is r y x, so now you apply this this.

(Refer Slide Time: 00:50:58)



This so now, so you you, what do you have? You have what is the optimal filter, it is S x y by S x x by solving them in a half equation in the frequency domain. So you get, so the non-causal filter is S y y, in this case the state is S y y by because, S y x is equal to S y y. And here you have to put S x x, which is which is S y y plus S v v, right. So if you have some knowledge about the noise; see if you know if you know this and this, you can you can compute this. Or in other words; if you know even this and this S x x and S v v, then also you can compute this, because because from because S x x if you know, how will you know S x x. This also actually take that measurements, x is a measurement so it is available. So you can estimate S x x and if you know, what is the characteristic of noise; maybe by maybe by using very very expensive and accurate instruments, once you have calibrated, what is the exact noise is. See you cannot have you cannot have that instrument every time, because it is very expensive.

So once you have calibrated and found out what is S v v, now in the field you cannot use an instrument; because it is so expensive because, it may be delicate etcetera etcetera. So in the in the field, you use a normal measurement, get the normal measurement from a sensor, which is not so accurate and then you use your filter, that is the idea. So so if you know S v v and if you know S x x, then you can compute this filter, using this formula. Now this is again, this maybe an arbitrary form; but then you can you can approximate the form using some

suitable relationship, that you will get some this will be a general function of z. So now you have to you have to you have suppose; you looking at the function and say that if I if I have to approximate this function reasonably, well I have to choose a six corner filter then, by choosing coefficients I can match this frequency domain characteristic, that you can do.

So you can estimate S x x and then fit as fit some pole zero model, eighth order, twelfth order, and then if you use that filter then at least theoretically, you can try to get x n. Now still it is a non-causal filter, it uses minus infinity to plus infinity those problems are there. So we will stop here today only, we will note that; see even for solving this problem, there are two central hard ways, which are not adjust in this class. That ever where we are using this, R x x this F x x x, where where do I get them from?

So there has to be a way way, if you have to know; if we if we have to apply these, we have to know that from a given set of signals, how we compute its auto correlation function, because autocorrelation is in terms of distribution and all nobody will give me. People can only give me data, from their I have to estimate it is statistical properties. So it is very important to understand, how do we, from given a data, how do we compute a power spectrum? Estimate a power spectrum and then, use it for say in such problems? And how do we compute the autocorrelation, that is very important to know and that will do... that we will take up in the next class. Thank you very much.