Probability and Random Variables

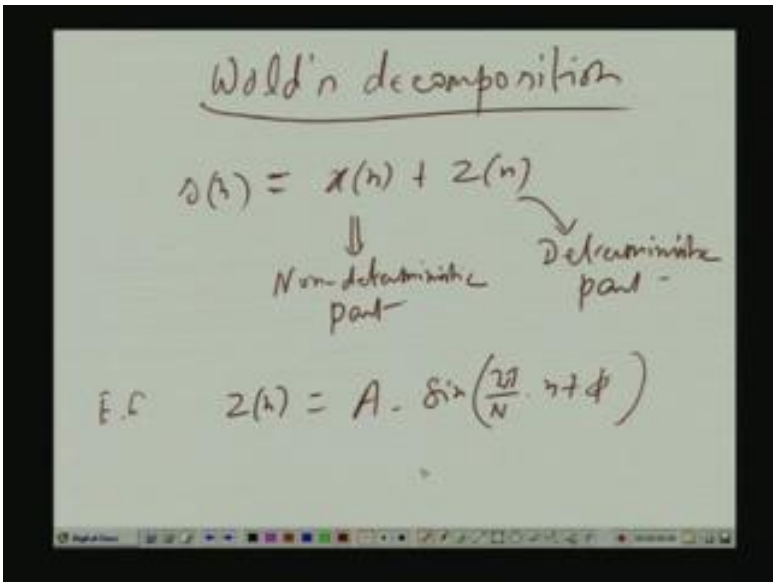
Prof. M. Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 37
Spectrum Estimation–Parametric Methods

So, today we will be considering as I told you in the last part of yesterday's lecture. I will be considering parametric method of spectrum estimation that is spectrum estimation by modeling. And I have already told you the elementary ideas of modeling that is you assume that the process is generated by passing a white sequence through a rational LTI system of transfer function saying z which can be all pole 0 or pole 0 both. And therefore, the power spectral density of the received sequence would be nothing but mod square of h e to the power g omega; h e to the power g omega being the transfer function.

So, mod h e to the power g omega square into some constant. The constant denotes the input power spectral density now input is white. So, input power spectral density is flat which is equal to some constant. So, in this case the entire spectral density estimation was done to identify that model or a estimating the parameters of those model. You know the coefficients that occur in the numerator and denominator collinearly and exist. Now, before I go further into that I mean I told you yesterday also that there is the justification for you know making this assumption. And going by this method and that comes from what is called Wolds's decomposition of famous theorem by a great statistics person.

(Refer Slide Time: 02:07)



This was Wolds's is called Wolds's. This says that given any random sequence sn you can always decompose it into 2 parts 1 is xn. And other is say zn where xn is called non-deterministic part and this is called deterministic part. And these 2 are mutually uncorrelated. What is deterministic part n also is a random sequence xn is a random sequence then if something is random what is deterministic about it. That is a very you know that that is a kind of puzzling question. That zn is random, because it is coming from it is basically originating from the given random sequence sn.

So, what is why do you call a deterministic what is meant by a deterministic random sequence. Well a random sequence is called deterministic, if any current sample say za can be accurately given or is given directly by a linear combination of all its past sample. All or part of its past sample that is if zn can be written as a linear combination of z n minus 1 z n minus 2 z n minus 3 dot dot dot up to z minus infinity.

That is if zn is equal to some c 1 times zn minus 1 plus c 2 times zn minus 2 plus c 3 zn minus plus dot dot dot dot. Then that means, that from the entire past of that process zn I can accurately describe zn without any errors. Such kind of process is called deterministic process. To give you an example suppose you take zn to be a sinusoid of this form where A is the amplitude is random. What is sin 2 pi by N into n? It is a discrete sequence which sinusoidal sequence is periodic over a period N and the amplitude A is A.

So, every time you perform an experiment you find this sinusoidal sequence, but amplitude changes, because amplitude is random. This process according to me is a deterministic random process is random, because amplitude A is random. So, sometimes it can be very small sometimes it can be very high. So, you get sinusoids from various amplitudes in that sense it is random. But, see if you know all the samples of a particular zn sequence belonging to a particular period take 1 period of zn.

Suppose, you know all the samples then you can accurately describe all other samples in terms of the samples taken over a period, because of the periodicity. Because, whatever you observe in 1 period they only repeat. So, if you know all the samples within 1 period then you can predict any future sample or past sample accurately without any error. Instead of this you could have also have a phase here some phase you know. And that phase could be a random variable or both amplitude and phase could be random variable, but still that does not take away the periodicity.

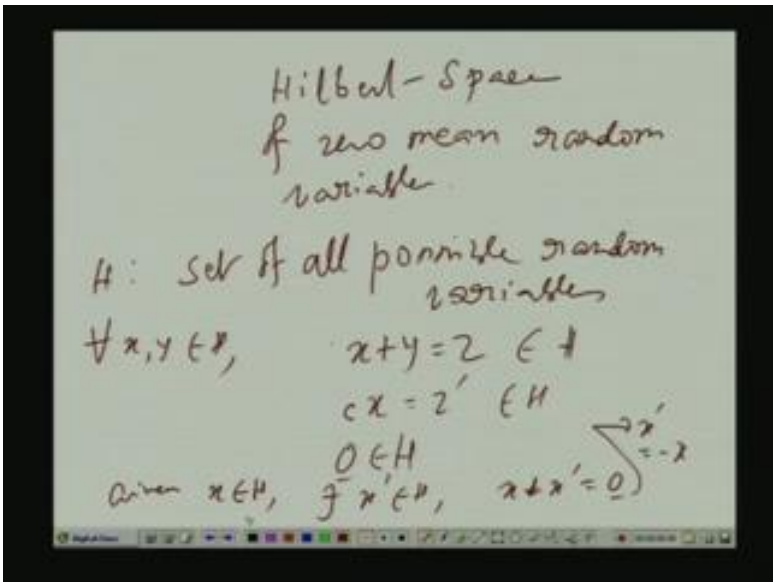
So, if you still know the samples within the period you know everything about the process. You know I mean I mean you know all the other samples of the process accurately. This kind of processes is called deterministic processes. If you take just spectral analysis actually they give rise to what is called line spectra, that is you know just a series of impulses derived delta function or impulses in the frequency domain.

Then, obviously if you take sinusoidal function like this and if you take a autocorrelation function which is also sinusoidal. Then take its Fourier transform dtft. You will; obviously, get impulse functions, because any sinusoidal function if you a Fourier transform give rise to impulses. So, in general this deterministic random part zn that give rise to lines. That is, impulses in the spectrum for the time being we will not be considering this zn.

So, we will considering we will concentrating, now purely on x of n that is which does not have any line spectrum. What is xn? Xn is the purely non-deterministic part. That means giving any xn you cannot write it exactly as a linear combination of all its past samples. That is if you want to predict x of n from all its past samples as a linear combination you cannot predict xn with 0 error. There will be some error of non zero variants or non zero power that is some non zero error I mean I would not care non zero error, because error for a particular case could be 0, but there is that error will be random.

So, its power variance sometimes it could be 0. So, sometimes it could non-zero, but the variance which is important you know that variance will be non-zero. Such a process is called non-deterministic process and for all this spectrum estimation by the parametric modeling parametric method will be concentrating on such non-deterministic part. I will not give you the proof of this Wolds decomposition, but I will consider this x of n. Now, since I am talking of predication and all that is better that, we again take resort to our previous notion previously described notion of is Hilbert space of random variables.

(Refer Slide Time: 07:47)



May be to make life simple we can assume all random variables including the random sequences xn to be 0 mean. So, Hilbert space of this is a set of this H set for all possible random variables 0 mean I am not writing 0 mean again and again it is 0 mean. Then there is rule of addition that is for all x y belonging to H. You know what is this x plus y which is z belonging to H is called we say we say that is closed under addition. What I mean be x plus y equal to z?

It is the usual way we add 2 random variables to generate another random variable. But since, I am considering a set of all possible random variables the resulting variables if you call it z stills belongs to H. So, when you add 2 elements of this H the elements sometimes are called vectors they are not like position vector, but just right terminoginally they are called vectors. So, if you take any 2 vectors that is any 2 random variables belong to H there is rule of addition involving the 2. And what you get of the addition that also still belongs to H.

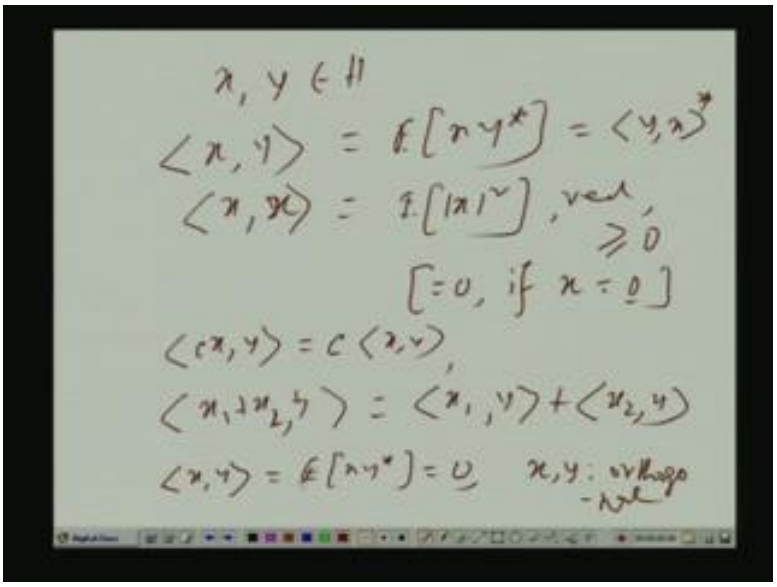
Similarly, there is a rule of scalar multiplication. If you take c c could be a real or complex number. In fact, for our case we will be assuming you know the associated field of numbers to be complex. So, c is it is any complex number. So, c times x if you call it say z prime that also belong to H. C is any complex number and there is rule of scalar multiplication. In the usual way like that we know if there is x is random variable and c is a scalar what is c x is again a random variable.

So, in each trial whatever value x takes that gets multiplied by c and that is the value assigned to z prime for that particular trials that is a physical meaning. Now; obviously, z prime also should belongs to H, because H consist of all possible 0 mean random variables mean is always 0 in all these operations. And the range is to 0 random variable 0 random variable. I will put a bar here to indicate it is a variable vector 0 random variable it means it is such a random variable which always takes 0 value.

We say that x 0 value probability 1. When you add 0 to any x what you should get back is x that is a property of 0 vectors. Then given any random variable x there must exists is this negative; there must exist given x there exist this is there exists say x prime element of H. So, that x plus x prime is equal to this 0. In that case x prime is called this means x primes is called actually negative of x. And we denote it by minus x. Obviously, I explain I am what is the explain physically. Whatever, value x takes negate that that value is assigned to x prime.

So, x prime is that kind random variable. That is physical meaning of 0 all rules of you know addition as multiplication like associativity distributive commutativity and all that they work there values here also there is a definition of Hilbert space. Like ordinary like our conventional 3 dimensional vector spaces involving I vector j vector k vectors. You know here also we define something, something called a dot product which in a general says is called inner products rather dot product.

(Refer Slide Time: 11:40)



So, for any x x for xy belonging to H we define the dot product like this, this is inner product and we define it like this xy star. And x we takes actually there are certain basic properties that this inner product should follow it you know those properties are satisfied.

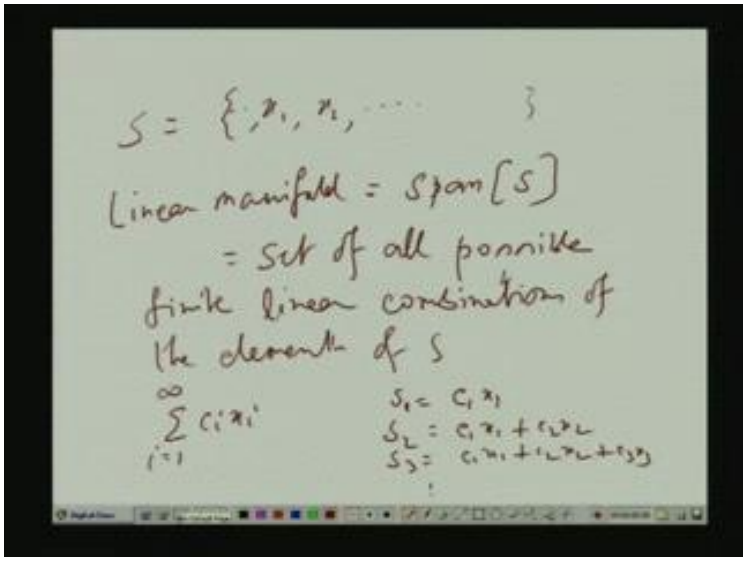
So, I will not go into that I just define inner product and if you take the inner product with itself it gives you mod x square variance. That for those properties are simple e of and x comma y should be conjugate of y comma x that is satisfied here Exy star is Eyx whole conjugate Ey star x star whole conjugate. So, this is actually this should be yEx star this is 1 property x with x must be real greater than equal to 0 equal to 0 if x is the 0 vector only then. If x is on the 0 vector even if you take take 0 value some times, but on some other some other occasions it does not take 0 values. Then mod x square can never be expected value of the mod x square can never be exactly 0. So, it is 0 only if x is a 0 random variable of that Hilbert space otherwise not.

And the other thing is cxy should be c times x comma y that is initially says y if here instead of x you call it cx c can be brought out. And then linearity x 1 plus x 2 comma y should be x 1 y plus x 2 y that is again satisfied by this inner product by this correlation definition if it is drawbacks if you put x 1 plus x 2 within bracket times y star you can break it up and we can write in this form. So, all these conditions are satisfied if 2 vectors if x y that is equal to Exy star 0 then we say x and y are orthogonal. That is in case they are uncorrelated also. So, if random variables are uncorrelated 2 0 mean random variables are uncorrelated they are orthogonal.

A set of vectors say finite set of vectors will be linearly independent if no vector can be written as a linear combination of the others. If that is suppose you are given 2 vectors 3 vectors x 1 x 2 x 3 that is 3 random variables. If you cannot write x 1 as a linear combination of x 2 and x 3 or x 2 as a linear combination of x 1 and x 3 or x 3 as a linear combination of x 1 and x 2 then this is a linearly independent set.

If it is not if any of them is linearly expressible as expressible as a linear combination of the other two; that means, this is redundant. So, this can be thrown away other 2 can be kept and again you examine the linear independents of those 2 and so on and so forth. You can reduce the set to as smaller set where the smaller set is linearly independent. We all know this, because we have discussed all this the earlier. So, I will not go too much into it.

(Refer Slide Time: 14:59)



Then suppose, you consider just a set x 1 x 2 dot dot dot dot dot dot the set can infinite the set is then linear manifold is actually we also called span of s. It means set of all possible linear that is I would say finite linear combinations combinations actually we do not it all this, but I am just take its opportunity to introduce few basic notions to you people. Finitely the set of all possible finite linear combinations of the elements of s that is what is this linear manifold a phase. That is it can consist of all the variables x 1 x 3 dot dot dot then may be c 1 x 1 plus c 2 x 2 or c 2 x 2 plus c 3 x 3. That is linear combinations involving 2 variables linear combinations involving 3 variables linear combination involving 4 variables set of all such finite linear combinations that is called linear manifold.

That is a vector space is it is, because if you take any 2 elements of that linear manifold 1 is 1 linear combination another is another linear combination. You add that 2 you get another finite linear combination. So, it is closed is it not? So, you still remain within the linear manifold, because linear manifold consist of set of all possible finite linear combinations. So, if you take 1 entry of linear 1 element of the linear manifold which is some linear combination of these elements. And take another element of this linear manifold which is again another linear combination of the elements of s.

Then add that 2 resulting thing also is a linear combination of the element some linear combination; some new linear combination of the elements of s. So, it also belongs to linear manifold. So, linear manifold is a vector space similarly if you take any element of linear manifold. That is, some linear combination finite linear combination of the element of s multiplied by scalar resulting thing is still a linear finite linear combination of the elements of s.

So, again it belongs to linear manifold which is closed. So, I you can verify all properties linear manifold is a vector space. It consist of 0 element, because it takes it manifold means set of all possible linear combination of s means combination of the elements of s means each element of s also belongs to linear manifold. So, the 0 also belongs if even if the 0 does not 0 is not in s you can multiply x 1 say by 0 scalar 0 you get 0 so on and so forth. Then, but remember linear manifold given consist of any series infinite series that, x 1 plus x 2 c 1 x 1 plus c 2 x 2 plus c 3 x 3 plus c 4 x 4 dot dot dot dot up to infinity.

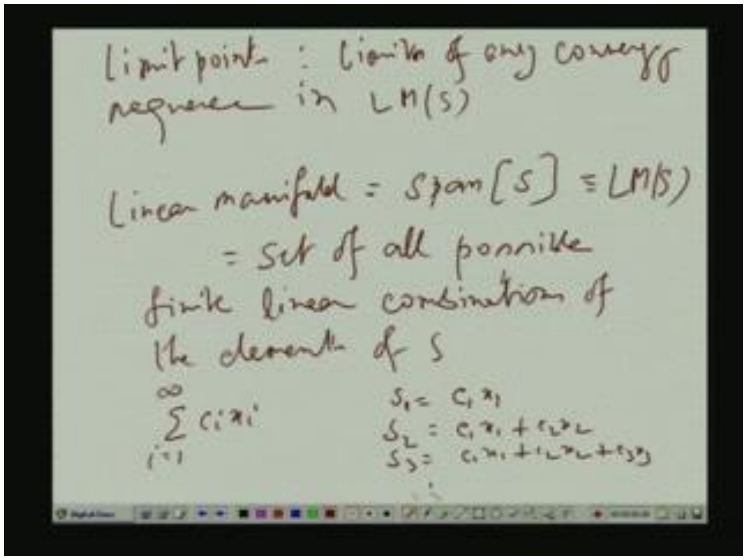
It only consist of finite linear combinations, but for dealing with these problems you know we need also we need to handle also a thing like series; infinite series involving: x 1, x 2, x 3, x 4 like that how to handle them. Suppose, you have got a series like this ci xi ci r is a sequence of scalars and xi is belongs to s. So, I equal to say 1 to infinity. Now, any set infinite linear combination actually it is not define it is not belonging to a linear manifold linear manifold consist of only finite linear combinations.

The way we write is we write it like this some finite sums you know s 1 you take c 1 x 1 s 2 take first 2 terms c 1 x 2 1 c 2 x 2 s 3 that is c 1 x 1 plus c 2 x 2 plus c 3 x 3 dot dot dot dot. You see s 1 s 2 s 3, they are all finite linear combinations of the elements of x. So, s 1 s 2 s 3 s four they all belong to linear manifold. And what is the limit of the sequence of s 1, s 2, s 3, s 4. That is what is given by this infinite summation, because from s 2 to s 3 when we go we get 1 extra term s 2 to s 4 another extra term.

So, as I go towards infinity I get all the terms of this infinite summation. So; that means, this sequence s 1 s 2 s 3 each element here s 1 or s 2 or s 3 or s 4 each of them belongs to the linear manifold. Now, if the limit also belongs to the linear now s 1 s 2 s 3 s 4 dot dot dot dot this sequence each element belong to linear manifold. Now, this its limit is limit of the sequence is this infinite sum. Infinite sum is not part of linear manifold, because their variables consists only a finite combinations.

Then what you say we not we not only keep the linear manifold you know want to expand it. We want to add some more some more variables or some more points to the linear manifold to make it bigger and those points will be called a limit points. Limit points are what you know I cannot give you all details here. But, just for mathematical correctness I am saying all these limit point.

(Refer Slide Time: 20:25)

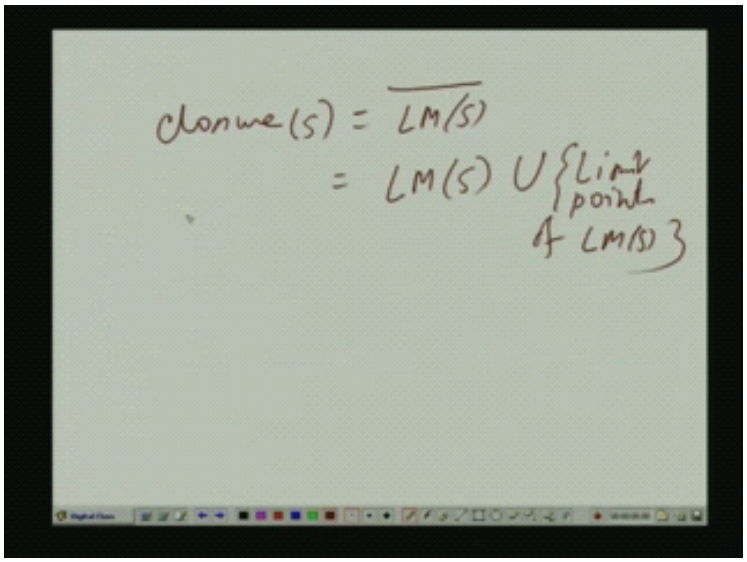


Limit point means, if you take any converging sequence limits of limit of any converging sequence in LMs LM means linear manifold this is LM s. That is if you take a sequence in a linear manifold like: s 1, s 2, s 3 and is giving the that they are converging that is s. It becomes bigger and bigger s 1 s 2 s 3 s 4 s hundred s billion s 10 billion as you go further further what happens that you basically converge. That is I mean even if you add new terms is not changing much like that if it is converging.

If that be, so that, the limit of that converging sequence is called limit point. Now, if we add all the limit points that is append not add not addition we append the set of all limit points to

this linear manifold of the base. Then what we call, what then it is called the closure of s? I will also call it the vector space of s the closure of s.

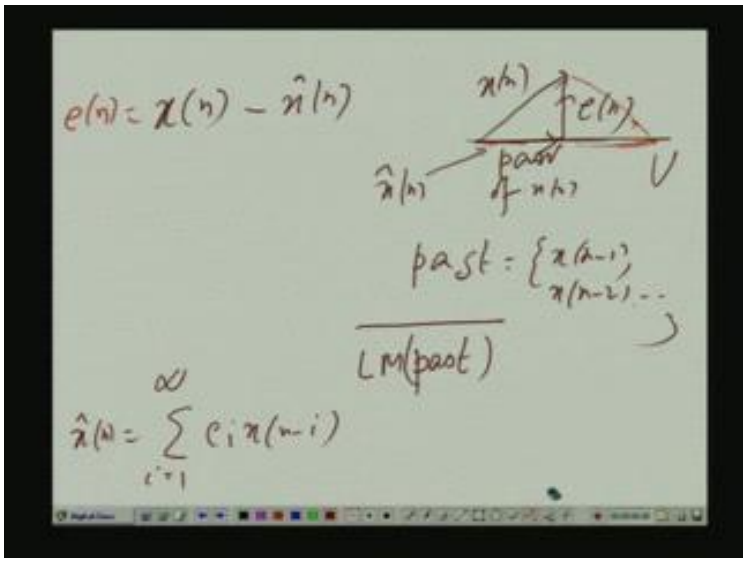
(Refer Slide Time: 21:57)



That is denote it like this, it is not closure you can say closure fine. We denote it like this LMs first linear manifold and then put the bar. That is take the closure of that is nothing, but union linear manifold of s union limit points of LM of s. Then within this closure thing we can write you know if you from any sequence any converging sequence taking the elements of the LMS. Or even this closure of s and the sequence you know is infinite it is an like if you form a infinite sum that is the series.

We can happily write we know what its meaning that we can given a series we can always write in the form of partial sums like: s 1, s 2, s 3, s 4 like that that is the sequence. And since, it is converging we know its limit limit also must belong to this closure, because all the limit points are included in closure. So, we know what it means it is not defined. So, just for mathematical correctness I did so much. Now, I come back to the vector spaces; that is Hilbert space of random variable issue.

(Refer Slide Time: 23:28)



We are now back to this random deterministic non deterministic random process xn. Suppose, you want project you want to predict xn what we do like you know the conventional vector space of this is a vector v. And this is 1 space say u. You want to u you known it span by some vectors you want to predict v in the best way from u. What you do you take an orthogonal projection? So, this is your prediction this is the original 1 this is the error. Error is orthogonal to this plane or to this space u same physiography works here. That theory for this I am not doing what I am saying is this suppose xn as I told you xn is purely non deterministic.

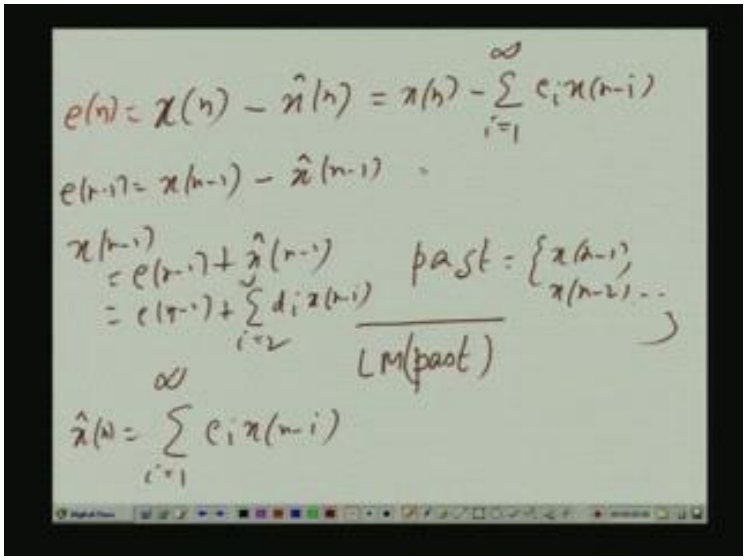
So, xn cannot be xn if this is like you know this is xn. This is xn and this is the past of past of xn that is it consist of xn minus 1 xn minus 2 xn minus 3 all the closure of that set. If you take a set if you take a set say s or may be past if we take we call it past is this set xn minus 2 dot dot dot. Then you take linear manifold of that this set and closure of that. This is given by u here. And you want to project you want to obtain xn as a best possible linear combination of all the elements of this past.

So, we will have a best possible linear combination like may be ci xn minus I equal to 1 to infinity. You see it is an infinite sum, but it is no problem it is perfectly defined, because I am taking the closure that is it not only the linear manifold of the set. But it also includes all the limit points. So, this infinite sum is defined and this must be the best this coefficient should be chosen in the best way. So, that this is the orthogonal projection that is the if you take the error between xn and this projection you can call it x prime n this part this is x prime n.

Then the error en this error, the error is orthogonal to you can say orthogonal to each element belonging to past and therefore, orthogonal to this closure of LM past. Only then only then this error is minimum in this variance. If it is not orthogonal you know from geometry if the error is not orthogonal if error was like this. Suppose, if the error were like this and this much is your prediction. Then there is off course length of this is more than the length of this. So, error variance in this other case is much more.

So, errors is minimum only when minimum in norm minimum in variance only when its orthogonal to the space on which the prediction is made. Same thing we will do here we will assume that xn is now projected on the past that is on the set LM linear manifold of past and closure of that. And the best prediction is x cap n given by summation of ci xn minus i, i equal to 1 to infinity. So, the error is minus x cap n.

(Refer Slide Time: 27:15)



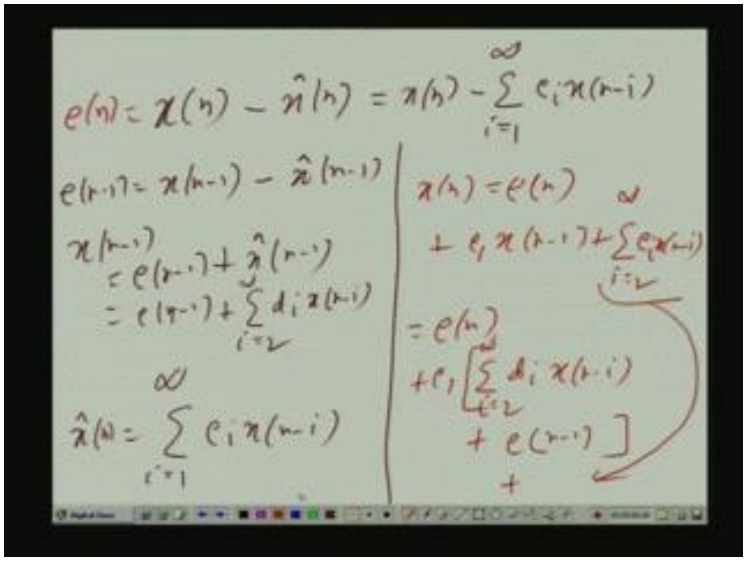
This error is en this is orthogonal LM, but this orthogonal that is uncorrelated to orthogonal to the entire past. So, it is uncorrelated that is its inner product or dot product with xn minus 1 or xn minus 2 or xn minus 3 all the inner products are 0. So, it is orthogonal. Similarly, so cap similarly I can do the same thing for x for xn minus 1. I can get that is if I predict xn minus 1 if I try to predict x minus 1 from the past that starts at xn minus 2 then xn minus 3 xn minus 4 and all that.

I get another linear combination of those elements that is my x cap n minus 1. And the errors is now you call it en minus 1 that error is again orthogonal to that error is orthogonal to what orthogonal to xn minus 2 xn minus 3 xn minus four like that. Now, remember 1 thing if we

know what is x is a first equation we know what is x cap n ci x n minus i. Here take out the first term c 1 xn minus 1 this xn minus 1 can be written as x cap sorry yeah that is right. Xn minus 1 can be written as en minus 1 plus x cap n minus 1.

Then x cap n minus 1 is again a linear combination mean be say di xn minus i, but I will start at 2 up to infinite. If you replace xn minus I by this, what you get what you get is is you get 1 term involving en minus 1 and the other terms starting at xn 2 xn minus 3 xn minus 4. Again xn minus 2 we replace by the same way you will get a term involving en minus 3 so on and so forth. So, that is I just I am just cleaning up and explaining.

(Refer Slide Time: 30:10)



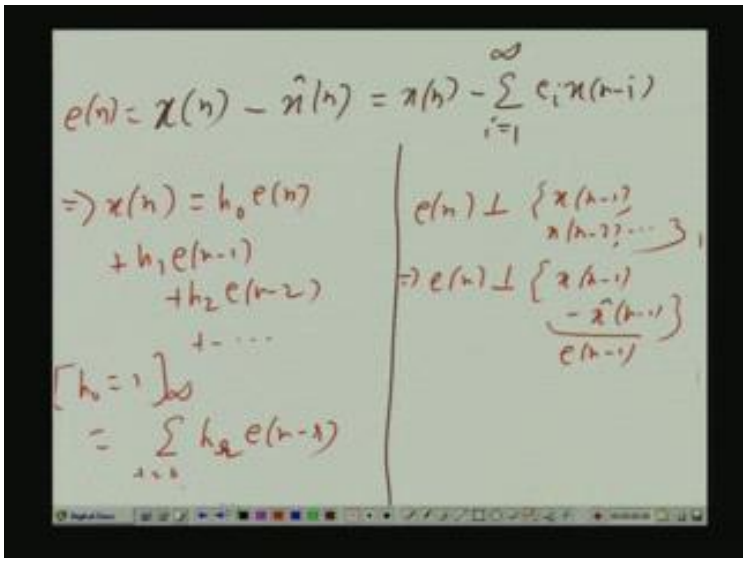
If you take this first equation xn you take this on the left side is en plus. So, I take out the first term c 1 xn minus 1 plus the rest ci xn minus I i equal to 2 to infinity and then xn minus 1 again I write as sum I equal to 2 to infinity. This is the projection may be some di xn minus i plus the error en minus 1 and plus the other term this term. You see here you will have en then some coefficient times en minus 1 and terms involving xn minus 2 xn minus 3 xn minus 4.

One more step xn minus 2 you project from project on its past. I will do the same exercise you will get again terms like en en you this right hand side will have en e 1 n en minus 1 en minus 2. And terms like xn minus 3 xn minus four xn minus five like that so on and so forth. So, essentially xn; given by what? If you do this exercise again and again I hope you are understanding this I erase this repeat sort I mean briefly what I am doing. First I wrote this

error en as xn minus x cap n x cap n was a linear combination of the past of xn out of that I pull out xn minus 1.

So, xn the other terms are kept as it is. Xn minus 1 again if you project on its past the error is en minus 1. So, xn minus 1 can be written as a summation of the error and the projection projection start at the xn minus 2. Now, if you simplify this we will get 1 term involving en another term en minus 1 and past xn minus 2 xn minus you go on doing it again and again. So, we will keep getting a I mean, I mean terms like you know en en minus 1 en minus 2 en minus 3 as a linear combination. Only en will have coefficient 1 others will have some non zero coefficients.

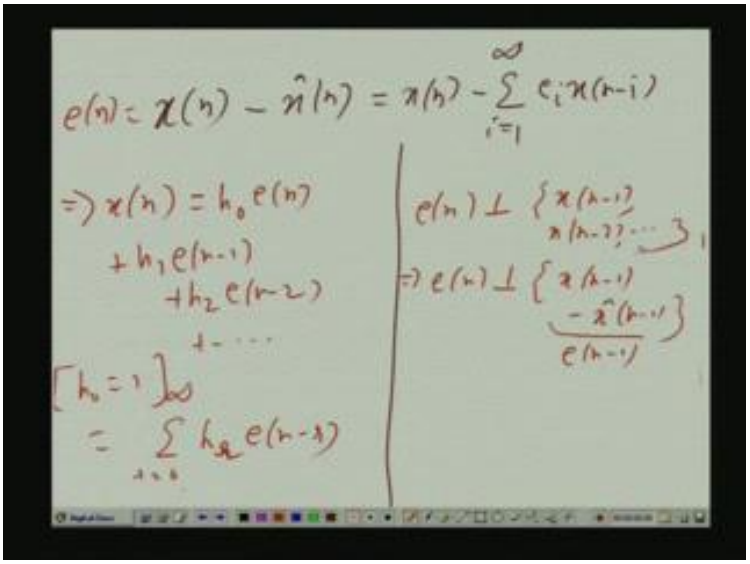
(Refer Slide Time: 32:21)



So, that means, xn can be written like some linear combination of say h 0 en. So, h 0 is actually 1, then another h 1 en minus 1 h 2 en minus 2 plus dot dot dot dot. I will of course, write h 0 is 1 as you seen here which is nothing but, a convolution between the sequences hn that is if you write hre n minus r r from 0 to infinity. As though the sequence en has been passed through a linear time causal linear time in variant system of response hr or hn.

Only thing is you see en we have seen is orthogonal that is uncorrelated to all the past of xn. Therefore, en is orthogonal to xn xn minus 1 xn minus 2 so and so. Therefore, en is also orthogonal to en minus 1, because what is en minus 1 en minus is an error between xn minus 1 and its past prediction. So, all the terms they are they are orthogonal to en therefore, en is orthogonal to the difference between xn minus 1 and x cap n minus 1.

(Refer Slide Time: 34:06)



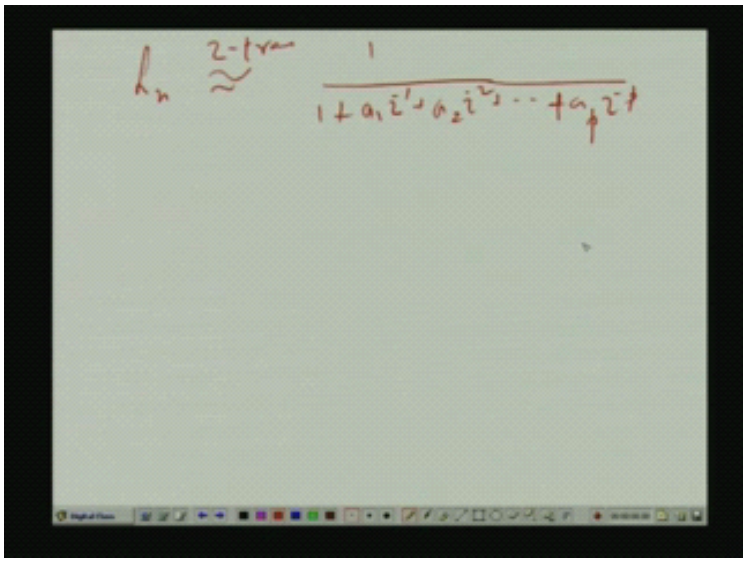
That is we can write en orthogonal to what all the past; that means, I can also write en is orthogonal to what xn minus 1 minus its prediction, because prediction comes from further past n minus 2 n minus their linear combination. So, all the terms on this in this bracket involved xn minus 1 xn minus 2 xn minus 3 like that. So, en is orthogonal to each of them. So, en is orthogonal to the error and this error is en minus 1. And you can record simply apply this en minus 1 is also orthogonal to en minus 2 en minus 2 is orthogonal to en minus 3 so on and so forth.

Orthogonal means, uncorrelated; that means, en is an uncorrelated sequence which is also called white sequence off course 0. Because, the basic data from which they are generated they are 0 mean. So, it is a 0 mean white sequence which is passed through a linear time in variant system. What kind of system causal infinity impulse response system and that generates xn that is the proof. So, any non deterministic process x of n can be given directly as I mean is always describable accurately, as though it is generated by passing a white sequence.

So, 0 mean white sequence through a causal IIR filter of impulse responses soon here by hn. This is very important this gives rise to the idea of modeling. That in that case can I now approximate that infinite impulse response causal in causal IIR filter by additional model az. Where, az is a ratio of 2 polynomial say z numerator polynomial denominator polynomial. If it is having the both numerator and denominator polynomial we call it ARMA autoregressive moving average model or equivalently pole 0 model.

Otherwise, we have got the autoregressive model that is AR model or we have the MA that is moving average model. In our case, we will be considering AR model. So, now, we have got a full justification for going for this modeling process. So, we will be assuming we will be approximating that that actual sequence hn.

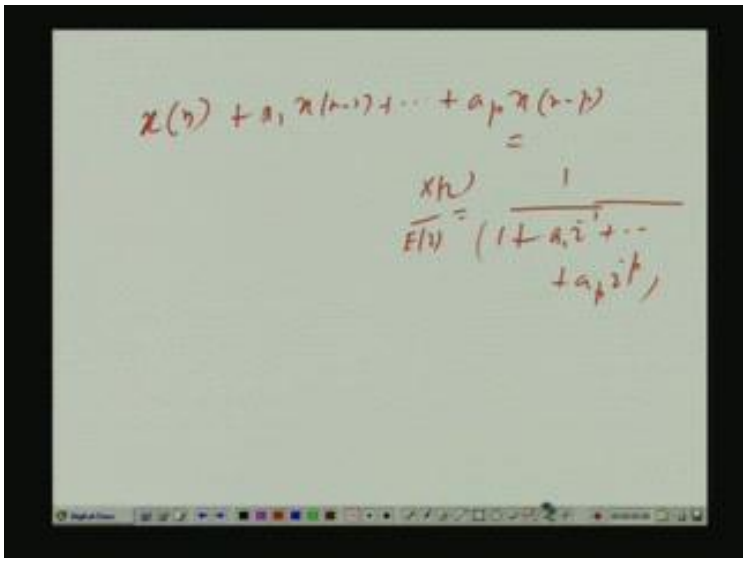
(Refer Slide Time: 36:27)



If you take it is z transform z transform is approximately equal to 1 plus this is a p th model I am taking p th order model. This coefficient is 1 and it is a causal system we should let z equal to infinite this set become 0 and any transfer of causal transfer function. If you let z equal to infinity what you get is h 0 when h 0 1. So, that value is 1. It is basic dsp for a causal transfer function hz if you let z tend to infinity what happens what is hz. Summation hnz to the power minus n starts at 0 goes up to infinity.

So, h 0 plus h 1 z inverse plus h 2 z inverse 2 plus dot dot dot dot. Here if you j allow z to go to infinity all terms disappear except for the first terms which is h zero. So, you get is the h 0. It is expression if you let z tend to infinity all this a 1 z inverse a 2 inverse 2 . And all that they disappear only 1 by 1 results and this is 1, because h 0 was 1 as we saw last in the previous slide in the previous page. So, if we can approximate this by choosing the model order p correctly. And if their approximation is clear enough or good enough then our problem is simply to estimate a 1 to a p from the given data record of xn and that is the AR modeling problem.

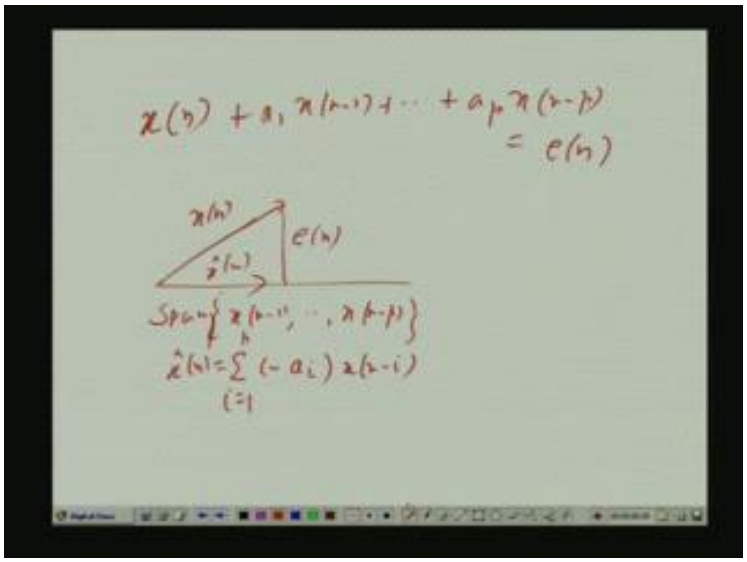
(Refer Slide Time: 38:04)



So, here what from now onwards what I will be doing, I will be assuming now that xn has been generated by passing such a wide sequence maybe en through an all pole system. So, in time domain what does that equation mean? You have something like this. You remember what is the transfer function 1 by 1 plus that is xz by Ez Ez is the z transform of the en xz is the z transform of xn. And this was dot dot dot plus ap z inverse p; that means, xz into this is equal to Ez into 1 in time domain.

It means xn plus a 1 a 1 z inverse xz means a 1 xn minus 1 plus dot dot dot ap xn minus p is equal to this 1 it is not z should be 1 you can take it a co coefficient b 0 here. But I will make it 1, because even if you put b 0 that b 0 can be absorbed into en. So, b 0 into en you call it e prime n only thing it will change is a variance of pn. So, is variance in either cases are unknown I will have these coefficients b 0 absorbing en. And you call it e prime n and why e prime again bring this a same notation en.

(Refer Slide Time: 39:26)



So, that is why I can afford to maintain a constant 1 here. So, it is simply en. And en is what? En is orthogonal to uncorrelated with all the past of xn and therefore, uncorrelated with xn minus 1 xn minus 2 up to xn minus p. Now, given the data record of the xn if I can find out a 1 to ap that is called AR modeling problem. Associated with this it is also the interesting problem of linear prediction. Linear prediction in in the case of linear prediction I do not care what whether xn has a model or not.

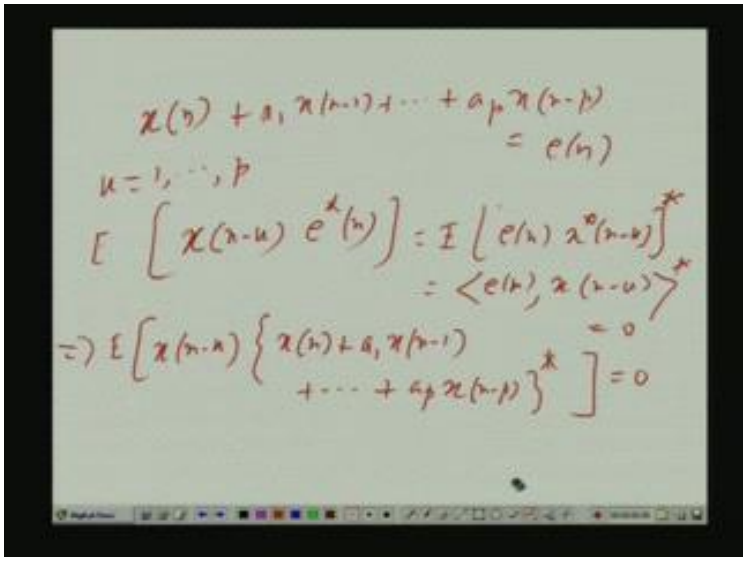
What I say is this I want to project xn into the space spanned by some finite number of past samples xn minus 1 up to say xn minus p. There is a vector space spanned by this a set of all possible linear finite linear combination after finite linear combinations, because we have got only finite terms that space. If you project it on this that projection will be a linear combination and the error is en. In that case, what is the error that is projection minus the linear combination for the projection?

So, xn minus that linear combination I write as I put a minus here just for convenience x n minus I i equal to 1 to p this is the projection this projection x cap n. So, what is en xn minus x cap n which is minus minus plus which becomes equal to this. So, in the linear prediction problem also the error en is orthogonal to to whom to xn minus 1 minus 1 up to xn minus p not the entire past. In an AR modeling problem en is orthogonal to the entire past not just to xn minus or up to xn minus p, but en terms further past. In the linear prediction problem which is a different problem from modeling en is just orthogonal to terms of from xn minus 1 up to xn minus p.

But, whether you are dealing with linear prediction problem or AR modeling problem I will be using only the orthogonality between en and this few terms xn minus 1 xn minus 2 up to xn minus p. Since, everything else is same identical model equation and all the 2 problems turn out to be same, because I will not be using in the case for AR modeling problem. Even though en is orthogonal to further past terms xn minus p minus 1 xn minus p minus 2 dot dot dot dot I will not be using this.

I will be using only the orthogonality between en and xn minus 1 xn minus 2 up to xn minus p as I will be doing in the linear prediction problem also. So, both these problems will gives rise to same set set of equation as we will see now and those equations have to be solved. We know that en is orthogonal to whom xn minus 1 to xn minus p.

(Refer Slide Time: 42:39)



So, what I do I make use of the orthogonal, I know that x I can write xn minus k to e star n. And by the way if you per permit me I will put a subscript p here, because if it is linear prediction. What is the order of the prediction? I am taking only p past terms. If it is an AR modeling problem what is the model of the what is the order of the AR model chosen p, if instead of p if it is p plus 1 and after all I do not know the exact model order.

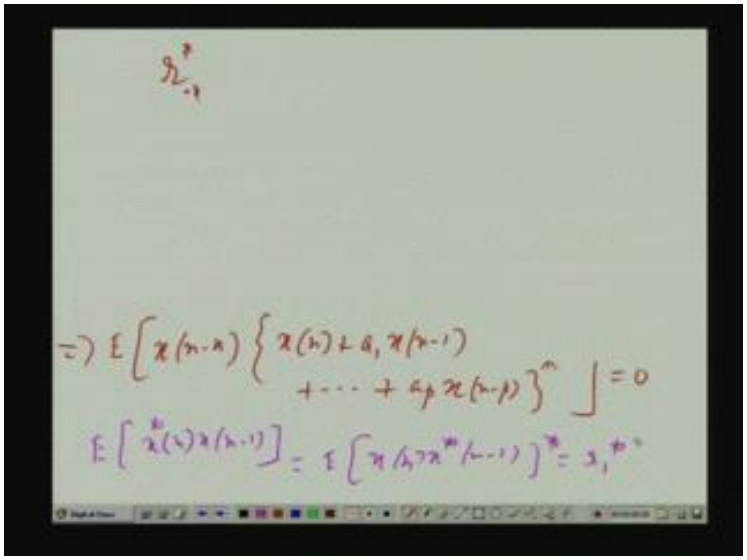
So, could have a also taken further term some more terms and I would have got another order. So, whatever equation I have the corresponding error and all that would have been different there the white sequence have been different there. So, just to indicated that a p th order model problem or p th order linear prediction has been considered I just put a subscript

p in all these. And here also or maybe I will do this little later just a minute, I will do the same thing little later when further clarity will come.

So, for the time being let it be like this and if you take this this is the inner product between xn minus k for k equal of course, 1 dot dot p. So, xn minus 1 e star n that must be 0, because that is what inner product xn minus 1 e star n means actually E of en x star n minus k star which is nothing, but inner product between en and xn minus k and then star. And this n is orthogonal to xn minus 1 xn minus 2 up to xn minus p. So, this is 0 conjugate of 0 is zero. So, this must be 0.

If I write it now if instead of en I substitute the left hand side here. What I get? E xn minus k into xn plus plus star that is, equal to 0. If you work it what we will be getting first term will be what. We started k equal to 1 say and then k equal the 2 then up to k equal p we will see some beautiful things. This is true for all k from k equal to 1 up to k equal p. So, let us start with k equal to 1 case first then k equal 2 so on and so forth.

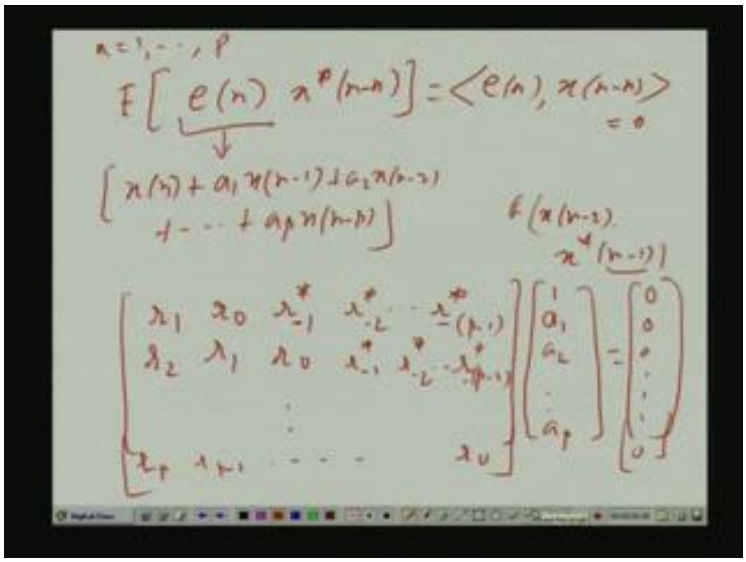
(Refer Slide Time: 45:38)



So, for k to 1 first term will be what xn minus 1 into xn conjugate expected value of that, that is is equal to what? R star I am putting the lag in the subscript r star minus 1 why. You all know this, but still for your sake I am saying. What is the first term? X star n xn minus 1 right which also we can write as we can also write as E of I made a mistake here x star n minus 1 star.

So, it is actually r not r minus 1 star r 1 star am I correct or may be if you permit me I delete the entire thing in a slightly different way that is mathematically more convenient to handle. We have because of the orthogonality, because of the orthogonality; what we have, I write the orthogonality in the other way. In fact, that is simpler way than what I did the earlier.

(Refer Slide Time: 47:31)



So, E of orthogonality means, I earlier wrote xn minus k what I wrote earlier was xn minus k e star n that was 0. But that also means that was actually bit roundabout I would I can rather take en here and x star n minus k which is inner directly product no conjugate anywhere. Then, that is equal to 0 for k equal to 1 dot dot dot p. Now, if I put here this terms xn then what I get for k equal to 1 for k equal 1 what I get simply r 1. Then for k equal to 0 for k equal to 2, for k equal 1 I am taking the first term xn then x star n minus 1.

If you take that correlation it is just clearly r 1 second time is n minus 1 xn minus 1 x star n minus 1 that is r 0 and then xn minus 2 I am just writing the correlation terms side by side first. Xn minus next 1 is a 2 xn minus 2. So, we take the xn minus 2 xn minus 2 star xn minus 2 x star n minus 2 n minus 1, because k equal to 1 case. So, that will give rise to give rise to what? X n minus 2 i I can write down, but just for your benefit I am working out those. This is permanently n minus 1 in the summation this next term I am considering a 2 xn minus 2.

So, the term is xn minus 2 xn minus 2 star xn minus 1. If you take the expected value what you get is simple correlation for the lag minus 1 minus of this right. So, what is that? R star 1. R star minus 1 if you call this index m I mean whatever index, then this minus this; n minus 2 minus within bracket n minus 1, so it is minus 1. Then r star minus 2 dot dot dot dot r star

when last term if you take this equal to what xn minus p that times x star n minus k n minus 1. So, this will be minus of p minus 1.

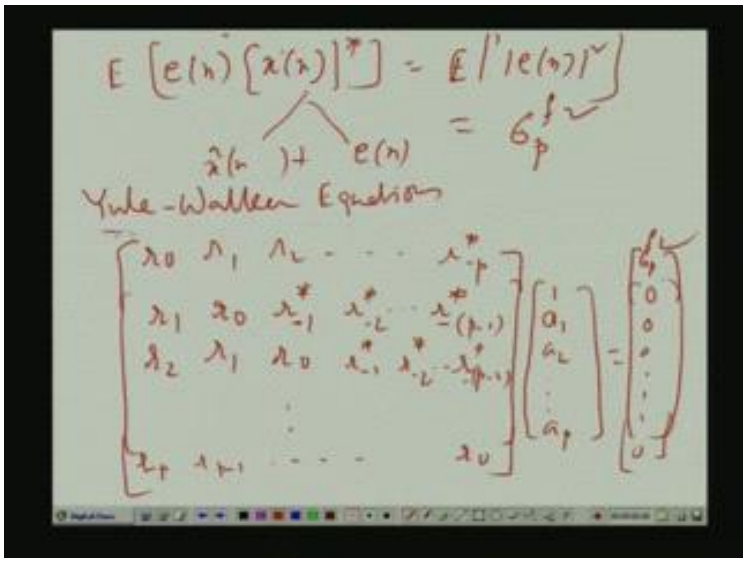
Then I put it in this kind of form 1 a 1 a 2 dot dot dot ap that is equal to this row times this column that is 0 the inner product is 0. Then take k equal to 2 case k equal to 1 and k equal to 2. What happens in k equal to 2 case? In the k equal to 2 case I have got about 5 minutes left. So, I just we will develop this equation and call off today. K equal to 2 case for k equal to 2 what do we have same thing you know I mean just instead of just we have x star n minus 2. So, this becomes r 2 r 1 xn into x star n minus 2 that gives rise to r 2.

Then next 1 is a 1 xn minus 1 x star n minus 2 that will gives rise to r 1 r what times a 1. Then r 0 and this will continue will be so on and so forth. What is the last or k equal to p case for k equal to p case that will be rp rp minus 1 dot dot dot here will be r 0 here all zeros. So, we see we get a set of equations whether you are doing AR modelling or linear prediction. You will get the same solution because; I am only using the orthogonality between en and xn minus 1 up to xn minus p which is valid in both cases.

So, equation is valid this equation is valid for both the cases. How to solve the equation actually right hand side is all 0. But, do not think that this equation is like ax equal to 0 means x equal to 0, because top coefficient here is 1. So, 1 into r 1 that you to take to the right hand side, so a 1 into r 0 a 2 into r 1 r minus 1 star dot dot dot you get 1 1 side. R 1 into 1 that goes to right hand side. Similarly, from the second equation also r 2 into 1 r 2 that goes to right hand side unknown quantities a 1 to ap they remain on the left side. You solve it.

I am just writing in this kind of form , but this equation you know I mean solving this the you know there are excellent techniques to solve this equations. And there is a famous algorithm called Levisohn Durbin algorithm this equation actually is called may be I just modified little bit. Just a few lines what we have suppose, I started at k equal to 1.

(Refer Slide Time: 53:34)



But suppose I do this en xn minus k and I also take k equal to 0 case k is a say equal to 0 that case I take. This is no longer orthogonal n is orthogonal to xn minus 1 up to xn minus p this is not orthogonal then what it is. This will be what, what is xn minus k xn minus k has got 2 components and their summation. What is the, what is one component? One component is a projection k equal to 0.

So, for the time being I can put k equal to 0 actually there is no point in keeping a k here. So, same here you do not have any k. Another is en itself projection and the error together forms x star n xn we have putting a star we have putting the star. Now, we see en out of this is orthogonal to x cap n, because x cap n belongs to the past. So, what results is e of expected value of en into e star n that is E of mod en square which is the error variance sigma we call it actually. I will I will put a notation I will explain later it is called forward prediction, because I am predicting xn from the past.

So, from past into future is called forward prediction and order p p th prediction or p th order modelling. So, sigma p f square this is just a notation it is a constant. Now, if you really replace here n by that expression xn plus a 1 xn minus 1 plus dot dot dot dot to ap xn minus pn and carry out this product. You will see; you will get this term r 0, because xn into x star n that gives rise r 0. Then r 1 r 2 dot dot dot up to r minus p star and this product will give rise to this term sigma pf square.

So, how to solve this equation first forget about the first row, because right hand side top quantity sigma f square is also unknown. Because, only thing I know is the data its

correlation values are r 0 r 1 r 2. I do not know sigma pf square that is the variance of that input noise or input white sequence. So, you take from the second row onwards solve it whatever value for a 1 up to ap you get put it here. Then take the product between first row and this column that gives you sigma pf square, these equation is called Yule Walker equation. In the next class, I will be considering first algorithm to solve this Yule Walker equations.

Thank you very much.

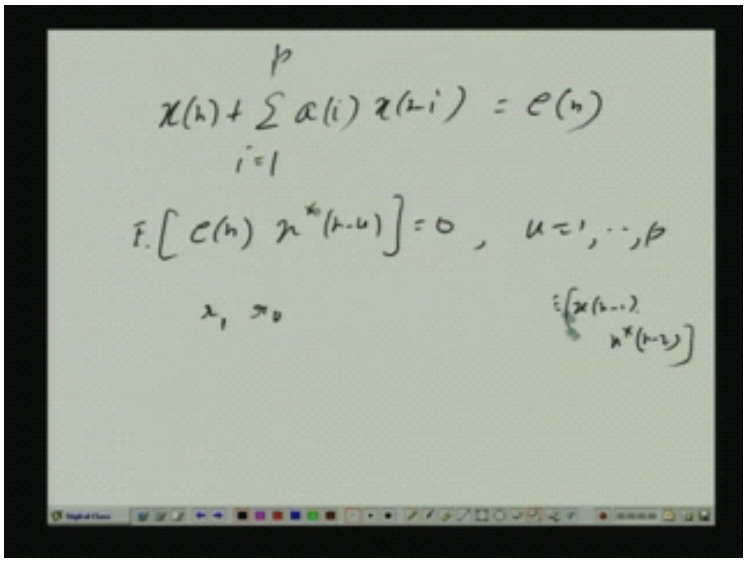
Preview of Next Lecture.

Probability and Random Variables
Prof. M. Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 38
Autoregressive Modelling and Linear Prediction

So, yesterday we are considering this autoregressive modelling there is a small mistake I made which want to correct actually. So, let me re do it again, because I do not remember exactly where it was...

(Refer Slide Time: 56:58)



So, let us redo read with we had a model like this Xn plus summation ai xn minus i, i equal to say 1 to p say p th order model you say en. In the case for autoregressive modelling en is of white sequence may be 0 mean white sequences. And it is a p th order model it it is an all

pole model if it to the z transform find out the transform functions are only poles. Alternately, we have also shown that even if there is no such model you want to just project you want to just predict xn linearly from its past p values xn minus 1 to xn minus p.

Then xn in the let the prediction coefficients be minus ai then the error will be what is what you see here on the left hand side. And that error then also you have got a similar equation. So, both linear prediction and AR modelling gives rise to same equation further. If it is linear prediction we know en is orthogonal that is uncorrelated with the past p samples. Because, what we are the doing is we are projecting x of n orthogonally on the space spanned by the past p samples xn minus 1 up to xn minus p.

The error then because it is orthogonal projection the error will be orthogonal to the total plane or total space spanned by xn minus 1 up to xn minus p. So, en is orthogonal to the past p values. Similarly, in the case of AR modelling we have see the yesterday that input en which is white sequence, because of model is causal and all that. En is orthogonal to all the past samples of xn therefore, the past p sample also. Since in the modelling problem where we will be estimating or finding out ai is unknown ai's from the given data. We will be using only the orthogonality between en and the past p samples.

It does not matter whether you are solving a linear prediction polynomial or AR modelling polynomial you get the same set of equations and same solutions. There we said that how to find out these equations, we know this orthogonality E of this is what we did yesterday and there I made a small mistake. So, I am redoing in this step, say x star n minus k that is equal to 0.