Adaptive Signal Processing

Prof. M. Chakraborty
Department of Electrical & Electronic Communication Engineering
Indian Institute of Technology, Kharagpur

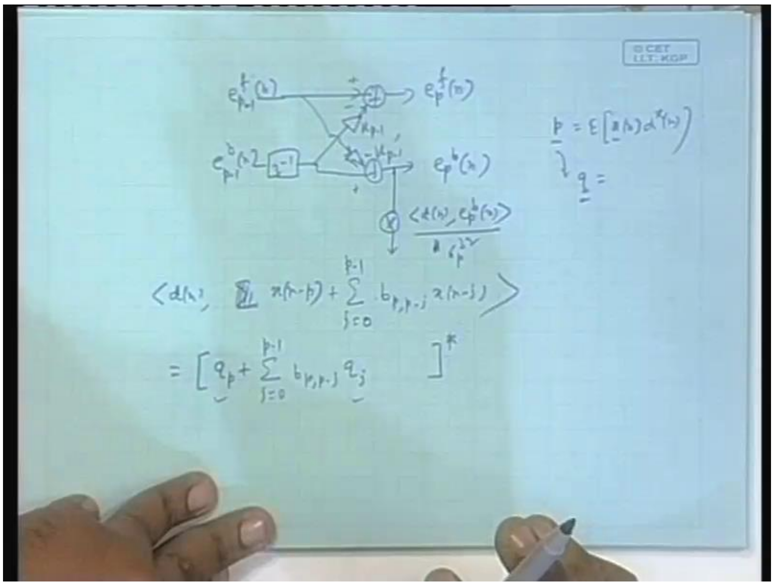
Lecture - 25

Linear Prediction & Autoregressive Modeling

You remember what is the lattice filter? One advantage of the lattice filter was that it was uniting the backward friction error to form an orthogonal basis of the single space. That is if we have x n, x n minus 1, upto dot, dot, dot, dot, x n minus p, by that backward friction errors, you get united basis for that space, for the space by the elements which is orthogonal basis. So, any external vector that is d of n, I mean, if you want to estimate it you should be able to project it along each of this backward friction errors, compute the projection as sum.

And how the advantage is that if you want to increase the order if you want to have 1 more element say, x n minus p minus 1, you simply find out the additional backward friction error component by cascading, by appending 1 more space to this; find out the new backward friction error and find out the projection along that as to the previous one; you do not have to redo the entire computation, that is the advantage, right.

(Refer Slide Time: 02:03)



So that, typically, you are, suppose, if one particular stage is this; suppose this is e p minus 1 b n; I think, this will be k p minus 1, k p minus 1 star, is it not? Because, starting at p minus 1 th order; so this component. And, multiply this by one factor; what is that factor? Inner product d n with e p b n divided by mod square e p b n that is sigma p b square. If you take this scalar put a multiplier, that multiplier tells times this. That will be the projection of d n along this backward projection error. This you have to do for all the backward projection errors. Find out the individual projections and add. Are you following me?

B n has to be projected on e 0 b n, e 1 b n, and e 2 b n, dot, dot, dot, e p b n, all the projections have to be added then we will be able to project it on the space span by x into x n minus p. And you can go on adding more stages, so the space will go on expanding; we will have higher order projections also. So, typically this is the projection coefficients. So, naturally you have to find this also, then only you can multiply e p b n by this, is it not?

You have to, but this coefficient is not difficult. This sigma p b square is already found out. It should be called sigma p square because sigma p b square and sigma e p f square they are same. And this square is nothing but this, d n; you write this, this way; e p b n means x n minus p minus, the error, an error had, it is used to be this way because of these terms had minus sign; minus and minus plus; j was from 0 to p minus 1, I think so.

J equal to 0 x n, j equal to p minus 1 x n minus p plus 1; b b p, b p 1, 2, of course; this is this; and you, this correlation; correlation will be what? Instead of, earlier we have given in p vector as the cross correlation vector between, what is that the real case or not even real say, what is this case? Cross correlation vector between this and this; x n vector into this star n. But I have already used the notation p for indicating the order. So, instead of p vector you give a, define a name to it; you call it q vector.

We have already used of this symbol p once. So, there is no idea; there may be confusion. So, what is this coefficient? D n with x n minus p, that will be actually q star because the way definition is x n and d star n, that is q; and here is your d n with x n minus p. So, actually, we can put a thing and then put a star here; this will be q p. This will, what? Q j; q j, x n minus j; x n minus j is component here, times this star n; this is a opposite here.

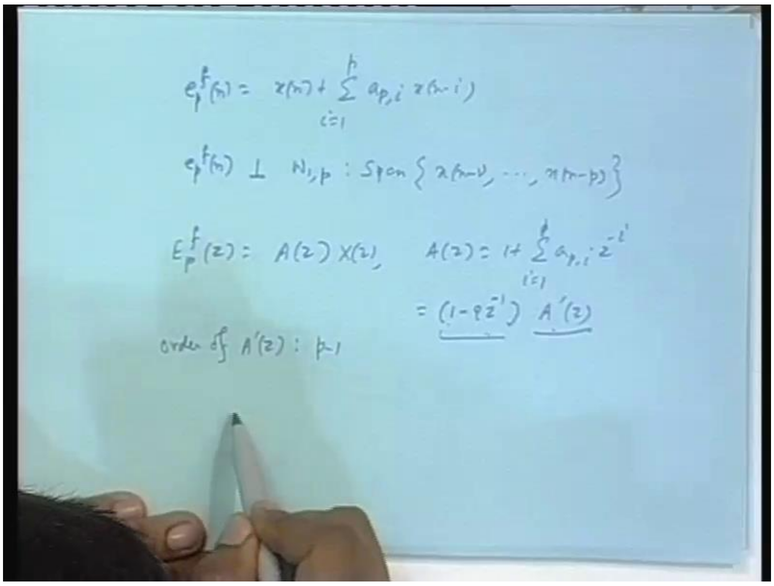
So, q j and star; if you know these terms; and this coefficients, coefficients you have already evaluated order recursively. So, we can find out this. Again, this is not for online computation, this is only for constructing the lattice. Once the lattice is constructed, this multiplies the stationary and fixed; then you go on filtering x n, and you get all the backward friction error, power friction error, and the projection of d n, and all those things.

This is only off line computation; that is to compute this various parameters recursively; or even directly, here it is direct. Because once you find out this coefficients b p comma anything recursively, then put them back here, carry out this overall correlation terms, summation of this correlation that will define this inner product. This is just the formula; this is how to get these coefficients because you must know these coefficients, then you can carry out the projections.

We will now see one other nice property. I think, this property can be proved in various ways. But, let me first consider it, prove it; discuss it, prove it, and then I will tell something more about it. Suppose, I am doing linear prediction; there is a process, random signal x n which is w s s, 0 means say, and w s s, and there is no really linear relation involving the samples.

So, no sample is a linear combination stuff the remaining ones, either finite ones or infinite ones, which means any projections, if you take any x n projected on the space span by any other set of the samples, the projection error is always having, I mean, nonzero variance, there is nonzero mod square. This is the one of the case we are considering. They are, suppose, I am considering a p th order prediction.

(Refer Slide Time: 08:29)



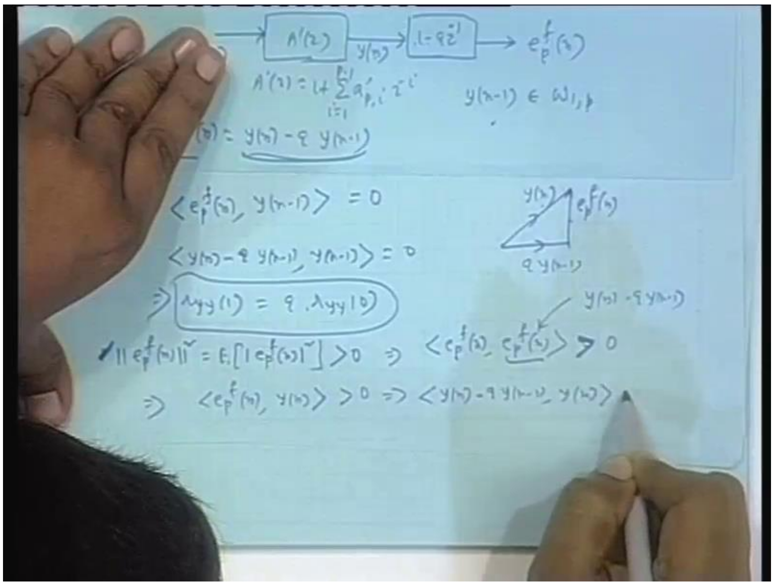
So, p th order prediction means the prediction error, forward prediction; prediction error according to us will be; because this is already we have made negative sign here, in the coefficients; so minus and minus plus; and e p f n is orthogonal to what? W 1, p. What is w 1, p that is span of, is it not? I am projecting the x n orthogonally, on space span by the x n minus 1 to x n minus p, I am taking the error, that derives orthogonal to this space; this space is w 1 p; so this is this.

Now, consider z th form on both sides. So, what you get here is as A z into X z square; A z is; so A z is a polynomial in z, of what order, p th order, is it not? So, I can factorize into p faster of factor, some factor might repeat, but p faster of factors. The roots mean complex, in general; but I can by the fundamental theory of algebra as long as this coefficients are more complex fields I can factorize it as well as a coefficients of complex, or I mean, complex numbers are allowed, I can factorize it into p faster of factors.

So, consider one factor at a time say, q z inverse into remaining one; remaining one is A prime Z. You see, factors I can always write in this form, 1 plus or minus something into z inverse; otherwise, if you want to do like a school boy, you can put everything as a power of z, I mean, z minus a, z minus b, z minus c, again you can write into this kind of form, what is the point, you can write it this way.

Because after all the whole thing is positively z inverse. So, 1 minus q z inverse into A prime z. So, A prime z is, its degree is, its order is p minus 1. One factorize has been taken out. Why do you call it z inverse? I called it A z; I called it j A z, A prime z also will be a polynomial in z inverse, is it not? While this is a polynomial in z inverse, I took 1 factor out; this is also a polynomial z inverse, leaving coefficient to be 1 like this, and it is, the order will be p minus 1.

(Refer Slide Time: 12:04)



So, this factorization of A z into this form, it means what, I can cascade it like this. Call it y n; and this is your e p f n, it cascade. The y n is a linear combination of what? This is a polynomial; you can call this polynomial A prime z; if you want I can write it explicitly; A prime z is something you know, 1 plus some A prime p comma i z inverse i, i equals to 1 to p minus 1, right, some new set of coefficients.

So, that means, y n is what, a linear combination of x n, x n minus 1, upto x n minus p plus 1; x n that is past p minus 1 number of samples because orders is less now; it is not p th order; it is p minus 1 th order, right. So, instead of y n if I consider y n minus 1, then y n minus 1 is a linear combination of x n minus 1 upto x n minus p; that means, y n minus 1 belongs to w 1 p; y n was a linear combination of x n to x n minus p plus 1; that is y n belongs to w 0, p minus 1.

So, y n minus 1, it belongs to w 1, p because y n minus 1 means it is a linear combination of x n minus 1, x n minus 2, dot, dot, dot, upto x n minus p. We are starting at x minus 1, is it not? So, that means, this is true. So, that means, e p f n which is orthogonal to w 1, p; e p f n is after all orthogonal; e p f n is what? 1 x n projected on these and the error. So, e p f n is orthogonal to y n minus 1 also; e p f n this orthogonal to y n minus 1 because y minus 1 belongs to w 1, p and it is a function of x n minus 1 to x n minus p.

And, this guy is orthogonal to each of them; b into x n minus 1, b into x n minus 2, b into x n minus p; that means, what? Now, e p f n, the beauty is I have found I mean y n is the, y n I made it y n minus 1; immediately y n minus 1 was member of these; not y n; y n is not a member of this; y n minus 1 is a member of this. And, this e p f n is orthogonal to w 1, p; not w 0, p minus 1; e p f n is orthogonal to w 1, p. So, these 2 I can connect, and immediately I can say e p f n is orthogonal to y n, sorry, y n minus 1.

But at the same time, e p f n is what, y n passes through filter like this. So, what is e p f n? Y n minus q into y n minus 1; simply, if a filter first order, y n minus q into y n minus 1. That is, simple b s p says, e p f n is what, y n minus q into y n minus 1 that is for b s p. So, this inner product which is 0, I replace e p f n by these expression. This e p f n is this. Geometrically, what is happening you know?

E p f n is orthogonal to this guy, y n minus 1. This is q into y n minus 1; together is y n; y n is summation of the 2; e p f n is orthogonal to y n minus 1; and therefore, it is say along y n minus 1; this q into y 1 minus is if you have added together is y n. So, e p f n is orthogonal to this, but e p f n is y n minus this. I have to finally, you know, write everything in terms of y n, rather than e p f n that is what direction of the movement is.

So, e p f n, you write as, this means what, y n with y n minus 1, that is r y y 1; this implies what? Minus q into r y y 0, that is 0, that is this is equal to r y y 0; this is one relation that comes immediately from the orthogonality of this error with y n minus 1, from this relationship immediately. So, this is orthogonal to this and summation of the 2 is this, immediately you get this.

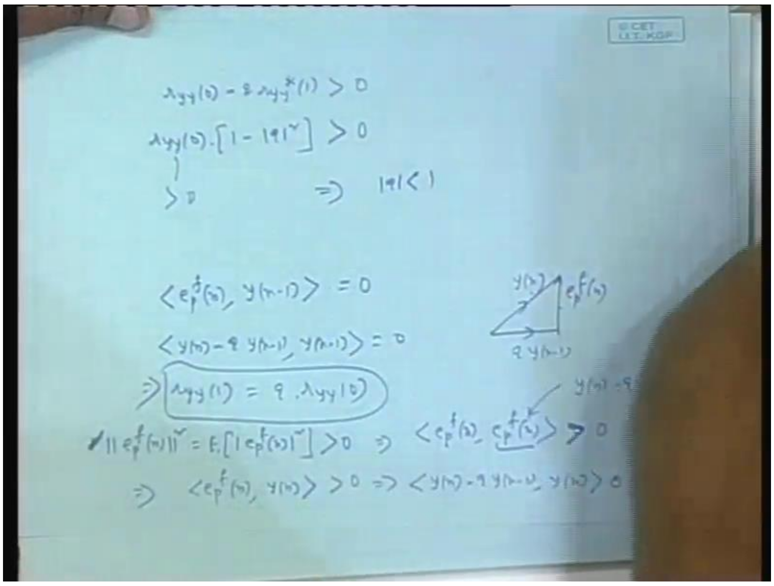
Now, non square of the e p f m, as I told you no sample x n is a linear combination of the, of its past or future samples, that is why e p f n is what, it is a projection error whose non square and variance has to be non 0, has to be positive; it cannot be 0. Our assumptions is this; this is a full gang process that is no sample is a linear combination of its finite number of past or infinite number of past, or finite number of future, infinite number of future terms.

There is no linear relation linking the elements; that one is a linear combination of the others. In such a case, and in this particular case, that means, x n is not linearly related to, x n is not lying in the space this, that is why x n projected on this as the error which is this, that is a non 0, that is positive non square. So, this which is equal to variances, it is strictly greater than 0, this we know; but this non square means, e p f n with itself 0. What happened?

Student: Greater than 0.

Sorry, greater than 0, yeah, thank you; greater than 0. And, now there are various ways you can go about. E p f n, you can write as y n minus this; this you can write as y n minus this, q y n minus 1; e p f n is orthogonal to y n minus 1, is it not? That is these dot property itself means is what; this that is how you see; this is nothing but this plus this; this implies that is, this e p f n you can always write as y n minus q y n minus 1. And e p f n and q y n minus 1 that we are getting 0; this fellow, this fellow is 0; so e p f n with y n that is greater than 0. Now, replace e p f n by this expression; e p f n by this expression, y n minus q y n minus 1, y n greater than 0.

(Refer Slide Time: 20:50)



This means, now you expand this; y n with itself minus, q what? Y n minus is y n that is r y y star 1; r y y is because y n minus 1 coming in the first coordinate, y n going in the second coordinate. Correlation would have been, I mean, r y y 1 is inner product within y n and y n minus 1; here it is the opposite; so conjugate. So, this is, q comes out, q times r y y star 1, this is greater than 0, right; but earlier I have found a relation between r y y 1 and r y y 0, it is the orthogonality.

So, I replace r y y 1 by this; that means, r y y star 1 means q star into r y y 0; r y y 0 is always real. Star of r y y 1 is star of the entire thing is product, but only q star because r y y 0 is the variance, if it is never, complex; r y y 0 is the variance, non square, inner property itself, it is never complex; this is the step. So, this you understand. So, I replace it. So, r y y 0 remains as it is. So, you get this term, 1 minus q into q star.

Now, what does it mean? R y y 0; what is y n? Y n is here. Now, again y n is what? X n plus a linear combination of x n minus 1 upto x n minus p plus 1; y n can never be a 0 random variable; that is, this is not that, that would be an x n is a linear combination of plus p minus 1 samples, is it not? Y n variance, this according to me it is not only real, it is always greater than equal to 0, but according to me this is greater than 0.

Because if it is equal to 0 that means, it is a 0 random variable; it always take 0 value which means x n is a linear co x n plus this; x n and the linear combination of x n minus 1 to x n minus p plus 1 that is 0; that is x n is lying within the space span by x n minus 1 to x n minus p plus 1, that is again contradiction. I am assuming our sample, no sample is a linear combination of other samples; no sample belongs to a space span by its past or future samples; that means, y n can be what would be a 0 random variable; variance of y n has to be positive.

So, that means, that this implies, mod q is less than 1, mod q square is less than 1; so mod q, mod as the positive number, mod q less than 1. What is the implication of this, that whenever you factorize this linear predictor polynomial, every root which is the 0, every 0 lies within unit circle. So, remember lattice has so many stages. At each stage we are computing say p th order, prediction p th order of poly polynomial; next stage, p plus 1 th order for a prediction polynomial.

Each of these polynomials has 0s lying within unit circle. A polynomial with 0s lying within the unit circle is called minimum phase polynomial, minimum phase. So, understand it; that is a cascade of stages, but each stage gives us to a minimum phase predictor polynomial, power predictor polynomial.

Because, what are, in any stage what I am doing, I am finding out the p th order. So, any stage is a p th stage; what I am doing? I am finding out the p th order forward prediction and corresponding polynomial, this coefficients; A p, i 1, 2, 3, upto p, those coefficients they give base to the forward prediction polynomial in z domain, and it is for any order p.

So, for every order I am finding out the linear predictor, prediction, and therefore, the predictor polynomial. But, by this theory, every predictor polynomial is minimum phase that is as it 0s within unit circle. So, if I master the system by taking the reciprocal of this linear predictor polynomial that will be causal and stable. These are 0s, suppose I take that A z and make a system 1 by A z. So, 0s become poles for further system poles like within unit circle. So, that means, there will be a causal and stable system, that will be guaranteed.

How?

Student: voice not clear.

You cannot.

Student: voice not clear.

No, but there are other proof. Now, I tell you one thing. There are various proofs given to this property. When I studied this when I was a student that time this proof did not come up. Actually at that time we used that property, you know, that each reflection coefficient has magnitude less than 1, using that and there is some theorem from complex analysis called roucet theorem.

There we found out the roots of, assuming that polynomials upto p th order are minimum phase, and then find out the polynomial for p plus 1 th order by the recursive formula we used; there you show that, that also is minimum. So, that is how it was done. Okay, but they had used some theorem of which whose proof I do not know, but that is called roucet theorem.

But, 4, 5 years back, I was just seeing one journal paper, I had just glanced through journal. So, you, I do not know whether you know of this guy, P. P. Vaidyanathan. Now, P P, we actually, to us he is known as P P. P P actually is not a person of this area. P P is a very famous person. Actually, he comes from Calcutta. He studied in Calcutta University and then moved to moved to Cal gate, moved to Santhovada, did PhD there and he is a celebrate faculty in Cal gate; very, very well known.

But, he worked primarily in filter lattice. He is a person of, key person of that area; he was considered founder of all research of that area. But, I do not know how in what context he was teaching this thing. He wrote in that; he published 1 short papers in, later which caught my attention actually, it was 1999 or something like that; that was the end.

In the end he wrote; actually, I never thought that I would have published this and I just derived it, but someone who is a father figure in linear prediction, his name is John Makhoul, you know, John Makhoul Makhoul. John has suggested that he must publish this result, so we published. And, I wrote that proof and I it got my fancy; I thank you, it is beautiful.

So, then I, when I teach, I consider the linear prediction orthogonality and all those things, and I prove. But, there is no simple proof. If you get come across one, I will let you, I mean, let me know, but this that is no simple proof like that, fine. So, linear prediction polynomials are minimum phase polynomials.

Now, we will do something related to this, not prediction, but something which is related to this. Again, there is a lot of theory in this; unfortunately this is not a course on; I will not go into it, but just hear me out. If you consider random processes say, 0 means random processes; stationary or not even stationary; then this kind of processes can always be decomposed into 2 things; that is, a summation of 2 different random processes.

One is called purely deterministic component, another is called purely nondeterministic. Purely deterministic, deterministic but still random, mind you. So, purely deterministic is one where any sample can be predicted exactly without any error from its past, either finite number of past samples or infinite number of past samples, but absolutely predictable.

So, deterministic random process is actually the predictable component. These components, you know, unlike this say, consider a sinusoidal random process, sin omega t plus phi; phi is random. So, every time you measure it, phi changes; and therefore, away from changes it is random, but sinusoidal nature remains. So, if you know it is valued over one period, any future period values can be predicted without any error because of the sinusoidal nature. This is a random process which is predictable.

This kind of process if you see its power spectral density, you know, it will be line because sinusoidal things if it in their Fourier transformer then they become impulses. So, similarly, here also if you take a sin omega t plus phi, phi needs auto collation function. It will, I mean, take, assume it is a stationery; actually, it will be stationery if you assume phi to be uniformly distributed about 0 to 2 phi which is always met.

Another one, always made in control communication signal processing, the random phase is uniformly distributed from 0 to 2 phi, then we can show that sin omega t plus phi into sin omega t minus star plus phi, if you take the product, take the expectation with respect to the phi that is multiplied by 1 by 2 phi, 0 to 2 phi, all those things, you know; then it will be some sinusoidal components of that gap tau.

See, if you take its, so it is basically function of tau lack; so this is stationery. If you take Fourier transform, it will give, obviously, the impulse; impulse means line. So, this kind of processes are called as processes which are having line spectra. So, that predictable component actually can be shown to be either a single line or is a combination of several lines. So, the spectra they constitute, they give us only to lines.

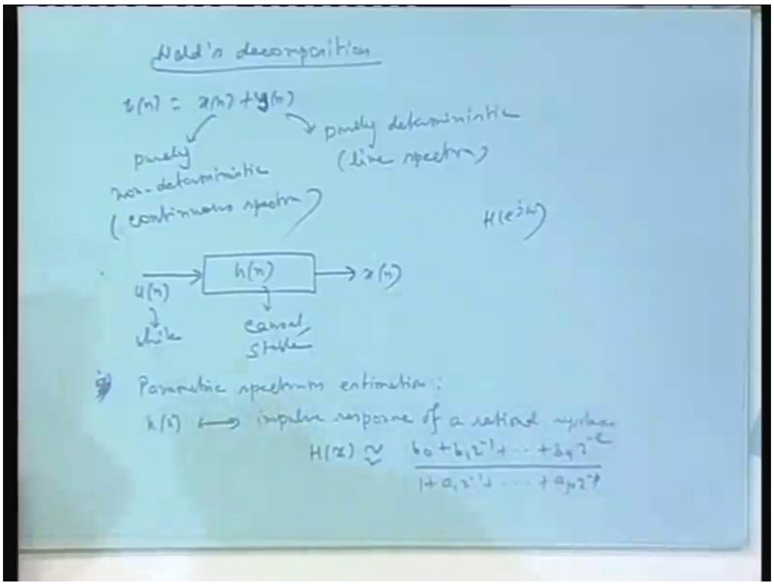
It may not be a single sinusoidal. It may be predictable which can be recomposed into some sinusoidal as, I mean, random sinusoidal processes is giving the extra line in the spectrum while in the stationery; unless stationery, there is no point in calling spectrum because spectrum, power spectrum or psg works are defined only when the process is stationery, you all know; auto collation function of only one lack variable. So, that is the line constant.

But, the other component which is called the purely nondeterministic component that interests us more because that is a kind of random processes we come across in life most often. Other one also come, we come across the processes, but they are synthesized by all; you see radar and all way you know originate sinusoidal pulses or some predetermined pulses, only this phase is random or amplitude is random, and when it comes back by reflected, by reflects of from an object and when it comes back amplitude changes, phase changes.

So, there the structure is controlled by the sinusoidal structure, the sinusoidal nature processes. So, in those applications these things come up. This predictable process is property and all that, but otherwise in communication specialty we come across random processes, noise, all those things, they are the there is nothing predictable. So, there is a nondeterministic random process. In general, any random process is 2 components - one is predictable part, another is purely nondeterministic part or non predictable part.

In the non predictable part, it is said, in such a one where any current sample x (n) cannot be predicted exactly without any error from its entire past. If you want to predict, the prediction error will have a non 0 that is finite, a positive variance. This is called, this thing, nondeterministic or unpredictable component. Now, in our, in this course where the unit purely nondeterministic processes.

Refer Slide Time: 32:40)



This decomposition actually is called Wold decomposition; please read this; Wold's decomposition. That is, given any random process say, v n, we will write it as x n plus say u n; this is purely deterministic. This is giving us to line spectra, purely to do have a spectra, general kind of spectra known continuous spectra. Line is discrete because it is a line spectra or discrete spectra, discrete all has to be line only; this is a continuous spectra.

We will be considering only x n. Further, this is not the end of the story; x n can be written as what, as generated by, a processes h n is generated by some, actually instead of u n let me call it y n, x n plus y n; u n I want to use here; it is generated by passing of white noise, white; white sequence through a system with this causal stable. The causal stable is linear, I mean, the system for the process h n;; u n passed through this, to give raise to x n.

This h n, so that means, if you know, h e to the power j omega, you know the power spectral density of x initially. It is some constant; it is a variance of these times, mod h e to the power j omega square. This is the starting point of one very, very powerful method of spectrum estimation called parametric method of method of modeling. There is, we assume the system to be; you know the system is purely nondeterministic; we can model as though it is generated by passing out white sequence of some variance through an lti system that is causal and stable.

So, then estimation of the power spectrum density is amounts to estimating, h e to the power j omegas, by what, this system; because mod square of that times is constant; that constant will be the variance of this; that can be separately estimated or even if you do not know that it is just a scaling parameter which will not change the shape of psg. Of course, the values might change, but the shape will not change, right.

Basically, it amounts to identification of the system; it becomes a system identification program. Then, in parametric estimation, what we do? We have to then introduce some approximation. In one approximation you see that, h n, is an impulse process of a model of a rational system; that is, parametric spectrum estimation; h n, you assume to be impulse response of rational system.

I do not know whether you know rational system; we have all studied in psp. It is called constant coefficient definite equation will be 0 initial condition. I do not know how it has been taught in psp, but in constant coefficient definite equation if you do not have 0 initial condition it does not become linear and time invariant processes. You have to give 0 condition; some other time I will discuss off line and not while in this course.

Rational system means, H z, is approximated as, approximated that is modeled as this kind of form, because this is general h n, upto this one is accurate because Wold gave us this; this is always true. But, Wold never said that this h n actually, if you take the j transformer of h n you will get a close form like this; this is something we are introducing. We are saying that by taking q and p, the 2 orders to be very large, will approximate actually Hz by something like this.

This model, if you take the model as it is, both with numerator, denominator, this got 0s also, poles also; this will be called the pole 0 model. Pole 0 model is also called; in that case if you have a model like this, then we will simply need to find out; here I have to find out all the infinite coefficients say, 0, h 1, h 2, h 3, dot, dot, dot, upto h infinity, then only I can find out h e j w is equal to 0. But, here the advantage is, you need to find out only this coefficients b 0, b 1, upto bq, and a 1 to a p.

If you sum up the estimate of this coefficient, transformer function is known. So, entire infinitely long impulse response is basically captured by a few finite number of coefficients. So, your entire purpose of this transfer of to be for giving data record of the process x n, sum of estimate by some technique, estimate b 0, b 1, b q, and a 1 to a p. If you can estimate them, then find out h e to the power omega mod square of that times, some scaling constant will be the psg; that is why it is called parametric power spectral estimation; this parameters need to be estimated.

So, entire thing transformed to be what? Find out better and still better thing to by which you can estimate b 0 to b q and a 1 to a p. Entire power spectral density estimation amounts to that plus, the input variance to be more to be actually accurate; you understand? Here, if you assume the models to be just like this, both 0s and poles by z, this is called a pole 0 model; pole 0 model also called autoregressive moving average, a r m a, autoregressive moving average model.

That is the, difference equation will be what? If it is y get will be what? Output is x n, input is u n. So, this will be x n output plus, this will be u, sorry; input u, output x; mind you, this kind of things, just writing this is not enough; we have to have 0 initial quantity that comes from psp that we will not discuss here. Please understand in the case of differential equation in continuous type system you have to ensure the initial condition then only it becomes linear and timing variant, then only output by input ratio turns out to be h of s which is the h t.

In this case 0 initial conditions we have to assume; that is when the system is switched on, all x n, its past values x n minus 1 stand x n minus, they all 0, then only it will be get separately it is lti. In that case, y x z by u z, x z by u z is always of this form whether you have linear 0 initial condition or not, but only if a 0 initial condition x z by u z will turn to be x z.

Because, under 0 initial condition, this lti; for lti system, output is nothing but convolution between the 2; in z domain, product from the 2 z transformer. So, output z transform by input z transform, turns out to be h z. So, only, because I am telling you again, even when there is no initial condition, x z by u z will be these, but x z by u z will not then be h z; it is because it is not a linear timing variance system. So, this is not a convolution; output is not a convolution between the 2. So, x z is not u z into x z, that work separately.

But, only when if you ensure 0 initial conditions then I can prove separately, irrespective of all these, that is in lti system. For lti system, output is a linear convolution between these 2, which means z transformed y is output get transform by, input z transform, will be s z. So, output z transform by input z transform leads to this kind of form, will it be indicating the s z for the system.

This is a separate story if you have not studied in esp, I am sure you have not studied, it is not given in book of course, we will not teach it again very well foresee. It is not linear and not timing variant, what it has to, unless you ensure that even if you take this kind of form, x z.

Yeah, otherwise not; otherwise this modeling collapses, anyway.

Pardon me.

Oh b 2 a into v n, sorry; this is b 2 a into v n, thank you.

So, this is called, this side is called the autoregressive part; the regression on itself x n; and the regression, you know, x n minus 1, x n minus 2, regression, autoregress on itself, that is the autoregressive. This is moving average because this is after all some kind of averaging, linear combination of e b m to n minus n minus q some kind of it, but average again changes from, because right now was, living coefficient is living terms is u n.

You are averaging, this term u n upto n minus 1, n u n minus q, but next time it could be n plus 1, then u n, this is the average is moving with time, changing with time, that is why you call moving average part, auto regression moving average. Sometimes, you can just take the model to be only this, as the denominator is only 1, in a special case of only moving average part and that is called ava model, moving average model.

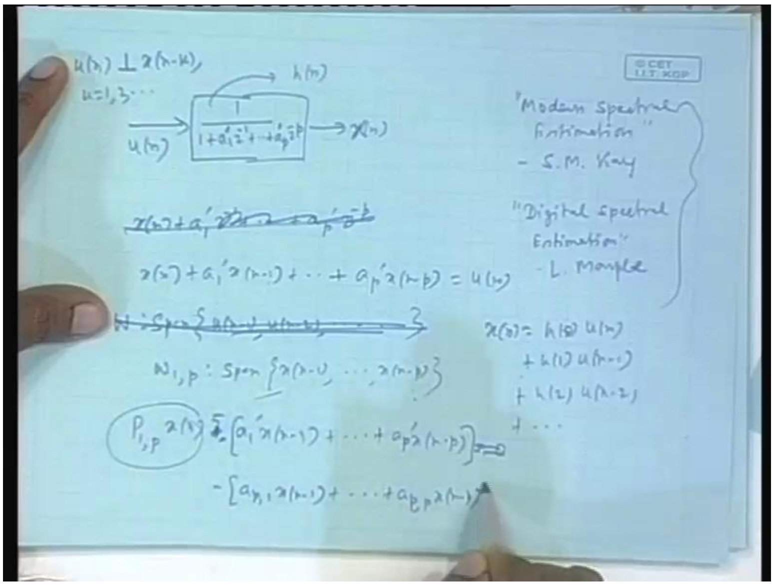
In that case, if it is only 1 x n output, this part is not projected, x n is just this. Remember, in both augma and na you solve, basically to estimate at b 0 to b q you are basically looking at a non-linear program. This is because input, you are only giving the data x n, you are only modeling it like this; you do not know which random sequence generate it. So, u n is unknown to you, but, so what this coefficients b 0 to b q, and here you have product of 2 unknowns.

So, whether it is av model or alva model, this is basically a non-linear estimation procedure that makes like l. But, when it is this part, that suppose you take only this, only poles, numerator may be a coefficients b 0, in that case you have got these term; it is a all pole model; it is a linear estimation problem because of the left hand side parameters are unknown, but x n is known to you.

This is the model which is a celebrated model which has been studied most extensively it is called autoregressive model, ar model. This b 0 can be observed in u n also, because if u n is wide, so is b 0; b 0 into un; but, we need to find out the variance of this overall process, b 0 into u n, because output power spectral density will be not only mod square of these, but also that times input variance, input is wide, it is variant. So, b 0 u into un, if you call u prime n, I must know the variance of u prime n. So, this auto regressive model we have considered.

This b 0 u n, I will be calling you u prime n; or, why in u prime n, if you permit, I will call it again by u n only. First you call it u prime n; b 0 u n you call u prime n, and u prime n you call as u n again, no problem.

(Refer Slide Time: 45:45)



So, that means, what you have got, please see that there are 2 books on this; they are the fantastic books, easy to read, but celebrated books on spectral estimation by the professor and experienced student, teacher and a student, each wrote separately. One is digital, one is called modern spectral estimation by the professor, a celebrated person and a wonderful man also, Stephen K, S M K, modern; they are in the library. You people worry doing research and all that who have this kind of things in mind, buy this book, atleast must have a look at this book.

Modern Spectral Estimation by someone who is a great researcher in this area and also a great person you know, so known to be a great person, S M K; and his PhD student also became very famous, his name is Larry Marpel; he wrote a book simultaneously. So, the books are more or less overlapping; each is great, great book by its own, but books have to be even defined names, you see. So, then Marple gave this name Digital Spectral Estimation.

Both; I personally I am more familiar with this casebook by Larry Marple, both are books; both are there in the library, and not difficult to read; it does not assume any extra knowledge and anything; very reader, reader friendly kind of books.

No, only normal to my knowledge; there will be 2 copies, but 1 volume, 2 copies.

Student: Marple, Marple.

So, this is our problem. We are finding out, this is the system 1 plus A 1, sorry, x n. So, in time domain this is what, and u n is wide; and, I am assuming the model to be most important; I am assuming the model to be causal and stable.

Now, see, one thing, before I proceed further, suppose I am interested in random process x n, after this random process exact samples are not any consequence to this, power spectral gets you that random, if I can generate similar random process; that means, something which has a similar statistical structure statistical thing that will be enough for me.

Suppose, x n implies what, iit, suppose. Every time you what are iit, you will get different x n's, but they all have some statistics, they all belong to the same statistical class, then there is, they are say, power spectra and all, they will be same; they belong to a statistic. Now, suppose instead of sending x n, what I do, I find out these parameters A 1 to A p and the input variance, that is enough for me to generate the random process; exact samples are not enough.

I should generate that those that random sequence whose psd is of that type. And then if I play it in a speaker it will generate the, what iit, very loosely speaking. So, that is actually, this is for all these linear predictive coding and all that, you know. And, instead of transmitting the entire block of x n, if it is a, it is relatively accurate, you can model like this by an AR model; you better estimate A 1 to A p and input variance and send them, that is all.

So, there is p plus 1 number of parameters. You do not need to sample all, send all the samples. Anyway, let us find out these equations. For the time being let me permit to, please permit me to let prime b here. Sorry, this will be a problem. And, this has got impulse response h n which is casual and stable, please see; I am assuming because I will be dreiving this model; and, this model must be casual and stable, otherwise there is no point.

Then, I estimated the model and parameter standard to be such the poles lie outside unit cicrle, and therefore, this is a casual and it is not a casual and stable system, then this is no good, is it not? System must be casual and stable, then only our all those formula remains valuable that output power spectral density is mod, this what l if the system is not working; system is nonprocess, and so it will not work. It has to be a real life system, causal and stable.

Now, if it is casual and stable, that means what? This is the model equation, but parallelly x n is what, h 0 into u n plus, h 1 u n minus 1, h 2 u n minus 2 plus, dot, dot, dot; causal, that is why it is starting at h 0 times, current 1, then h 1 times u n minus 1, h 2 times u n minus 2, then can I say that u n is orthogonal to x n minus k, k is 1, 2, dot, dot, dot. So, it starts with k equal to 1, x n minus 1. What is x n minus 1? It will depend on u n minus 1, u n minus 2, u n minus 3, dot, dot, dot; u n is white.

So, u n, correlation between u n to n minus 1 or minus 2 or minus 3 or minus 4, they are all 0; so that means, u n is orthogoanl to x n minus 1, because after all it is x n minus 1, here it will depend on u n minus 1, minus 2, minus 3, like that. And, current one is u n. So, u n is orthogonal to each of these terms – u n minus 1, u n minus 2, u n minus 3; that means, u n is orthogonal to x n minus 1, by the same logic u n is orthogonal to x n minus 2, x n minus 3, etcertra, etcertra.

That means, if I project u n on the space span by all the past sign, u n minus 1, u n minus 2, dot, dot, dot, dot, space span by them project u n on that, that projection will be 0; u n is already orthogaonal; it is orthgonal; if it is already orthogonal, projection is 0; and, it itself is the projection error. If suppose, somebody has projected, that somebody has already orthogonal that projection error is 0, is it not?

Now, suppose, in this model, on the left hand side and right hand side, both I project on the space say, w only, which is span of; or just a minute; no, I can make it simpler. Suppose, left hand side and right hand side, I consider this space with which you are familiar; what is w 1, p, this span of; actually I am hurrying because of this signal I got that to wind up; this is w 1, p.

Suppose, left hand side I project on this, right hand side I project on this, projections will be same. For the right hand side, u n is orthgonal to this x n minus 1, just now I have seen, because x n minus 1 depends on u n minus 1, then u n minus 2, all that; u n is orthgonal to x n minus 2 also; u n is orthgonal to x n minus p also; that means, u n is orthgonal to this space. So, projection is 0; projection is 0. So, right hand side becomes 0. On the left hand side, projection is a linear operation. So, over all projection is projection of x n on this plus, projection of this on this, projection on next I want this, and so and so.

So, first one will be P 1 p, x n, P 1 p, very quickly I do not have time now; then can you tell me, what is the projection of a 1 prime, a 1 prime is forget, forget, it is a constant x n minus 1; on this x n minus and on if you project x n minus 1 and w 1, p; what is that projection? If itself, it is lying there. So, if you project u this is the that is the one which will be the minum error; error is 0 there.

So, projection is one which gives rise to minimum error, now x n minus 1 is such that the error between x n minus 1 and x n minus 1 is 0, same for others. So, you see, right; that is what is the projection, projection is I think minus here, this is; now can you see, but this projection and all considered a linear prediction; x n projected on these means it is a linear combination of those terms that is minus a p 1, a p p. So, you can identify easily this coefficents; these cofficents are the predictor coefficients.

That means, if there is an AR model satisfied, AR model coefficietns are nothing but the corresponding p th order linear prediction coefficients, that are not 2 the same. I stop here to again next class I will coninue from here. I just I do not know how much it went through actualy; this relates to linear predictions to AR modeling.