Adaptive Signal Processing

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

Lecture - 2
Introduction to Stochastic Processes

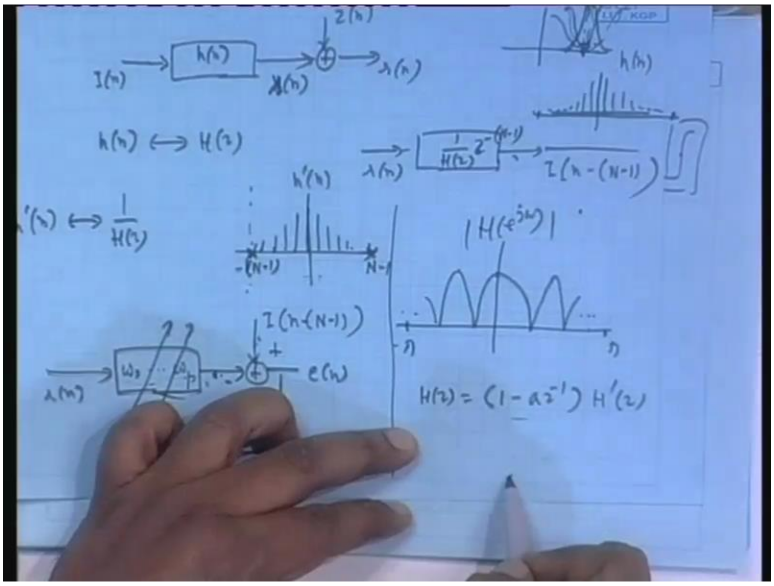
So, we attended the last class, at that time we gave you the motivation for adaptive filter with had some examples. You know examples were taken one from adaptive line enhancer that was if you have got a sinusoidal, which is noisy and you have to filter out you can use a band pass filter. But if the frequency changes from time to time; then you cannot go on redesigning and reframing the filter. The filter should have the learning the capability from the data so that it adjusts its set of frequency accordingly. So, there should be an adaptive algorithm adaptive mechanism, which will learn from the data and adjust itself that was one example.

Another example, which I gave was from echo cancellation, which is very you know I mean valid for telephone and data modem. That is, when an incoming signal is coming which could be either speech in case of telephone modem or more importantly, which could be data for data modem. Then, instead of receiving instead of going to the receiver through the hybrid part of the data comes back in the return circuit as an echo, which gets mixed up with the data flowing from this end already, so that gives to rise to error.

So, then we model that hybrid as an echo generating system as a linear FIR filter of very large order and by using adaptive algorithm what we did we estimated, we identified those filter coefficients. So, basically we emulated this hybrid performance. So, we generated an echo and subtracted the echo from the incoming signal that was in and then since hybrid characteristics changes from time to time this has to be done adaptively that was for adaptive line echo canceller. Another very important example of course, I cannot get into that now very important example was discussed was did I cover fully adaptive equalizer.

So, that where did I stop last time; because, so quickly I will go through that example that was adaptive equalizer, but you know I mean those who are coming for the past time today I cannot for that sake go back. So, the channels model this only I am telling you for the consumption of those who are coming only for today only today or joining today.

(Refer Slide Time: 03:18)



These only are discrete time model; this is not the actual physical channel. The model was this that you are transmitting suppose some sequence I n I n I 0 I 1 I 2 I 3 like that. So, this I n I mean I 0 I 1 they can take some discrete values from a set of some finite discrete values that set is called alphabet. It has a finite number of symbols; if you are taking r bit together. So, two to the power r symbols are possible say two to the power levels. In general, this could be complex in phase a quantized component, but for our business we assume this this should be real for our business here.

If I n there is an then equivalent discrete time model of the channel, which you replace by h n and what you get here is y n or maybe I called it x n that day right what did I call s n is it, this is s n. This h n in general is non-causal, but that is coming because of some the way it was generated the symbols were transmitted. That is you take a pulse you trigger the pulse; the pulse goes you get pulse gets multiplied by a symbol say I 0 and you measure it at the midpoint of the pulse at a point capital T, but even before that point arrives you switch on the next pulse may be after second of tau at a gap of tau.

So, when you measure the first pulse height, second pulse already is present there. So, contribution from future comes that is how the non-causality came up that is how this model came up. This is an equivalent model that was derived; it is not the physical channel it was non-causal in general. Then the whole channel is noisy, there is an additive noise instead of showing it present all throughout we take the entire the noise effect and add it separately as a noise component and you can call it z m. So, this is what you received r n for the time being, if you assume that z n is not present for the time being if z n is not present.

Then if h n is H z z transform then what you do; you pass r n through if we a pass one by H z you should get back your I n. Suppose, z n is not there. H z 1 by H z this 1 by H z is the basic equalizer it is called linear equalizer, linear equalizer it is a linear system, but there is some problem. Problem in the sense, you know they are the things you know I mean not necessarily part of adaptive filter theory, but I am trying to convey this info to those who work in communication area and not these are not given elaborately in books. Normally, when this is h n comes you remember how the h n came it came because of this that suppose there is a pulse then further pulses past pulses still past pulses still future pulses they all interfere here.

Their interference goes down as the pulses are further and further away from the central pulse. The interference component is reflected by the coefficient h z 0 is a pulse height at the center point of the basic pulse. Initially, we said it to be one, but the channel distorted it,, so it becomes h 0. H 1 was the value of the future pulse or may be the past pulse the immediately previous one as it passes through the point in the center point this much height then h 2 h minus 1 h minus 2 like that. So, you understand that this h sequence, since this is a repetition I am going to fast all this we discussed in the previous class.

H sequence is such it is a dying sequence on either side; that is h n could be like these; h n could be like this. Now, so you can approximate by an FIR filter by an FIR filter only thing is this this filter is non-causal; so you can place the origin here all that is fine. So, FIR filter means it has it has got only zeros. So, if h z zeros 1 by h z only has poles and there is no guarantee that the zeros of this filter or poles of filter will lie within unit circle, some could be within unit circle; some could be outside unit circle. So, 1 by h z is not necessarily causal and by h z is in general stable it stable filter; because any real time filter is a stable filter, but it is not non-causal; that means, 1 by h z.

If the corresponding impulse response this h prime n for 1 by h z. How will h prime n look like? It will be going both again in both direction, but it is a stable system; stable means mod of h prime n if you sum it is finite; that means, the h prime enhance to die out. It cannot remain steady; then it is not summable. For any stable system, impulse response is absolute summable means impulse response is always dying down; it has to be always dying down. In fact, all the in the analog case if you deal with r c circuit r l circuit; you see all those e to the power minus t by r c r l by r t by l e to the power minus r all the signals are dying. So, that will be basically impulse responses they have to die out; so that they are absolutely summable.

So, that means this is this will die out. So, you can trinket it up to some point n minus 1 and on this side also its not necessarily symmetric on both side up to this. So, that means 1 by H z is this filter and which is non-causal. Non-causal filter, in general cannot be implemented, but suppose I shift the origin here. I shift the origin here; that means, what output will be delayed that is if I delay this if I shift it if I delay this entire thing; so that this fellow goes here. Are you following this? If I shift it to the right by n minus 1 point; that means, origin actually will go here it becomes causal, but if i delay it by n minus 1 point; that means, here I will be multiplying this by z to the power some amount.

Time domain shifting right means this and therefore, r n if it is z transform is R z capital Rz Rz into this into z to the power minus of this; that means, from previous output, whatever I was getting that will get delayed by n minus one. So, previous output was exactly I n now it will be delayed. So, what will come out is I n minus this; but this is fine with me, because it is only a delayed version of what actually was transmitted no loss of data. So, this but this introduction of this factor makes it causal and therefore, realizable this is where I think I stopped last time introduction of delay factor that makes it this thing realizable.

Now, how does if you want to construct this filter you must know H z. H z comes from the channel model; that means, you must know the channel you must know how the pulse values are what are the pulse values are how the pulse values came originally there was a pulse of p of t that passed through a channel physical channel of impulse response h of t convolved. Original pulse was distorted zero forcing condition was destroyed new pulse p prime t came. That pulse values at various points they gave rise to the samples H zero H one H minus 1 and all that all that were discussed last time.

So, to know those values you not only have to know original pulse p t, but you must know the channel then only you can find out p prime t. But channel is not known to you; further channel characteristic changes from time to time. That means I have to try on adaptive means, which will learn from the incoming that about the channel impulse response and adjust itself.

That means, what I will do this r n I will pass through an appropriate filter FIR filter of coefficient say w 0 dot dot dot w p. How many coefficients the w 0 will be for this w 1 for this w 2 for this dot dot dot w p for here. I want this impulse response to come up; I want this impulse response to come up there, but I do not know this impulse response because I do not know the channel. So, what I do I have to arbitrarily start with some values for this? Total number of points here, how many 2 into n minus 1 plus 1; whatever it is you call it p plus 1 w 0 to w p p plus 1.

So, I have to find out these once. So, what I will do I do not know what their ideal values as shown in figure r. So, I start with arbitrary values. I filter them find the error between what I want the output to be this, but I did not get the actual impulse response of that this filter. So, I will something else not this. So, this will be what is called desired response. That is in during the training phase when transmitter transmits universally known sequence pilot sequence training phase receiver also known. So, receiver knows what is the transmitted it is recorded with itself.

It just delays it by this amount gives it at a desired response measures the error between the actual filter output plus and minus this the error; and this error is actually used in an adaptive algorithm that in that is the topic of the that is the subject to be studied here adaptive algo. This adaptive algorithm what will generate it will from this weights actually we show by this what it does you know. Suppose, at a particular time index the n th point of time, you have some weights w 0 to w p; some value you use them to filter get the output measure the error take the error in the algorithm, which algorithm uses the error.

From the current value of the filter weights, it generates a new of set of filter weights that is called weight updation or weight adjustment hopefully better choice of filter weights. Then, again you filter this, again measure the error and again feedback to this adaptive algorithm; it again adjust it further purpose is that to bring this one as close as possible to this. Because I want this output to be this actually I want this filter to be this only filter. So, the error will be minimized in that direction we will move, but in iterative manner with time. At one point of time, you have got one set of weights filtered using the input data using that find out the output take the error use the error in the adaptive algorithm; then adjust the weights to get another set of weights hopefully better set of weights.

Again filter again measure the error again passing the feedback to this adaptive algorithm further adjust the weights so on and so forth. It goes on in an iterative loop for several rounds of iteration and hopefully it convergence that is after sometime you really get this weight is very close to this and we say it is converged. Once it is converged at sufficient training time for this training is given. So, it is already converged then again normal business starts you no longer; I mean transmitter no longer transmits the pre-designed training sequence.

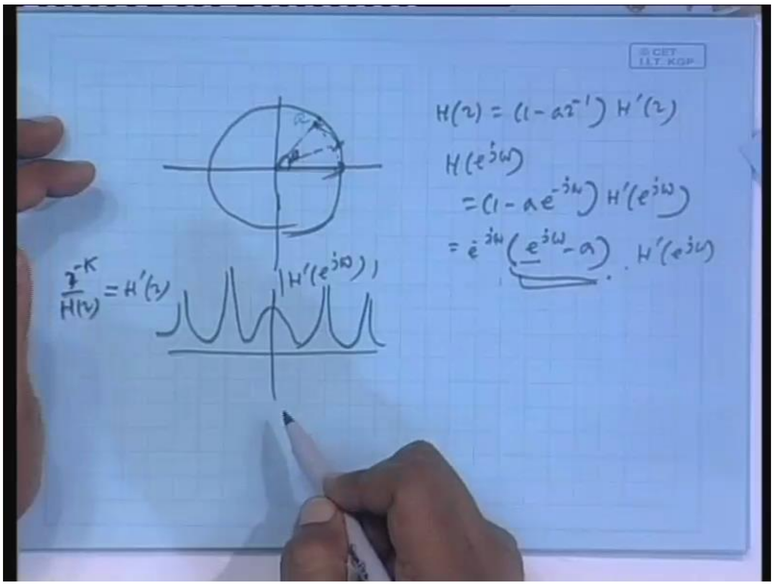
It transmits normal sequences, but now I can assume that I can trust the channel; channel will start fluctuating now at least for some time. So, this will be the filter weights that are converged; they will indeed replace indeed represent that ideal equalizer and I will whatever I will get that is indeed the one that is transmitted. Problem comes when you have the noise because so long I took it only r n, but r n plus z n that actually goes through this equalizer; this after convergence you get this equalizers. But r n plus z n after during training you are training, but after training you are happy that I have got a filter weights a set of weights and they got converged and all those things fine.

I want to trust it; I want use it, but now not only r n r n plus z n are passing through it; that means, there is a component from z n also at the output. So, far I mean often or many times this is not a problem, because there is a noise at the input then noise will be passing through this filter; hence there will be some noise at the output I cannot you see I have to leave with it. What I do? This will be then passed through a quantizer, because after all input levels here I n there are not any continuous valued variable. Either this symbol or that symbol or this symbol like discrete values like say various levels one volt or two volt or three volt or four volt like that.

So, if you have a quantizer like that and suppose indeed it is one volt, but because of noise it has become one point one point one; still between one volt and two volt there is a huge gap. So, you can still make a correct decision. In the decision process itself some effect of noise can go, because say I have one volt you have got a slot you have got a band 1.5 on the top and 0.5 on the bottom. So, one within that band anything falling will be equated to one. So, originally that I mean receive sequence should have been one receive valued would have been one, but because of noise it is a one point something 1.15 1.3 1.8 or 1.9 all will be quantized to the correct value only. So, by the quantization process itself much of the noise will go; that is what we hope we could have been living happily with it.

Sometimes, there is a problem that sometimes what happens you know? This equalizer they show and special this H z channel mode; H z shows that it has got something called spectral null. Spectral null means you know something like this; if you take H e to the power j omega it deeps it almost touches zero; if we take the mod of it touches zero like that dot dot dot at phi like that. Sometimes it happens it either goes absolutely to zero or very close to zero. What does it mean on H z? H z is FIR and the FIR filter suppose it has got a zero like this 1 minus say a z inverse into the remaining part; say H prime z concentrate on this.

(Refer Slide Time: 18:39)



Suppose, a is 1 minus a z inverse means, where is the root 1 by a at one by a at root is at a; it is z inverse root is at a. Now, suppose a is this is unit circle and a is within unit circle, but very close to the unit circle; this is where your a is this is a and a is a complex number. So, therefore, this much is the magnitude of a and this is a phase angle could be say theta. Now, H z I am repeating a inverse this some H prime z. So, if you want to find out the frequency response it will be 1 minus a e to the power minus j omega.

These now suppose omega is changing, I move omega from zero to two phi or minus phi to phi any DTFT, any frequency response is periodic over two phi all of you know. That is basic DSP, so it is enough to plot it from minus phi to two phi. So, suppose I have started at zero so; that means, 1 minus this a e to the power minus j omega. If you want you can write it this way also e to the power j omega e to the power minus j omega e to the power j omega minus a; this part concentrate on this factor say at omega equal to zero. You see at any omega this magnitude is one this is one phaser; this is one phaser this is a phaser difference, this much when omega is zero this minus this, so this much.

Then as omega increases you are say here; then this much, so the corresponding magnitude and phase that will determine if you write it as a complex number. So, the corresponding magnitude part will come and a phase part will come. Slowly omega is increasing, as you see as you are going closer closer closer to the point the magnitude part of that phaser difference phaser is decreasing and when you are aligned that time the magnitude part has the minimum value.

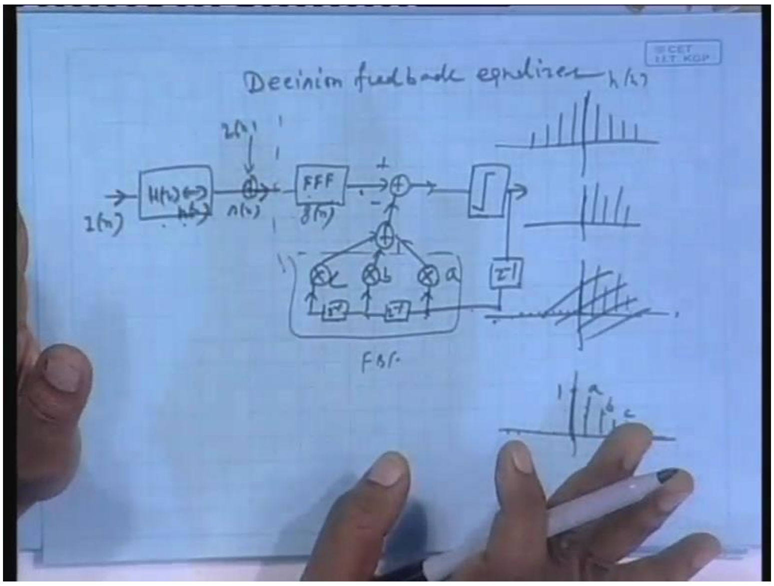
In fact, if this a is almost on the unit circle magnitude part of that phase difference phaser will be almost zero, which means you will have a situation like this; net magnitude if you take the mod of this that is mod of this mod of this mod of all the factors. So, mod of this transfer function will touch zero and very will become very close to zero. So, that means this spectral nulls indicate presence of roots that is zeros very close to unit circle in the z plane.

Now, if that be the case I was discussing H z, so this equalizer is 1 by H z so; that means, there will be spectral peaks zero has become poles. Earlier, it was just the phaser the magnitude the difference phaser magnitude that is close to zero. So, overall magnitude of the response was zero or very close to zero, but now it is not H z it is one by H z. So, it is like one by zero or one by something very close to zero. So, that instead of infinity is the peak so that means, 1 by H z into z say to the power minus something; say some constant k if you call it say H prime z; that means, this will have peaks like this, so which means noise will be enhanced.

You see here, if this has a peak it has nulls H z has nulls this has peaks they will cancel each other. So, signal I n was passed through first H z, then 1 by H z Hz and 1 by z canceled I should come out as it is, but look at z n z n does not pass through h n z n passes only through over this filter. So, z n finds there it certainly is enhanced, because it passes through a filter which has only peaks. This is called noise enhancement effect in equalizers noise enhancement effect. To cancel that a modified form of equalizer is used, which I will just discuss and then I will get into the proper theoretical beginning of this course, because we are still going through the examples.

That equalizer is called adaptive decision feedback equalizer. Normally decision feedback equalizer if it is made adaptive it is called adaptive decision feedback equalizer. For the time being again, I was not dealing with adaptive, because I assume 1 by H z to be given H z known one by H z known adaptive part, I did not I had it separately.

(Refer Slide Time: 24:25)



So, decision feedback equalizer decision feedback equalizer here, what we do this suppose this is the channel H z which is equivalent to h n. I n is going through this, what they do and here this is your s n for the time being I do not add the noise part. Otherwise, I would be adding noise part z and all for the time being I am I will add later. What they do you? First, uses these are the receiver side first use what is called FFF? FFF means a FIR filter called feed forward filter just the FIR filter. You have to design it carefully this FIR FFF may be you call its impulse response; you can call it this is h n here you give a name here g n.

You design it such that what happens; this h n and g n they come in cascade. Channel h n followed g n; g n is new construction, but that is at the front end of the equalizer, so channel immediately after that g n. So, what is what the impulse response is? That is h n convolved with g n at least you have done some of you have done analog and electrical people already know some DSP I suppose. Our students do not know that is the big disadvantage electrical will know. So, there will be this discrete convulsion between h n and g n you design the z n; so that after see h n is non-causality this is h n.

This is you h n say something like this and you are convolving h n with g n. What you do g n is a FIR filter how do you do convulsion. Suppose, the FIR filter was like this something like this. You reverse it and then go on shifting this way or that way and find out overlap from top and sample wise multiply and add that is the way to convolve. You design it such that, on this side this is our result; on this side you have got very low value almost zero and at this point from this point onwards you get something some values up to some point.

After all two FIR convolved will give rise to an FIR only, so it will go only up to some point. Suppose history was so being I do write like this say only three point and one this side zeros that is how you design there are design equations and all. We design so that means, together it is a causal system of this much impulse response; this is the sample of this equivalent impulse response at zero th point it can have any height, but I can normalize to one. That if it is say ten I will have just one by ten factor, so it becomes one that is important.

So, I will construct this to be one and then suppose this values are a b c. Then what I do this is my quantizer here only output decisions will come up, after all it is an equalizer. So, finally, your process sample has to be fed to the quantizer that will say this is closest to that symbol or this symbol and that symbol addition will be made. This decision, I will pass it through first a delay. So, past decisions will come up and suppose I have a thing like this. What does it mean? Suppose, past decisions are correct suppose till this point of time your past decisions that is decisions are correct.

Right now, you will be the correct decision present decision that I am coming later, but after this delay here past decisions will come. Suppose, past decisions by hook or crook are connects; the indeed correct values of I n, which you have transmitted are where detected like that at this past three. That means, here I have got correct value of I n here; I have correct value of I n here I have correct value of I n. They are multiplied by a b c and added and then if this addition I subtract from this guy. What will I get? Look at this. This two together we are following this cascade one a b c; that means, I n passed through just a simple system FIR system one a b c that came up here what is that then.

I zero into one plus a into I n minus one; I n into one can you see or I have to you know convulsion or you have to reverse it and shift it to the point n. So, I n into one plus I n minus one into a I n minus two into b I n minus three into c that summation will come up here, that summation will come up; but here, if I assume correct value I I n minus one n minus two n minus three available here. Then, there will be multiplied by a b c, there will be multiplied a b c added; there will be multiplied by a b c added and subtracted. So, what will be left here?

No, I n into one only I n will be the left only I n will be present, past portion I n minus one I n minus two I n minus that component is exactly subtracted. So, I n will be left and you pass it through the quantizer; it will happily quantize into I n it will be forming loop and actually it will work. Here, I am no longer nowhere inverting the channel. So, even if his channel has zeros close to unit circle, I am not having that constraint that zeros here become poles here and therefore, spectral null means spectral p can therefore, noise enhancement.

So, even if there is noise now added; some noise will propagate here and therefore, here but the spectral peak kind of thing will not come up. So, in that quantization process I can assume that noise effect to be minimized. It is a beautiful technique you know; this is called decision feedback equalizer. Now, here again FFF g n and this coefficients actually this is another FIR filter. Input multiplied by a one delay multiplied by b another delay means this is another FIR filter, this is called FBF filter. So, this g n and FBF this coefficient how they are found out first here, I assume that h n is known and therefore, you design a g n.

So, that you get things like this and once this is known then you find take out a b c and form this coefficient, but that means, you are assuming h n to be known. But, when the in the case where h n is not known then again adaptively, you have to find out g n and this that will make it adaptive decision feedback. So, this is another very important example for I mean adaptive filter. There are plenty one another example I forgot I forgot to mention today, another example I gave that day that was a smart antenna beam former.

That is in the same sell I can give same frequency to two users located at two different directions, but by adjusting the antenna gains I can form a beam in one direction and null in another direction.

So, that even though two users are using the band I can here only another person cannot interfere and if this person is in motion my beam should local to them. So, it has to adaptive. That is called adaptive beam forming in the last I give that example too. This gives you enough motivation, but I suggest that you go through the book by Haykin or Farag or there is a book by the father figure of elimination algorithm called Widrow, Bernard Widrow and Stearns. That is adaptive filter book plenty of copies are available in library and it gives lots of examples very practical very useful examples on each example there can be a course.

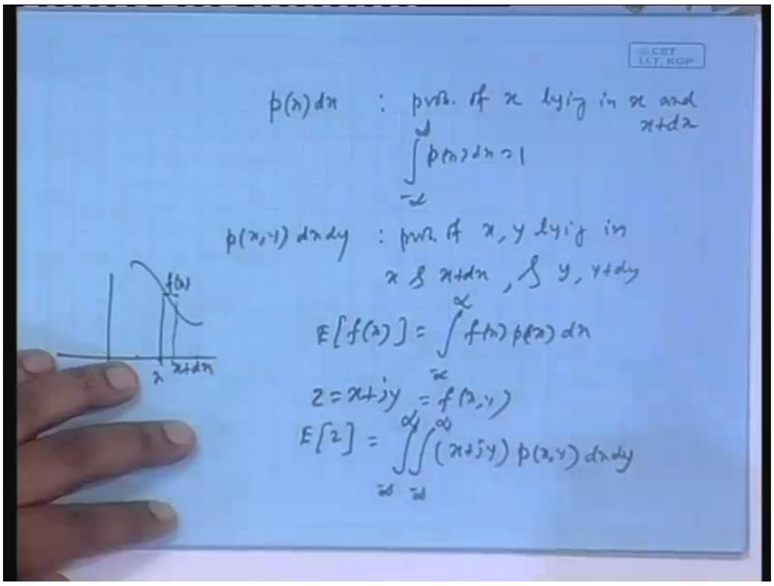
So, just I mean to get more and more practical motivation you can go through some of the examples you know very nice examples and you can implement them in lab also. So, I will not go through this example thing anymore; what I will do I will now get into the theory of the course. I tell you beforehand that this will be a mathematical course; any question there. No you can go if you want can go, but let me go on with recording. This is a mathematical course, the mathematics that is required will be developed almost like physics we develop without diluting any rigor everything will developed everything with proved.

But you have to have that love or interest in mathematical think, mathematical thinking mathematical derivations our purpose will be to derive adaptive filter algorithms. This is a course on derivation of and also some performance analysis of adaptive filter algorithms and dominates mostly statistics or stochastic process, because most of the time we will be optimizing something called mean square error. And the moment you have mean square error variance immediately, the e operators type will come up and therefore, all statistics will flow in. So, we have to start with some elementary notions of stochastic process and all.

Another thing, which I will need to make this course very neat and very nice, is linear algebra, which all electrical and electronic engineers must know. But unfortunately here I mean do not know, but you have other department in our department; we do not have this course. Linear algebra is a key to signal processing communication control today.

So, that also whatever linear algebra needs I will develop in this course. So, whatever be the amount of time that is left today; let us start with this stochastic thing, I have assuming you already know probability statistics and all, but then I have to develop the things the way I want.

(Refer Slide Time: 35:30)



You all know say, if we take a continuous valued real variable x; you know it has got a probability density p x, which is a non-negative function. What does p x means? p x into dx will you give the probability of the random variable, x to lie that maybe I am not writing everything because then on writing I will be wasting too much of time. I am trying to go fast on this initial portion I will not go, so fast definitely when I go further. Because, things what I find there are repetitions of things, which I expect students to know I do not want to you know I mean go slow.

Anywhere, this will be probability of x lying in x and x plus dx, that is p x dx and it is such that p x dx integral is a total probability of x lying within the range. So, it has to be global constant independent of whether it is x or y or any other random variables that constant is taken as one all of us know that. Then, if x and y they are jointly random variables two parameters which are random, but were which have got some dependence between them say temperature and humidity. So, then p x y is a joint probability density what does it mean; that this into dx dy means the probability of x y lying in x x and x plus dx simultaneously between y and y plus dy and simultaneously.

So, once again if you double integrate it, what the entire range of x and y? You will get you have to get the same constant y and all that. Then, you instead of double you can have multiple p x y z u v like that. Next, is if instead of p x suppose you every time you are the random variable x and there is some function given to you f of x sin x cos x any some function. Every time x takes a value evaluate the function. So, that function also is a random, because this argument is changing randomly. So, what is the expected value of the function when x is lying between x and dx.

You can approximate the function piecewise you know constant function; it will take the value f x, but why there are how much what is the probability that times p x, f x into its value will be f x between x and dx the function will take the value f x, I am doing the piece; are you following me or i have to draw. Suppose, it is like this you are here x and between x and dx you are approximating it its value is f x because dx is an infinitely small. But what is the probability that I measure x and I get this much of f x; that means, that into p x dx. So, that is the probability of x lying in this range.

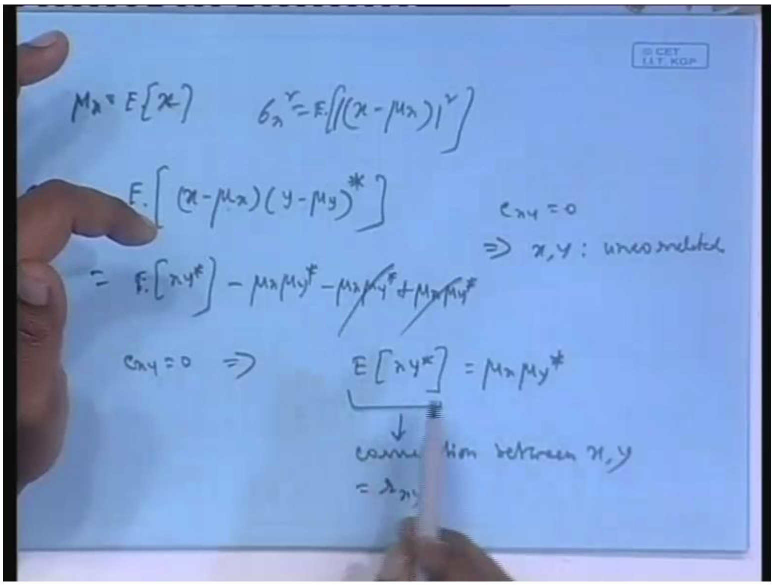
So, what is the expected value that will be integrated over the entire range? That is within this range expected value of f x is not just f x; f x I have to multiply f x by p x dx that is the expected value. Because it is not that every time I measure I will get the function value equal to this much f x, sometimes this is this much sometimes not this. So, what is the probability of that function taking this much value? It is not f x because always you will not get f x. So, it is qualify f x qualified by another amount modified another amount that is at p x dx that is for any two between any two value.

So, if you want to do it from minus infinity to infinity then you will get the average value of f x average value. So, E and then this is called E f x. E f x is f of x p x dx over the entire range, that will give the average value over the entire range, earlier it was giving the expected value only in the small range. If you integrate over a finite region, it will give you the expected value within that region. That is, if you integrate it between one range then and f x goes from one is to another then what is the expected value of f x; you will get that is now integrate it this all of you know.

In general, instead of f x if it is f x comma y then E of x comma y will be f x comma y into p x y dx dy double integral and you can extend it to any dimension. As an example, if z is complex if there is a random variable z which is complex and both x and y are random; then E of z again this z you can call as a function of x y after all. Then, E of z which is will be nothing, but z or x plus j y times, so this z will be called E of z will be called the mean or expected value or mean of the random variable z. So, instead of z you can just come to the original value x.

What is E of x? x into p x dx that is on the function is the value itself; f of x is x where f of x will x itself then E of f x means E of x, that is x into p x dx integral that is called the mean of x average of x you all know that it is just bracing up. Then, so you know also know what is mean by E of z if z is complex. Because soon I will be dealing with complex valued random variables.

(Refer Slide Time: 41:44)



So, now suppose you have got a random variable x, which could be now real or complex I do not care E of x we all know. What is E of x? So, you cannot ask me what is the meaning E of x, I have given you the definition you knew already, but I am doing in a systematic way. You call it the mean or expected value of the random variable x and you call it mu x. Then around the mean if you take it is this difference this difference in generally is complex. So, if you take the mod square of that you will get a real quantity, which is random because even though mu x is constant this.

So, this differential it is a difference like AC mu x is DC this is the AC, you take the square of that every time that fluctuates. Because x value changes and you take expected value of that that will be a measure of the power AC power which is called variance. A variance is denoted like this; then if you take two random variables, joint fluctuating random variables may be complex x y and if you take this quantity this is called covariance; c x y covariance between x and y. What is the meaning of this? If this quantity is high suppose x and y are such they are very much you know, I mean related to each other when x is going above its mean all that y also going above.

So, both the differences are positive positive into positive positive and when the x is falling below the mean y also they are hand in hand. So, falls below the mean, so both negative differences again product is positive. So, either case if they are related to each other you get a high value, but if they are very much uncorrelated then this can this difference can sometimes be positive this can be negative or positive vice versa. So, if you take a many a huge number of cases and average out you will get less value; that means, this will be a measure of the degree of correlation between the two random variables x and y.

Two random variables x and y will be called uncorrelated, if this quantity is zero zero implies x y uncorrelated. What does it mean? If this is you can expand this x into y star mu x into y star like four terms; you can take this product now there will be four terms. So, you have got one term like E x y star minus mu x y star E of that mu x y star E of that E you can take inside mu x constant goes out. So, E of y star E of y star is mu y star. What is E of y star? E of y star is same as E of y then star; because E of y star also means y star into p y dy p y is real integral p y is real. So, you can put the star outside in that way. These things I do not think I do not tell you people you know; it is very simple.

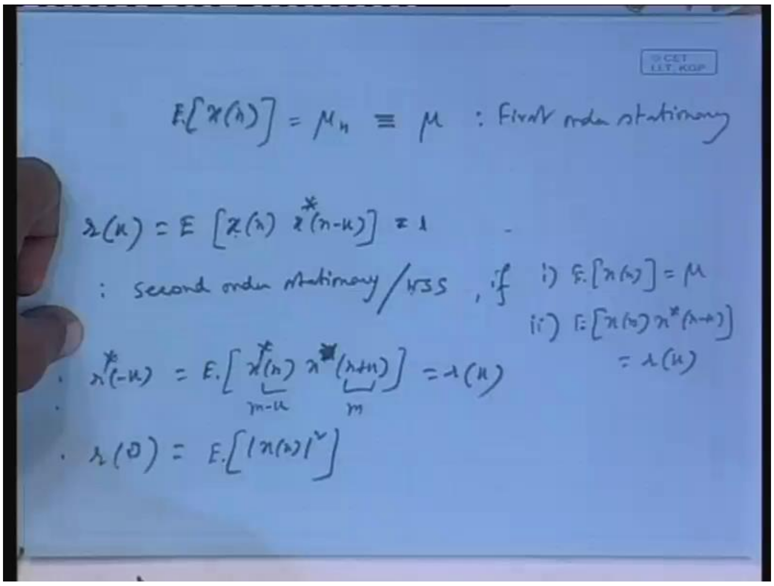
So, mu x mu y star and then E of x into mu y star, that is E of x mu x again mu y star and mu x mu y star both are random constant, so no E on that. So, these two cancels you have left with only this. So, if this is zero because if x and y are uncorrelated the c x y zero implies; if uncorrelated, then this quantity this quantity E x y star this is called correlation between x y you denote it as r x y correlation between x y. The order is important when they are complex with x between x and y; that means, x first y second star on between y and x means y first y second star on x. So, this order is important in the case of complex variables.

Of course, for real x y and y x are same no conjugate all that. So, you get the same thing, but for the complex case x y star and y x star are two different things. So, if two random variables are uncorrelated then their correlation is product of their mean, first one as it is second one conjugate it; often we deal with zero mean random variables. Often we deal with zero mean random variable in that case, if they are uncorrelated; if two zero mean random variables are also uncorrelated; that means, correlation will be equal to zero because zero mean zero mean zero into zero.

So, you can say that two zero mean random variables are given and there are correlation is zero. Then I say they are uncorrelated because this zero can be written as zero into zero and zero is you can write it as mu x you can write as mu y star. So, this formula is valid so; that means, they are uncorrelated. If two zero mean random variables also have zero correlation, then I can say they are uncorrelated if there correlation is zero they are called orthogonal. Now consider a discrete sequence, which is random that is you are measuring some waveform. Waveform means sequence you get a pattern next time, you measure you get another pattern next time you measure another pattern.

Every sample, say n th in each pattern it is not repeating its value in all the cases its fluctuating. So, any sample is a random variable changes with observation to observation. So, every sample is a random variable that is a discrete random process.

(Refer Slide Time: 48:15)



In that case, you take out say any particular one E of x n you all understand the meaning of E; now E changes from case to case mind you. Like here, simply if you take call this variable x multiplying that by the corresponding probability density and integrating; if I bring in many then joint densities all those can be avoided clearly by using E, because before that is why I explain the meaning of E beforehand. E of x n what is x n, you have got a sequence n th sample, which is the random variable. Because next you measure the sequence it changes the values next time it changes its value so on and so forth.

So, E of x n will be the mean of the n th sample you can in general call it mu n, because that should depend on the point of measurement time index. But often we model this random sequence is a stationary process. If it is first order stationary, it means this fellow is random, but its randomness is homogeneously distributed along this axis that is there is no bias. In the sense, that may be sometime in one experiment I get do a sample to take very high value next time. In next all occasions, it does not mean your sample will still have very high value often it may have very low value; somewhat else may have very high value that could be at hundred location or two location or n location .

So, the fluctuating property is not localized that around this zone of the time axis the random variable will take very large values. So, mean is high around this z it will take very small values. So, around this zone this has a different mean very low mean like that.

If it is purely random it does not know, where its origin it does not care for this time axis it its randomness is going on as though from minus infinity too infinitely and that is homogeneously. Then the main thing whether you measure at n th point or m th point or k th point, then you should ultimately get the same thing; you can model that way and this model is not is only an ideal mode by the way.

Then, it is called first order stationary process, first order means only x n first order no product no square term or no product of, so first order stationary. So, this becomes mu means first order stationary n goes. Because this is x n I am not saying x square n; x cube n or x n into another value of x, degree of x is only one, so first order mean is first order thing a stationary in the first order quantity. Actually, this is called first order moment this E of x n is called first order moment; E of x n square n is called second order moment so on and so forth.

So, if first order moment is stationary independent of n then it is first order stationary that is our language of pure statistics. Now, we have done fine. Another thing I discussed that is the correlation covariance and all that. Here, if you take two samples one is at n th point another at a gap of k and they are in general complex say and you want to find out the correlation between them; one is the n th sample another is the n minus k th sample. So, first guy second guy second guy put a star if the in general they are complex and take an expectation.

So, this quantity in general should depend on what the two points n th point and n minus k means k th point k th value n is already known. So, it should be a function of two indices n and k. But suppose again I feel that say that you see this degree of correlation which is given by this, this is after all degree of correlation between these two samples. If I say that look the process is such its correlation property between samples should depend only on the gap how far or how close not, where exactly they are located whether there are this two or this two or this two or this two; because that correlating property also does not depend on particular zones of the time axis.

That is two samples are highly correlated or very less correlated that should not depend on where in the time axis you are observing them; that should only depend on how at what gap or they located straight. If you model in that way then it should become only a function of k and then it will be called and this is a second order thing product. There is only one term here is x x x square actually second order. See, if this becomes a function of k and you can call it r k then it is called r k. Now, if the process is such this is one extra thing that this both first order stationary and satisfies this; then only it will be called second order stationary when it is a second order it is curved that this is a subset of this.

So, second order stationary or we call it wide sense stationary wide sense stationary WSS wide sense stationary if both the things are satisfied. E x n is just mu independent of n there are stationary in first order and E it is only a function the lag k the gap k is called lag lag variable. This correlation has some properties and for this time. So, there is couple of minutes. You can see one thing what is r minus k. If I put a star here the star here star can go inside after of E means an integral with probability. So, I can put the star inside because probability is real function.

Star goes means from here star goes star comes here n plus k; if you call m this becomes m minus k. So, x m into x star m minus k again it is a function of k only so r k. So that means, for WSS processes r k and r minus k they are conjugate symmetry this is called Hermitian property Hermitian Hermitian; r of k is same is as r of minus k star. So, if it is real valued no star means this r of k and r minus k are the same. So, it is what do you can even function or symmetric function; r of k and r of minus k if they are same they are symmetric even function even function and it otherwise it is conjugate symmetry. In any case this is called Hermitian property.

That is one thing another thing what is r of 0? r of 0 is which is the average power both DC AC put together no subtraction of mu, but if it, so happens that you are we are dealing with zero mean processes. We are dealing with zero mean processes that is all the samples have zero mean and therefore, mean is stationary in mean and then this is satisfied is a function of k only. If we consider this kind of processes, then E of mod x n square is same as AC power only. Then, we will call it variance for zero mean processes this should same as variance there no need subtract the mean and this is always greater than equal to zero real.

Correlation, in general is complex please note this properties; I will make use of this in my derivations in future. Correlation could be complex, but r of zero is non-negative real it cannot be negative square; if it is equal to zero; that means, at any n E of mod x n square equal to zero means at every trial take up any sample it would take only zero value; then only expected value also can be zero. A random variable say a random variable x n mod of x n square or x mod of x square I am saying average is zero; that means, whenever you try x its value should be zero only then only average will be zero.

That means, it is a random variable which always takes zero values and then you can ask me; if it always takes zero value what is random about it? Now, this is something which you have to except as a random variable. So, what statistics people will do they apply a trick. They call it look actually it is a random variable, which takes zero value with probability one; you know this is just a trick. So, here I will stop I will discuss some more properties of this correlation and including Cauchy Schwarz inequality and all and a little bit of spectral analysis.

Then, I will directly move into what is called Wiener filter or optimal filter; and from optimal filter to LMS algorithm for the real case and complex case. So, we will have ahead long you know entry into adaptive filter.

Student: Power is something related to time right.

No, power actually for random thing I should only have this; the time thing is a follow up. Power means what? A random variable is given I will treat at yes I will treat look at the sample purely, it is a random variable located at n th point. Because of stationary assumption; you get the same value whether it is n or m or l or k, but this value is changing from time to time. I mean this observation this much value the next observation high value. So, it is changing, so what is the average power of that particular sample? Then there is a concept called ergodicity that since you are asking I will use the notion of ergodicity in the next class.

They are under I have to explain the physical meaning of ergodicity. There this ensemble average can be obtained by time average and then you use there. Because in real life you cannot go on waiting for many experiments and then calculate the ensemble average and all calculate the ensemble average and all. So, ergodicity using an ones side one sequence only using temporal average time average you can find out this. But anyway time is up I am recording. So, either this questions I repeat that when you cross the time we cannot enter any question within the time frame we can discuss outside offline. So, that is all for today.

Thank you.