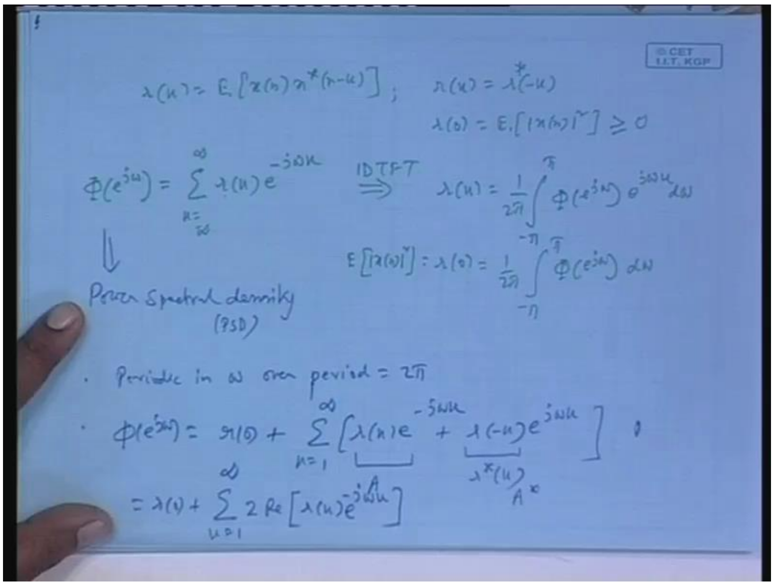
Adaptive Signal Processing
Prof. M. Chakraborty
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 3
Stochastic Processes (Contd.)

So, yesterday we were discussing **y**ou know the elements of probability and stochastic process. We went up to correlation and then we took up discrete random signal, and then for that we discussed what is called the stationarity, wide sense stationarity correlation and then autocorrelation function has one property, which is the Hermitian property.

(Refer Slide Time: 01:32)



That is say, we defined r k as E of x n x star n minus k; this is just for recalling. This is what I did you see function of only the lag or the gap k is called the lag variable r k. We found we established this property Hermitian property. This r k is same as r minus k star we proved this yesterday it is very easy. If it is real valued signal the star has no business, which means r k is just r minus k it is just an even function; otherwise it is conjugate symmetry. Another thing is another property is this r 0, which is E of mod x n square; you understand if you put k equal to 0 x n x star n, which is mod x n square this has to be real and non-negative.

So, this is what we saw yesterday. I think yesterday I stopped here only. Can you present? So, I start from here. There are some more properties of this autocorrelation, but that about that I mean I will let you know there is something called Cauchy Schwarz inequality and all. So, that I will present later when need arises not now. Because afterwards I will presenting this entire stochastic process in a vector space frame walk geometric frame work and there you know lot of things will be very clear from geometric insight using geometric insight.

So, that time I will develop some more I mean properties specially some particular inequality, which is called Cauchy Schwarz inequality, which is very useful in proving things very very useful property and key in communication mostly, so that I will not go into. Now, this r k is a sequence it is totally random of sequence x n is random; if you are going to apply E over it you get a deterministic value. So, r k is sequences the movement you have a sequence you can be tempted to take discrete time Fourier transform DTFT of these. Suppose, just we take a DTFT; so r k if you do not know DTFT I mean those, who do not know they you have to read it I mean sorry it is not a course on first course on DSP.

So, I cannot help that. If you take this discrete time Fourier transform omega is the digital frequency, whose unit is just radian; if you know take this k equal to in general minus infinity to infinity and even though it is a function of omega in DSP the style is to write it as a function of just not of omega, but e to the power j omega. That is, we can call it phi and by the way phi here; you shall write phi e to the power j omega though e is not a variable; j is not a variable; j is not a variable; e is not a variable; j omega is a variable, but still this is the notation we write like that.

And you have left mobile turned on I should switch on my mobile too you know because that will get recorded. So, just one minute those who are with their mobiles please sorry for this deviation. E is not a e is a constant; j is a constant, but still we write phi e to the power j omega actual variable is omega, which is digital frequency, but this is just style this is we write it like this just to indicate that phi e to the power j omega is actually a power series in terms of e to the power j omega which it is as now, this thing has some nice property any DTFT as such you know is periodic function of omega over a period two pi.

So, they if you want to plot it is enough you will plot it from say zero to two pi you have minus pi two pi. We normally will plot it from minus pi up to plus pi and then that will get repeated and that is common for any DTFT, which is valid here also. But what does this give; what is the meaning of this and what are the properties this has let us find out I am we have just taken the DTFT, but is it defined from is it something special does it have some extra property. What does it what is the physical meaning of this particular DTFT then let us find it out?

Now, you know inverse DTFT if I want to find out r of k; you mean this there is an inverse DTFT that is IDTFT is r of k where this is standard. So that means, what is r of zero r of zero is, but what is r of 0? r of 0 is E of mod x n square; because of stationarity this quantity does not depend on n. So, at any index n mod x n square is an instantaneous power for a particular waveform and if you take expectation there is an average power and because of stationarity that does not depend on the n you choose everywhere it will be same.

So, it is the average sample power per sample average power that quantity is what the integral of this DTFT is; so this quantity phi e to the power j omega. What does it indicate? What does is indicate is this you know that it kind of gives a distribution of the total power r of naught r 0 over the whole frequency range from minus pi to pi, this gives a distribution. Once you integrate net integral forget about the one by two pI scale factor; the net integral gives you this average power. So, this quantity gives a distribution; that means what is this quantity if you take a zone from omega and omega plus d omega within that small small strip.

How much is the power? phi e to the power j omega into d omega. What is the total power I integrate; so that is where it is called power spectral density PSD power spectral density. This is called power spectral density is what PSD. What are the properties? Firstly, periodic in omega over period two pi that is standard property of any DTFT; we know there is nothing special. But other important thing is, we know that this quantity r k; r k is Hermitian sequence it satisfies Hermitian property what does it mean?
If you really take this summation divide into two half one from say plus one to infinity another say minus one to minus infinity and take r zero separately.

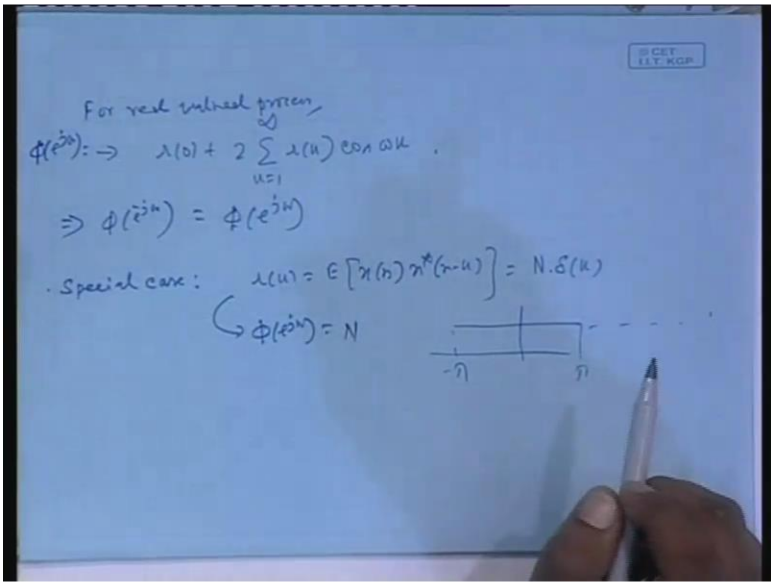
Then phi e to the power j omega it will be what r 0 into e to the power zero that is one plus; what will be from k equal to one to plus infinity r k something e to the power minus j omega k another will be from same thing, but k minus one to minus infinity. Suppose, I am not writing some steps you here be suppose if that another summation instead of k; I call it minus k instead of k replaced by minus k prime. So that means, the summation will be from one to infinity; it was minus 1 to minus infinity k is replaced by minus k prime. So, in terms of k prime it will be from minus one to infinity that is all and wherever I have k that will be replaced by minus k prime and why again minus k prime call it minus k.

So, put the two things under the same summation r minus k e to the power plus; because it is already minus here and k is replaced by minus k prime or rather minus k. So, minus minus plus e to the power j omega k and what is r minus k? This quantity from the Hermitian properties r star k from the Hermitian property here r minus k is r star k. So, if you call this quantity as A this is A star A and what is A star; r star k e to the power plus j omega k. So, A and A star when summed is a real quantity. So, it is a real function and r 0 is always real it non-negative, but real.

So, this is what does it give then actually the summation will be like this; r 0 plus twice real part of r k e to the power minus j omega k. If r k is real so it is real functions mind you; because of the Hermitian symmetry of the autocorrelation sequence; this is a real valued function. Obviously, it should be because when you integrate you get power. So, power is not complex its real quantity. So, this integral is it is should be real it is because integral is giving you real. Now, this is a real function we have proved, whether you have got a complex valued sequence or a real valued sequence this is a real quantity. This is proved.

Now for the particular case, where x n is a real valued sequence and therefore, autocorrelation function r n also is real valued; there you should the inside quantity real valued real part of r k e to the power minus j omega k, r k can be taken out because r k is real. So, only real part of this term e to the power minus j omega which is cosine omega k? So, this leads if you permit me writing here, if I write can you see this if I write something new. So, can you see this here?

(Refer Slide Time: 11:46)



So, for real valued signal sometimes that random signal is called random process. Real valued process mean a random process random real valued signal, this will be equal to r 0 plus. So, one thing you have seen this is a real valued function. This is another property we will show that is actually a non-negative function, it is not a real. It can be never become any point if you plot it from minus pi to pi, it will never become negative e it is always non-negative that proof we will look at some part some more things. Actually, I am forming some background for taking a plan into this statistical signal processing or adaptive signal processing course. So, it is just background.

Some of this may be familiar with you, but I have to put them in the framework I prefer. Because this is I would say quite elementally and for those who are not aware of this you know it is I think, it will be a good exercise will be good exposure. Now, one thing you see for the real case if it was not real this guy phi e to the power j omega is I understand real valued, but phi. If suppose omega is replaced by minus omega. Here, it will be e to the power class j omega k. but this entire thing will be different quantity again. So, that overall function may not have anything to do with you cannot find out adequate relations between that function and phi e to the power j omega phi e to the power plus j omega.

But suppose, I am dealing with real case then r k will come out and you are you have got this formula. In this case phi e to the power minus j omega that is at a frequency minus omega. What that will be? This was same this was originally because cosine minus and cosine plus just same. So that means, for the real case it is not only a real function for the real valued signal or real valued process cases power spectral density is not only a real function it is a function which is even. At omega, whatever is the value same value you get at minus omega. So, it will be symmetrical around the vertical axis.

We can also show and we will show that it is it is a non-negative function, but you have understood the physical significance of this. Suppose, you have got a process, you have got a random signal, where every pair of every two samples whether they are adjacent neighbors or they are far away from each other. If there is no correlation that is suppose special case. Suppose, r k which is E x n it is some constant capital N times delta k. It means and I am dealing with zero mean random variable process mind you; I am dealing with zero mean I am also dealing with zero mean process.

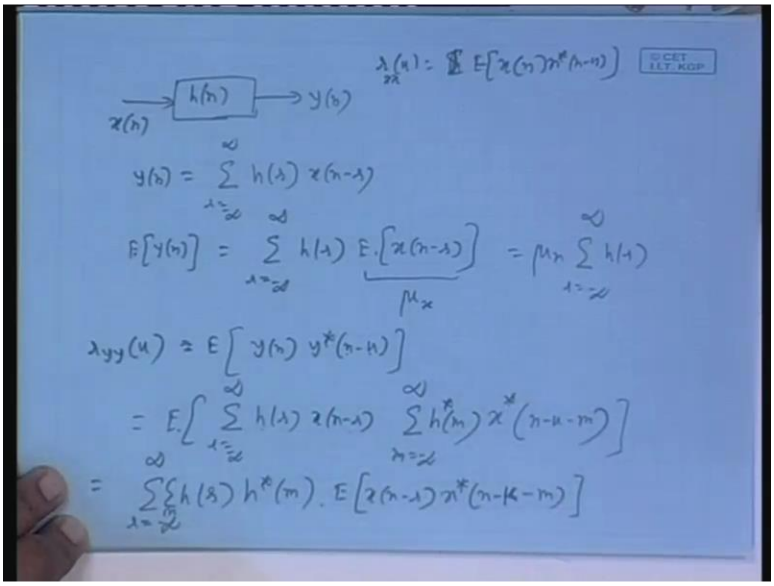
So, suppose r k is like this. What does it mean? When k equal to zero when k equal to zero then right hand side is capital N, which means E of mod x n square there is a variance is a constant independent of n because of stationarity and that is equal to capital N. So, capital N is the variance of average power and otherwise the correlation is zero. For any two samples, whether k equal be one or k equal be one million the correlation is zero since I am dealing with zero mean random signals correlation zero means; they are uncorrelated I told you yesterday. In general, two random variables x and y are uncorrelated if E of x y star is equal to E x into E y star what when you are dealing with zero mean; case then E of x that is mean is zero Ey study is zero.

Conversely, for zero mean case if E of x y star is given to be zero you can interpret in this way that E of x y star is identical to E of x into E of y star, because E of x is given to be zero E of y star k is given to be zero. So, you can write it like that and therefore, it stands to be uncorrelated this I did yesterday so for a zero mean random process that has this kind of what a correlation function. It corresponds to a random process, where there is no correlation zero correlation between any pair of samples. Only one sample is multiplied by itself that is because variance it has to have a value capital, N which is non-zero greater than a positive number.

N is a positive integer positive number not integer, for such case for such kind of random process, which is actually purely random absolutely random; no correlation between any two persons highly random highly random that is what it means no correlation total chaos for that. What is the power spectral density? If you take the DTFT of this, what do you get can you tell me? If the DTFT formula N into delta k into e to the power minus j omega k summed, but k equal to zero case only has to be taken k equal to zero means e to the power zero one and a delta zero is one so N.

So, it is a flat thing from minus pi to pi and therefore, it will continue to be flat. This process is called a wide process wide signal or wide random process discrete time wide random process. Wide process comes as you from the wide light the phenomena wide light consist of all frequencies possibly with equal intensity and all… So, here also it consists of all frequency. So, that is the reason why they call it wide. This is the special case a wide process means there is no correlation with any pair of sample.

(Refer Slide Time: 18:11)



Now, one more thing supposes there is a zero means or let it non-zero mean for the time being here in this context it is more general. So, it is not zero mean there is a random process, which is stationary WSS. So, a stationary mean stationary in correlation mean is independent of the index n, which is the everywhere same correlation is a function of only on the gap or lag k. That kind of process you are passing through a linear time invariance system, you know linear time like if I FIR filter and all linear time invariant system of impulse response h n in general.

I am not saying it is causal or non-causal. It is not important I am here in very general and you are getting one output y n. Since, input is random output also is random question is output also WSS number one then number two if WSS; then what is its mean? What is its correlation autocorrelation? Let us examine that because adaptive filter will do this is this could be a filter. Can you find out output statistical property? That is the reason I am doing all this, you know making a kind of groundwork selectively. So, let us find out first what is y n convolution between x n and h n?

You have no convolution if no you have to read yourself, because this is not a DSP course; convolution say h r normally you can write either way h and we are can write x r h n minus r summed or this way both are same. Because and this is question of substitution variable, but always prefer this for most of your sums prefer this in this statistical signal processing not the other way. h r x n minus what is the more preferred version then x r h n minus r for simplification of calculations and ease in calculations nothing else. Choose this version that means, E of y n what is E?

E of a random variable means you have to multiply by the probability density integrate.
If the y n is a summation remember, then summation times a probability density integrate means you can apply the probability density function on each term under summation integrate. So that means, it is a linear thing you can go on applying that I mean taking each term multiply by the probability density and integrate so; that means, expected value of that term. So, E is a E is a linear operator E of say x plus y is E of x plus E of y. So, that thing if you apply here you can push the E operator inside the summation then it becomes h r is constant.

So, first I apply E over this product, but h r is not random. So, h of r can come out. See, E of two x is two of two into Ex like that. And what is this guy; this guy is a mean of x independent of a naught r; because of stationarity WSS nature of the input. So, this is what mu you can put a subscript x means of x so that means, what is the summation? So, this is independent of n constant times a summation; whatever will be the value that comes up. So that means, of any output sample at any n is a quantity independent of n.

So, it is stationary in mean and if mu x is zero that is input is a zero mean process output also zero mean.

Please remember all this. This will be like you know every time we will be using making use of this even without explicitly mentioning. So, these things should go down in your mind. These are some basic properties assuming the filtering of random process. Next, let us find out the autocorrelation of the output? Whenever, actually you have you know more than one random signal when I had only one random signal I used this notation r k is equal to say I said this is E of x n x star n minus k and all that I said. But suppose you are dealing with many many random signals, you know then to distinguish you put a subscript here r x x.

Why x x, because autocorrelation means product of one x with another x; so x x. In case of mean you have one x r x, similarly you can find out r y y and then there is something called cross relation r x y r y x. You know r in the case of r x y this is x this is y and vice versa; that cross correlation I am not going into now that is a very important topic, but not related here at the moment. So, I will not say I am not providing that statistical signal processing entire thing. I will I am only taking thing selectively which are useful for adaptive filter.

Suppose, I want to find out r y y k; so E it is pretty simple; y n now you replace by a convolution sum this also you replace by a convolution sum. You write h r again follow this form h r x n minus one not the other one not x r h n minus r this form I I am taking; h r x n minus r r is the variable and in this case first y n minus k n star. So, in this case h do not bring r here r has been used up in this summation, r is a local variable in this summation. Why use the same index? Because there will be a confusion you can use any other index.

So, use some other index say x star n minus k that is the index you have to find out y at n minus; forget the star for the time being. You have to find out y n minus k so it could be h n x into n minus r so n minus k minus m so the star has come. You know the star a b conjugate is a conjugate and b conjugate so this. And then what the meaning of this kind of summations is. You start with a particular r for that h r into x n n minus r; n is your choice from outside n is fixed. For that again you choose one m find out these another m;

find out this another find out this summation that entire summation multiplied by h r into x n n minus r.

Then, again take another r for that r you find out this product and then hold it. Again evaluate the sum for all m then multiply. Again take another r that is the meaning of this.
I remember that such double such process can be reversed also instead of first taking a particular r and then taking the entire summation over m and then taking another r again taking the this inter summation for all m. I can do the reverse. I can take one m and move over all r then another m again move over all r you will have the you will cover all the same points.

This is basic thing, two doubles summations can be interchanged and you can extend this here this is discrete summation; if one is integral or both are integral they also can be interchanged by the same logic. You follow this step this is basic step. Now, E will be working only on where not on h, but only on x; that means, this x and this x you put together E applied on this. So, it will become like summation of r minus infinity to infinity h r just a minute you have another summation m same range into E of this quantity.

Now, what is this this correlation autocorrelation of x, but what is the value? This index minus this index this minus this was that thing you remember. The lag the gap between the two time points this is one time point this is another time point and n n cancels that is the important thing. The moment n cancels the entire summation becomes independent of m, which means this r y y k this thing. In fact, I should not have written as r y y k the moment I wrote as r yy k it gave the implication that the entire thing depends only on k not on n.

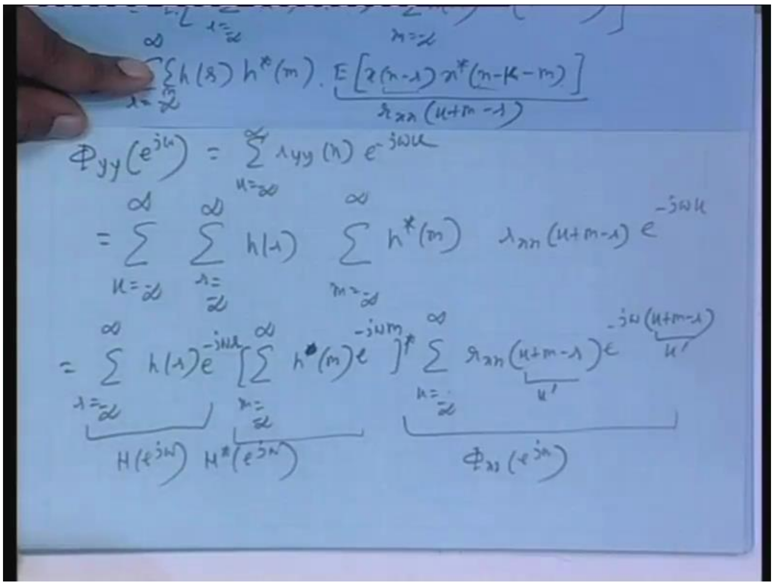
So, maybe I should not have written like that I should have started like this because without proving stationarity of output, I am using the notation that implies stationarity.
So, let us start with this thing I was trying to find out if this depends only on k and not on n if it is independent of n then stationary in correlation. Now, here the moment you subtract this from this, what do you get for this entire quantity what you get, r xx k plus m minus r not visible. You see this entire summation if it is r and m; so when you

completer this summation there is no r there is no m. So, entire thing is a function of k only.

This is h r h star m and k and m and r. So, when you complete your summation means just m and r are no longer present there are local variables. So, entire thing is a function of k only, which means output also autocorrelation also depends only on the lag or the gap variable that is k. So, I can write it as r y y k, which means output is stationary. What it means m equal to autocorrelation, which means output is WSS wide sense stationary.
Immediately, then the question comes it is WSS. So, its autocorrelation function is actually sequence of variable only k; if that is. So, what is that corresponding DTFT, because I can take remember one thing I can take power spectral density only for stationary processes; because only for stationary process autocorrelation function is a function of only one variable k.

So, it becomes a sequence of one variable and therefore, I can take this DTFT; if it is not stationary the question of DTFT has no meaning there. Power spectral density applies strictly to stationary processes WSS processes. So, now it is established that output also WSS so; obviously, I am be tempted to find out the output power spectral density to see what it is and I will try to relate output power spectral density with input power spectral density. So, starting from this expression I want to find out this quantity, you can see this is r y y k this summation this r y y k.

(Refer Slide Time: 30:03)



So, I want to find out p y y, earlier I did not use the index; I simply said phi e to the power j omega, but the movement you have got different processes; I told you that I will bring in the subscript either x x or y y not only for autocorrelation also for power spectral density. Because after all what is power spectral density it is coming for autocorrelation. So, if autocorrelation has x x this also should have x x if y y this also should have y y if x y; there is a cross power spectral density x y here.

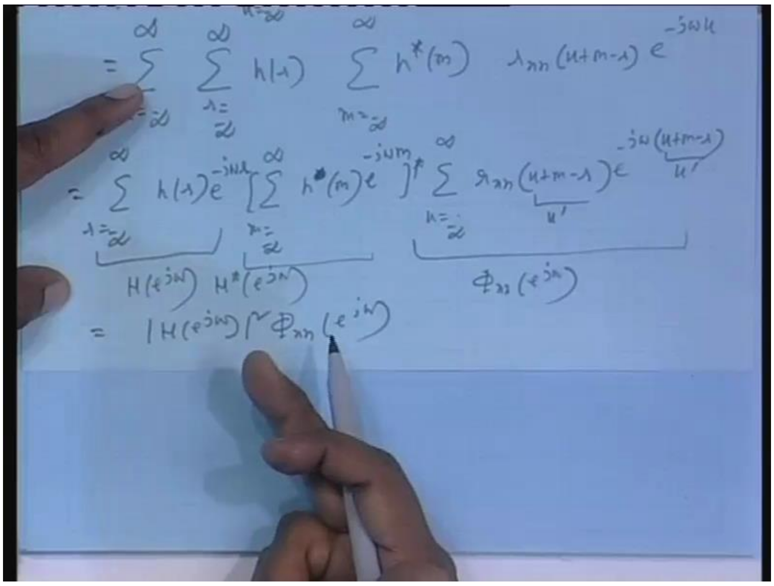
The phi y y e to the power j omega is this r y y k and r y y k this huge expression we have found out I have no other choice, but to have to replace this r y y k by this huge expression. So, it will become a triple summation; h r r x x and this and this goodwill fellow e to the power minus j omega k; up till this you are r y y k up till this by the formula r y y k this much only I have brought it down. The two summations that split h only had r. So, summation across r comes across I mean comes over h r h star m sum over m goes there is no point in having a sum over m before h of r; because h r does not depend on m.

So, I pushed that inside now; I tell you there is a rule of thumb in DSP whenever you come across double summation or triple summation the next step should be interchanging of the summation; if you have a double summation that is very simple just interchange the two again real will start following. But in this case there is triple some interchange will be required you bring in this one from outer to innermost. Because h r does not depend on k h star m does not depend on k. This fellow only depends on k r x x k plus m minus r. So, bring this summation over k from outermost to the innermost one. If you do that now e to the power j omega k just see this here please it was k, but suppose instead of k I call it k plus m minus r.

So, that this two and this two are same. So, I have to them cancelled. So, I should have e to the power plus j omega m that depends only on m. So, that I can put it here and another term will be e to the power minus j omega r. So, that depends only on r I can put that here. This quantity for your particular choice of r and n from outside this quantity then becomes depending only on k, you are moving k from minus infinity to infinity you can call this k prime. For any choice of r and m there is the r and m is fixed inside and then k is only varying from minus infinity to infinity.

So, k plus some constant you can call it k prime; so as k changes from minus infinity to infinity. So, does k prime for minus infinity to infinity; so r x x k prime e to the power minus j omega k prime. So, what is this quantity this is nothing, but phi xx e to the power j omega what is this quantity this is the DTFT of h n. So, this is capital H and what is this if you take the conjugate out, if you put it here and take this out and conjugate. So, this is H star.

(Refer Slide Time: 34:44)



That means, the entire thing everything as what mod; so mod H e to the power j omega and that is a real quantity. So, real and real output is real output based e; so you see this gives an idea about filtering. Suppose, input process was wide, you want the output random process to have its power if the frequency they were localized around some zones and not present in another zones. You can design a filter with appropriate transfer function and that will shape the output p h d that is the idea of filtering. You see the filtering here, is more general then the deterministic filtering, which you are familiar in normal signal processing courses; low pass filter high pass filter, because those are all meant for deterministic signal.

Deterministic signal means signal is either having this frequency or that frequency and so I remove this frequency and pass, but when it is random. It is fluctuating in that case;
I have to see power and always in a random from this stochastic point of view. This power spectral density not instantaneous power and find out average power how by integrating a quantity called power spectral density; there is the average power that has meaning when for random processes instantaneous power has no meaning. It is got a sin omega t as a sin omega t its r m s power is a square by two it is not. It is a random process it does have a deterministic power.

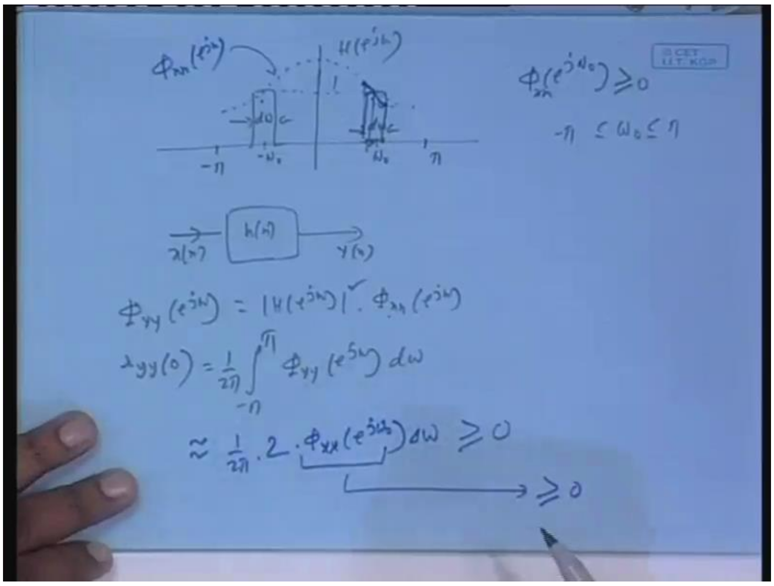
Every time the observation is changing; so only thing that matters to be is if it is stationary has been stationarity average power partial for any sample and that is integral of a quantity power spectral density that will give you; so distribution of the average power across various frequencies. See, what the output power to be localized in certain zone and gone and elsewhere; so you can boost up powering those components and elsewhere you can make zero. So, some kind of filtering will come up by you suppose e to the power j omega.

So, this is not designed with respect to a particular deterministic signal in mind with a random process in mind and it is average power in mind rather. Are you following? This the average power in mind not instantaneous power of a particular sinusoid or particular signal. Average power of a stationary random process; I want to wish to change its distribution in the frequency domain how that power is distributed. If a signal will be called low pass if that average power, when you see its power spectral density is localized the around origin average power.

Because average powers means what this power spectral density when integrate gives you average power. I am always repeating average power that has this kind of distribution, then it will be low pass random signal. So, any band pass random signal and high pass random signal and all that; no that has that would not have that meaning. You can have there is called time then you have to have. So, these are more generalized thing time varying the power spectral density you can assume that to be constant that. You can assume this autocorrelation function initially to be a function of n and k and that will not vary over n piecewise stationary kind of thing you know.

It will be a function two function of two variables, that you call time varying spectral density and all. So, that will not have this meaning. You can try yourself to apply the argument there it is not like an image signal that for an image signal. I agree it is a function of two variables and therefore, omega one omega two will come and all that count that thing will not come up that meaning will go signal is one dimensional. Anyway see one thing, then very quickly just for the case of say a real process; suppose I am dealing with a real valued process.

(Refer Slide Time: 39:20)



For a real valued process will be of power spectral density symmetric around origin. I am just proving here something in a very I mean using some simple steps, but to prove that reality in a more general case we will look at some further steps, which I will avoid here. Because I am trying to prove that power spectral density is non-negative. So, for that I am doing for the time being for real valued process for which PSDs are symmetric function of omega. So, suppose you want to prove you are in this zone from minus pi to pi you want to prove that phi e to the power j omega at any particular omega within the range. Say may be omega naught, where omega naught could be any frequency this is real we know.

We have to prove that this is this this for any omega naught is this is what we have to prove for any particular omega naught of your choice prove this. What I will do then this for the random process x phi x x you have to prove; what I do then you have you are having omega naught. So, around omega naught and over a slot infinitesimally thin of width d omega; I will construct a band pass filter suppose of height one. This is also d omega, call this system H e to the power j omega; this is a system linear time inverse.
Suppose your random process x n, I am passing through this process this system. So, output is y n output is y n.

So, what is output power spectral densities that will be mod H into this we have just now proved; we have just proved this mod transfer function square into this, but look at this filter even though they are appearing wide here they are infinitesimally thin. So, phi x x e to the power j and then what is then output power output power will be the integral of this, but integral means what this right hand side has to be integrated; right hand side is to be integrated from minus pi to pi. But you see even though omega varies from minus pi to pi; you have to consider only these zones around omega naught. There were infinitesimally thin strip of width we know d omega.

So, what is the area of suppose your phi e to the power j omega is something like this real; so it is symmetric. So, if you want to really multiply if you want to multiply phi x x. So, suppose this is I am just plotting suppose this is your phi x x e to the power j omega you have to multiply the two and then integrate over this range. What will you do? If you multiply this function with this, so this part and this constant height one will be multiplied. So, you basically get back this times d omega. So, area of this strip similarly on this side, but suppose this is very thin. If it is very thin then this will not be so long it will be so thin that this strip area will be approximated as the value of phi function at the center point say phi e to the power j omega naught into d omega.

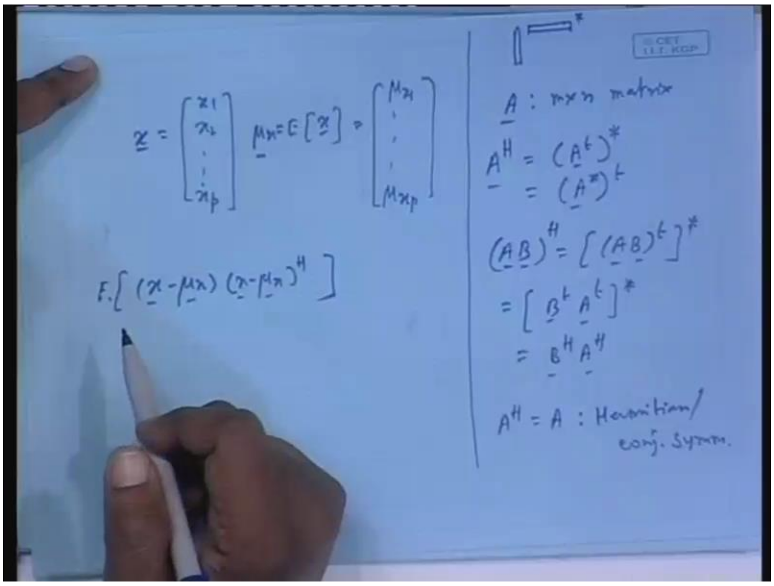
Very simple calculus because this is so thin here I have made it wide. So, you can see such a variation, but if it is infinitesimally thin area will be phi e to the power j center frequency omega naught times the d omega approximately. Here also because symmetric; so same value will come here; that is important same value phi e to the power j omega naught or phi e to the power minus j omega naught same that will come.

So, what will be the next the net thing that will be approximately one by two pi into twice because this and this twice phi x x not d omega say delta omega d omega is we used under integral and all delta omega fine.

The same thing because symmetric I took now, therefore, the height here and here will be same. So, I am proving giving this prove based on that real case only, but it can be generalized to a complex case that I will not do here. This is the quantity, but I have told you any variance is real and non-negative. So, this is greater than equal to zero. So obviously, this will be this quantity also is greater than equal to zero delta omega positive is just a constant. So, is do so is to pi, so this is the proof that power spectral density is indeed a real valued function.

I think I have got another ten or fifteen minutes. So, let us do some more serious business now. Suppose, I have got a set of random variables like yesterday I said x then x y then x y z then x y z dot dot dot dot single density joint density and all those things. Suppose, I have got a set of random variables x 1 2 x p one to p say they are random. It could be such that x one could be the n th sample of a random process x n x two could be x n minus one x three could be x n minus two so forth so and so as a special case.

(Refer Slide Time: 45:22)



In general, let there be a vector x my vectors are always underscored matrices are capital letter underscored. If you read books and papers they are not underscored they are bold faced, but I cannot go on spending you know making its bold that will terrible. So, I use this underscore thing. So, suppose I have got a set of random variables A in general complex valued; E of x is what is a vector; you can call it mu x vector that will consist of what E over this E over this dot dot dot E over this that is mu x 1 up to mu x p a mean vector. Then, you find out this if you take this vector and take its conjugate transpose. What is conjugate transpose? Transpose and conjugate each element that is called Hermitian transposes.

So, for your sake let me, I am just it seem you do not to Hermitian transposition. So, just couple of minutes I am showing something. Suppose in general, matrix A m cross n

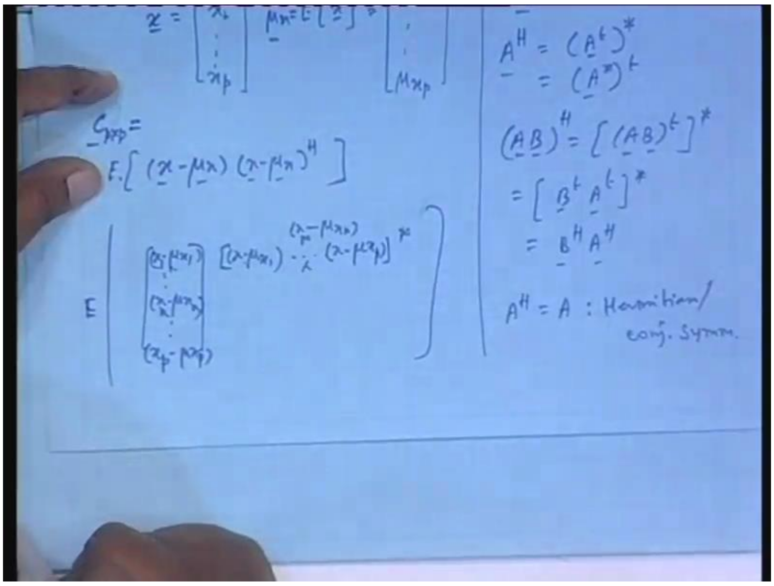
matrix complex valued; then A we say Hermitian. A Hermitian means you either you can take transpose of these complex the conjugate or by service; A conjugate transpose either way you will get the same thing, this is called Hermitian transposition. For real valued matrix Hermitian transposition and simple transposition they mean same. Another thing, you know this if you have A into B, suppose A and B are such matrix they can be multiplied with each other; if it is m cross n and this is n cross p something like that they are comfortable for product.
Suppose you want to do this multiply and then take H; that means, you can say AB first transpose then conjugate; I should put underscore everywhere. Then B transpose A transpose conjugate, but conjugate of a product you can suppose this two matrixes are multiplying. So, terms are multiplied and then added conjugate of; that means, you can apply the conjugate on each term and all that you can do this. So, this is nothing, but B transpose conjugate times A transpose conjugate, which is B Hermitian A Hermitian. So, what you know for transposition the AB transpose is B transpose into A transpose that gets easily generalized to I mean Hermitian transposition and a matrix A it has to be a square matrix actually.

If it so happens that A Hermitian is A; obviously, it has to be a square matrix m cross n becomes n cross m. So, they can be same if n and m are same. You transpose and conjugate and if this turns out to be this this is called Hermitian matrix or conjugate symmetric matrix, for the real case it is simply symmetric matrix. You know if you take a square matrix transpose and if you get the same thing it can be it has to be a symmetric matrix only. Do you know or not? A i j after transposition becomes A j i; if A j i is get originally at the A i j goes to j i th position, where A j i is located, but if you still get back to same guy A j i then A i j becomes A j i. A i j you are transferring to the j comma i th position and originally that was that was A j i.

So, if these two things are same that is after transposing you find that you are hitting by the same quantity there is no change; that means, A j i and A i j are same that matrix is called symmetric, which is real Hermitian; the real Hermitian it is called Hermitian matrix. This is the key of this course Hermitian matrix, because plenty of properties I was to begin this properties of this matrix Hermitian matrix. We will deal with what is called autocorrelation matrices, which are Hermitian it has got beautiful properties those properties have to establish, and then I will go into this further Hermitian.

That is conjugates or conjugate symmetric; if that in a real case it is simply symmetric this for this. Now, suppose I take this difference vector x minus mu x, you remember what was covariance two variables were there x and y. So, you take x minus mu x y minus mu y multiplied x minus mu x times y minus mu y star there are expected; suppose, I take this vector and take its Hermitian transposition. Firstly, this is a column vector; first one is a column vector; second one is a transpose conjugate. So, it is a row vector and conjugate. It is shape will be like this column vector followed by a row vector with conjugate there is a same. So, it will be a matrix of size p cross p and we apply E over it.

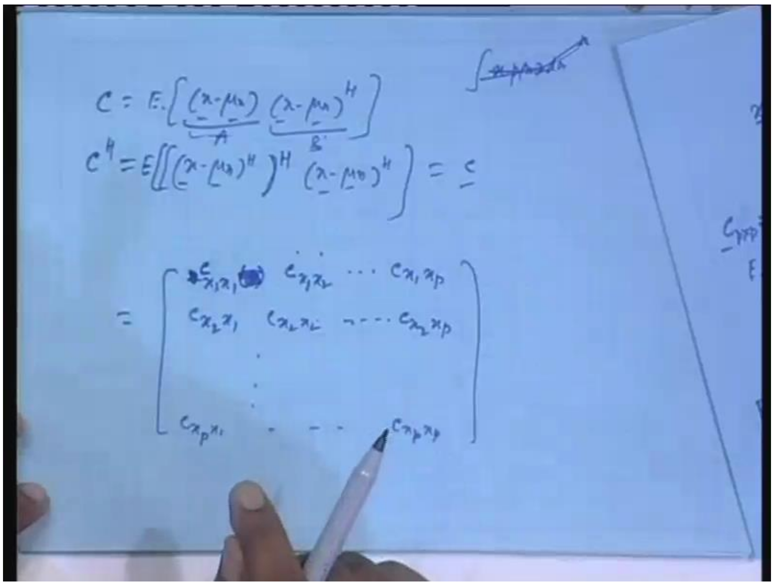
(Refer Slide Time: 51:37)



This is my claim this is for the covariance matrix or sometimes we call it auto-covariance matrix covariance matrix. C matrix C and you can write C p cross p just to indicate the size. Why it is auto-covariance? What is the nature of this matrix? You see you take this terms first one is dot dot dot x minus mu x say this is x one say x k minus y mu x k dot dot dot x p minus mu x p. And this side is conjugate of that and one general term, you can say x m minus mu x m star. So, see what is happening? You know this column vector into row vector product this term will multiply this.

Then the second term third term all of them E over each of them that we call the first row. You know this is very simple you know this or not you know this. This will multiply each of the term and separate into the first value formed, what is the first term.? This into itself variance this into itself, but this into itself means this into itself conjugate there is a conjugate here. So, x minus mu x one minus mu x one mod square E over that E is coming from behind, so that will be variance what variance.

(Refer Slide Time: 53:31)



That will be your r x 1 x 1 zero, then x 1 minus mu x 1 into x 2 minus mu x 2 star that will be what covariance between x 1 and x 2. Yesterday, I told between x and y there is no x and y here x 1 minus mu x 1 into next term is x 2 minus mu x 2 star. So, that is at the E over that there is a covariance between x 1 and x 2. I will not call it is let me call it C. C x 1 C x 1 x 1 is actually a variance with itself taken out the mean mod square of that expected value is a variance. And then you have got C actually this bracket should go, no point in putting a bracket here I was mixing up between x 1 and x 1.

So, already lag is zero one to one is one, so no point in bringing a zero there. Then, C x 1 x 2 dot dot dot dot C x 1 x p. You see these elements then again next what will be C then C x 2 x 2 dot dot dot C x p x 1 dot dot dot C x p x p. You understand this matrix is what is diagonal into variance is. So, real non-negative other elements are complex. But remember one thing, C x 1 x 2 and x 2 x 1 they are conjugate of each other. Covariance between x and y or y and x are the conjugate of each other covariance between x and y; I told you order is important just I told you.

E of x minus mu x times y minus mu y star that is between x and y; I mean y and x E of y minus mu y into x minus mu x star. This is the conjugate of the other one you see that is that what you are having here. In one case this quantity into x 2 minus mu x two star another case x two minus mu x two into this quantity star. So that means, this conjugate this i j th element say one one two th element two one th element they are conjugate of each other one and three element three and one element conjugate of each other like that so and on so forth. In general, this is the conjugate symmetric matrix; that means, this is the covariance matrix is a Hermitian matrix that you have to you do not have to show really taking.

You know this thing that take x i and x j c x i x j that is in the i j th position and j i th position means C x j x i and they are conjugate of each other that we can show, but very smartly you can show this way that what is C after all E into. What is C H? Each means this matrix is to be transposed Hermitian transposed conjugate transposed, but E over this you take E of a matrix then do conjugate transpose or better do conjugate transpose, then E operator we will get the same value E will not give E operator will not give defined values. So, you can apply the Hermitian operation inside this inside the E operator.

You understand this; you are applying E operator on a matrix, but this consists of random thing. You find out expected value then take the Hermitian transpose or better just take the Hermitian transpose then apply E you will get the same thing. See, E of x and then conjugate is same as E of x conjugate because E of x means x into p x conjugate means p x is real I told all this yesterday; x p x dx is the E of x the basic physics is x p x dx. Say, this is E of x if want to conjugate this this kind of thing; you have to conjugate this conjugate will come only on x star, which is same as saying that I am taking the expectable value of x star. So, expected value of x star or star of expected value of x both is same.

So, that will be also applicable there, because this is a conjugate symmetric thing; so here instead of taking Hermitian over this entire expected value of the matrix. Hermitian means transpose and conjugate you first transpose interchange the positions and conjugate and then take expectation you will still get the same thing. I told you expected value of conjugate or conjugate of expected value are same. So that means, H operator can go inside and H of any product of I told you A B there are two matrices not necessary square, any general matrices ABH is BH AH.

So that means, this into this H this H over that H that will come; this H will come first this is B this is A. This H over that another H is you following me and this guy H. So, you get the same thing, you get back C. So, C is a Hermitian matrix remembers this is auto covariance and particular case where all the means are zero, then it will be autocorrelation matrix also. So, then correlation covariance are same. So, we stop here today. So, Thursday morning we will continue from here; this matrix has got beautiful properties, we have to discuss because they are useful that is all.

Thank you.