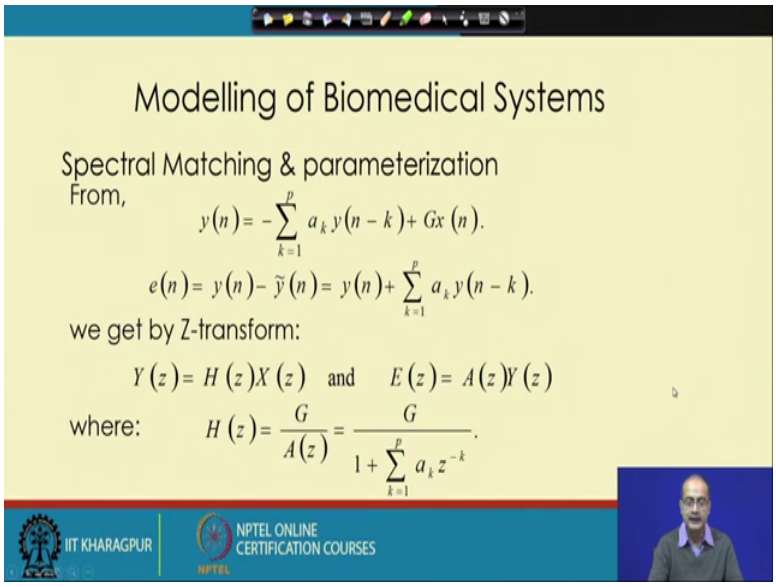
Biomedical Signal Processing
Prof. Sudipta Mukhopadhyay
Department of Electrical and Electronics Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 44
Modelling of Biomedical Systems (Contd.)

So, now let us look at that parametric model from the spectral domain or the frequency domain.

(Refer Slide Time: 00:27)



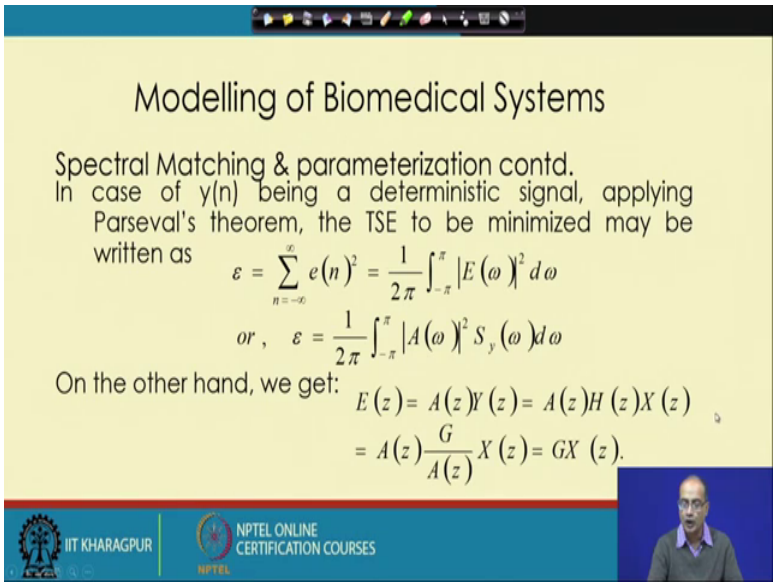
So, what we get here that from the equation of the AR model and that the error term that with the help of which we are actually extracting the parameters. The first one is the synthesis model that if we drive it by a white noise and we know the AR parameters how we can get a signal like yn.

The next one is a analysis model, that where we are trying to extract the model parameter because we do not know that driving sequence and the coefficient that is aks. So, given the signal y we are trying to do the linear prediction, given the fact that for the timing we have assumed that we know that model order and one way to make that that we have taken a sufficiently large model order. So, we can compute the that prediction y tilde n and we can compute the that prediction error.

So, if we take the Z-transform in both the case of the equations we get one is for Y z in

terms of H z and X z, another is for the error term E we get E z terms of A z and Y z we get they have some relationship, because A z can be represented as that G by A z ok. So, these two filters that H z and A z they are inversely proportional to each other and they have a relationship in that way.

(Refer Slide Time: 02:40)



Now, if we take Y n we have actually captured that signal that output, if we treat that as a deterministic signal because we have just one set in hand even if it is a stochastic signal or a stochastic process we do not have multiple that the shapes of Y n. So, we can treat that as a deterministic signal and using the Parseval's theorem, what we can get the total square error it should same in the time domain as well as the frequency domain.

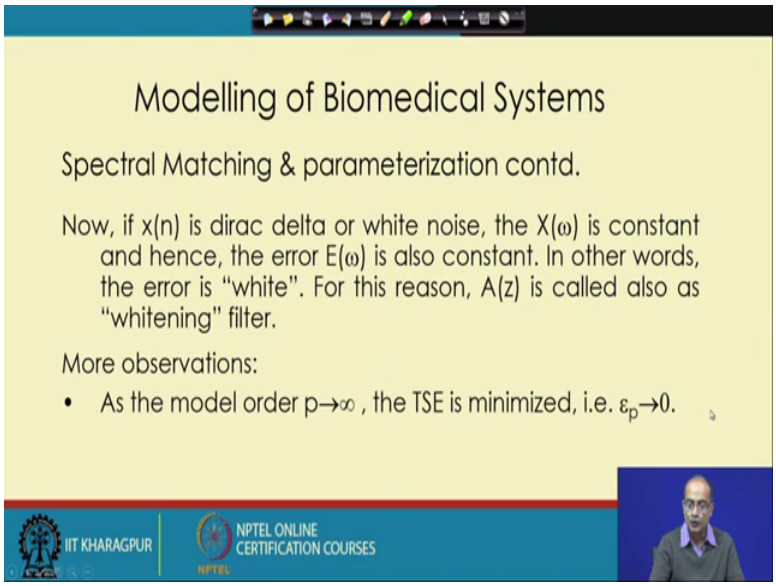
So, we can write eta that energy of the error in the time domain that is same as the energy of the signal in the frequency domain and from there E z or e omega to be more precise in this case, where z is taken a point on the unit circle. So, we can represent that in terms of A and Y.

So, it can be represented E z square or the spectrum of the error in terms of the square of the that polynomial or filter A omega, and the autocorrelation of the output S y or spectrum of the output to be more precise. So, it is a convolution of that on the other hand what we get we can do some mathematical manipulation as E z equal to A z into Y z.

Now, we can replace the value of Y z we know that the value H z into X z and H z can be represented as G by A z. So, we get from the right hand side and A z will cancel because it appearing both in numerator and denominator. So, we are left with G into X z.

So, the prediction error is a scaled version of the that input signal.

(Refer Slide Time: 05:20)

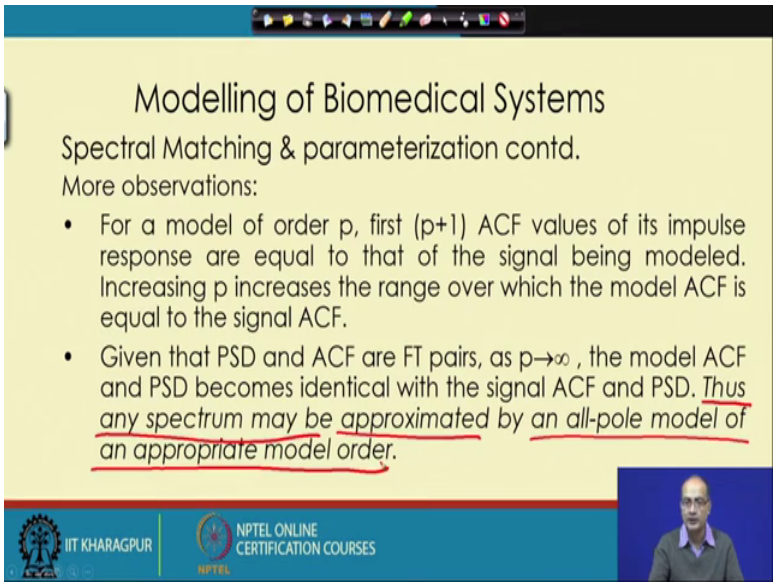


Now, in this previous analysis what we have seen that X n is either that X n is the driving force it is either is a impulse function or direct delta function or a white noise. And for that X omega is constant. Hence, as the E omega is also a scaled version of X omega it will also have the same characteristics; that mean, it should be constant over any frequency.

So, in other word what we can tell the error is also white and for that reason that the filter A z, that what we are determining here using the autocorrelation method or auto covariance method, we call it as a whitening filter, because the output or the prediction error what we get that gives us a wide spectrum or white noise ok, that is why it is called that whitening filter.

Here some more observation as the model order increases the total squared error that is eta p it actually decreases eta p tends to 0.

(Refer Slide Time: 07:09)



So, what we can conclude from there a multiple things we can get the first thing for a model order p, if you look at this model we are making use of p plus 1 autocorrelation values to compute the model. So, the model impulse response has actually same ACF as that of first p plus 1 ACF of the signal and as the p increases, which means the model ACF becomes more and more equal to the signal ACF.

First we are matching with p plus one ACF starting from 0 and as the order is increasing the ACF of the filter impulse response and that of signal that is actually we are finding a match we are making use of them. So, in that way increasing the model order we are getting the same value for the filter ACF as well as for our that model ACF this.

Now, ACF and PSD they are just Fourier transform pairs. So, they are if we can catch one we get other also. So, as the p increase increases and it tends to infinity what we can tell the model ACF it becomes identical to that of the signal ACF and model PSD would be identical with the signal PSD.

The simple reason is we are making use of more and more lags ACF with the increase in the model order. So, we get a identical matching between them. So, in other word we can what we can infer that given any spectra, we can actually approximated by an all-pole model of an appropriate model order; that means, if we do not restrict that model order.

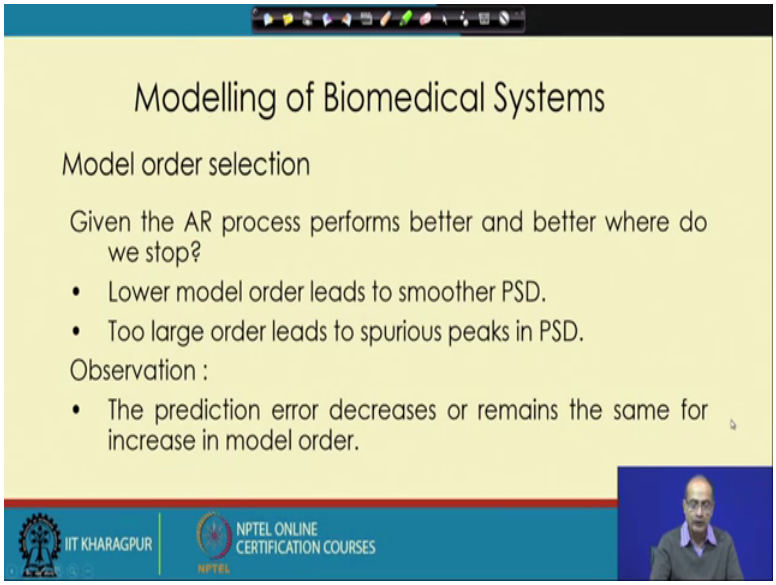
Any model order is fine then given any spectra we can actually model by all-pole model;

previously we have told in case of biomedical signal AR model is good, but there may be case that where the ma model would have been more appropriate we may not encounter such situation that is why we do not know or ARMA model should be a better model in terms of a compact representation.

But we started with an AR model because of our lack of knowledge. So, what would be the situation in that case that is an important question, because what we have told that parametric model is very powerful. So, long the assumptions are correct if the assumptions are wrong it can actually lead us to very dangerous situation. So, we bound to check that what is the penalty we can have to pay by taking only the AR model in case of a more general model like ARMA model.

So, what we see that if we are ready to increase that model order and give a flexibility in that given any signal or to be more precise any signal spectra, we can represented it by the AR model of a sufficiently large model order ok. So, that is the importance of this derivation.

(Refer Slide Time: 12:15)



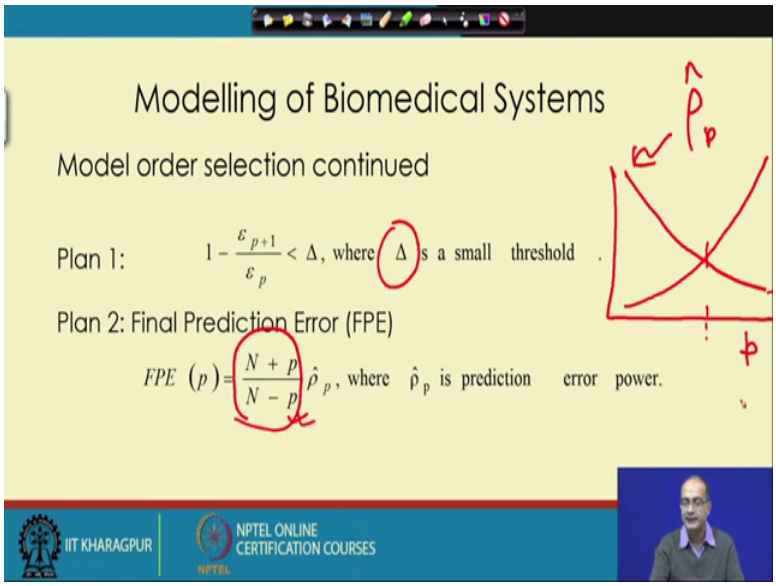
So, now we come to another important question that how do we select the model order? Given the AR model performs better and better where we stop, we can put it also in that way because we know as we keep on increasing the model order say using the Levinson Durbin algorithm or any other technique we see the error becomes less and in that case where we can tell that the right time to stop to avoid over fitting.

The first thing to note here lower the model order leads to a smoother PSD we get a smooth PSD, but if it is not again appropriate them by over smoothing we may loose some actually prominent peaks. Now if we keep on increasing the model order what will happen? We would get number of actually spurious peaks, which will be more specific to that actually that n symbol; that means we have taken a set or a vector Y n or a set of values of Y n. If we take another set of observations of the same stochastic process, we will find the spectrum will change those peaks will change their location and that is why we call them as spurious peaks ok.

So, increasing the model order what it is doing it is trying to do over fitting in that signal though it is reducing the error and matching more and more to that given signal, it is not actually giving as more new information about the underlying process rather it is suffering from over fitting.

So, we need to find out an way the first observation what we get that as we increase that model order there will always be it decrease in error or error will remain same. So, we can say error is actually a non-decreasing function or rather non-increasing function of the model order either it will decrease or it will remain constant.

(Refer Slide Time: 15:13)



So, based on that fact all the model of selection techniques have been proposed the first one is very simple based on that fact, that it will remain same or it will reduce with a increase of the model order the prediction error.

So, the plan one is let us take the ratio of the error for order p plus 1, divided by the previous actually error for model order p and this ratio is expected to be less than 1 or very close to 1. So, we take the difference from 1 now when this value this ratio comes to very close to one determined by a threshold small threshold delta, when it is coming closer than that then we should stop. So, that is the first simple technique.

The next one is using again the final prediction error the plan 2 it is suggests that we should take again the prediction error and we take a multiplayer along with that. So, if we look at this term N plus P by N minus P we get the numerator is monotonically increasing or linearly increasing the denominator is again linearly decreasing.

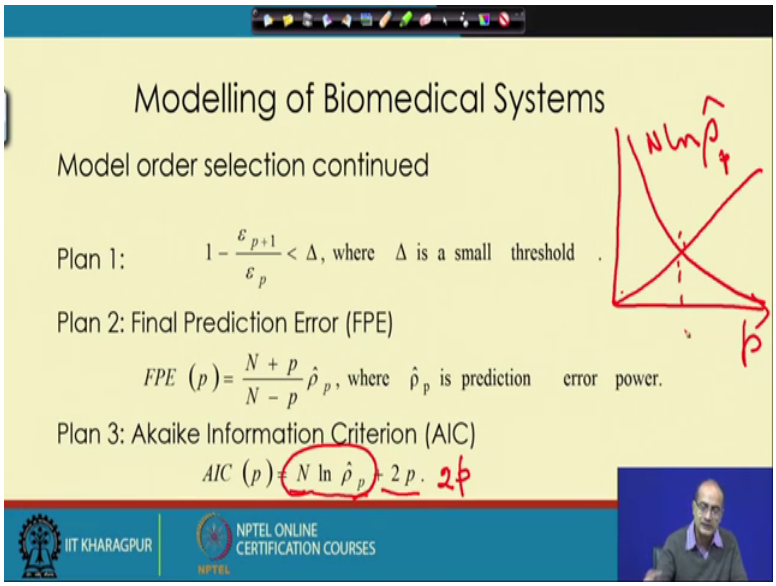
And initially as the N is very large that change will not be prominent, but as the P is growing it will become more and more prominent and 2 together it will give a non-linear increase on the other hand the prediction error that is rho hat P for the model order p what we get if we draw the curve that with respect to p, that the prediction error is actually decreasing in this way this is the prediction order rho hat P and the first term it is increasing in this way.

So, the minimum will come at certain point in other words actually what we are doing here we know the final prediction error it will decrease very slowly and to come to a constant value it may take very large model order. It can keep on actually reducing in very small way and in that case to find out that threshold delta is a difficult choice, because that depends on the signal and that depends on the process in hand.

So, what we should do we need to add some penalty for the increase of the model order. So, for that we have taken the correction term here, which is increasing with that model order. So, that as the p is increasing that penalty is also increasing and we stop at some point when the 2 curves are cross each other or restrict from a from the over fitting.

So, both these techniques are actually empirical techniques and they are have actually less theoretical background.

(Refer Slide Time: 19:52)

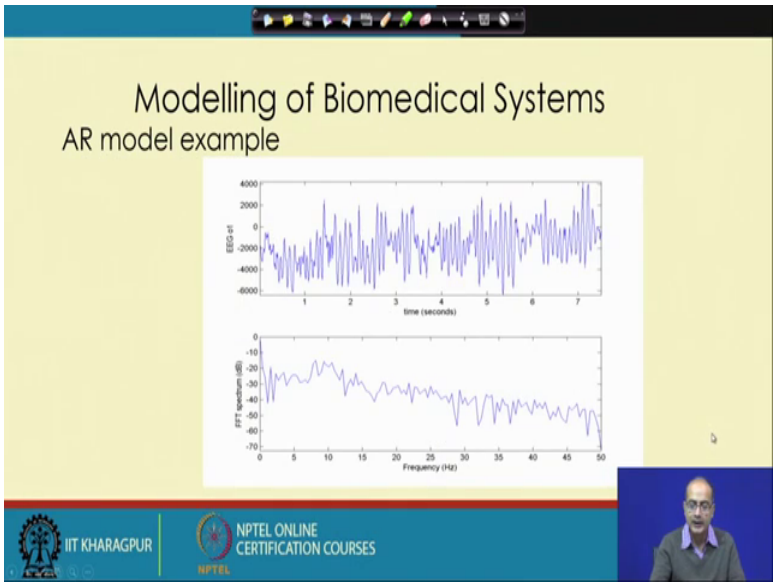


So, there has been a proposition based on the information theory proposed by Akaike. So, it is known as Akaike information criteria which has given a technique to select the Akaike information way that it depends on again the prediction error power we take the log of that and multiply with N. So, these term that will go that will in k actually decrease with the model order and the second term will linearly increase. In fact, for the theoretical derivation Akaike has found that this term should be p.

But when people try to use that they found that if we use a scaled version of it that is if you use 2 p we get more accurate values. So, that is what is used in practice. So, in this case what will get that again the curves would be like this with the increase in the model order that, the first term will decrease that is N log of rho P the other one is 2 p is increasing linearly.

So, we will get the point where the minimum will get where the 2 intersect with each other. So, that will give us the optimal model order ok. So, that is about the model order selection. So, we can use any of this technique and out of them the last one is the theoretically more sound and more accurate one though it is a bit more complex.

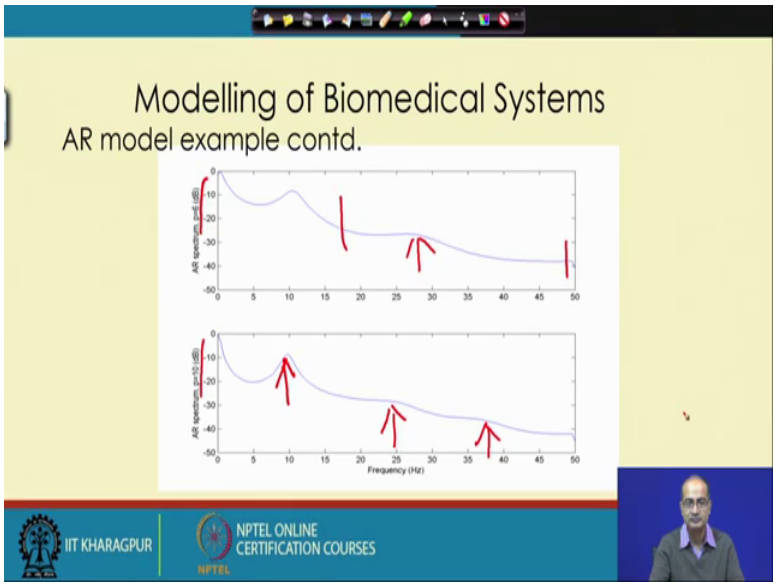
(Refer Slide Time: 22:10)



So, here first we see some example that what is the difference we get when we take the FFT and directly FFT that is periodogram technique or if we go for the that AR model to compute the spectra.

At the top here we are showing the signal at hand that is an eg of channel O 1 the corresponding that FFT spectra or periodogram spectrum is given here and we know that the periodogram; it gives us a more clean view than the autocorrelation function to check that where the signal energy is more concentrated, that ACF also has same amount of information, but the representation is not. So, clear or direct.

(Refer Slide Time: 23:27)



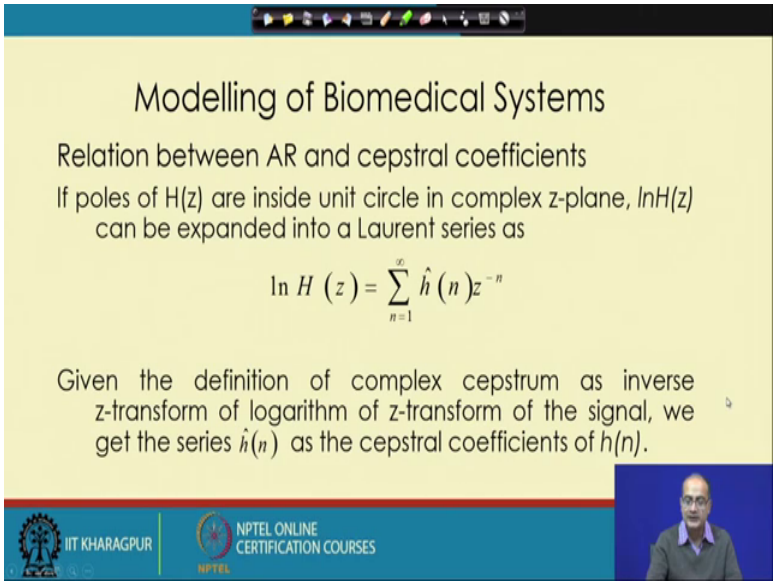
Now, let us look at if we compute the same using the AR coefficient what we get? Using the model order 6 here we have taken AR model we get a very clean peak very near to 10 hertz; that means, alpha components are there you get it very clearly. And it is much more prominent than what we could get in the periodogram ok. We can check that in the previous that FFT base spectrum if you compute here also we are getting that more accumulation at this place, but the peak is not clear there are multiple peaks that makes a difficult to find out that where is the location of the peak and to decide on it whereas, AR model makes it very clear.

Now, if we increase that model order instead of taking the model order S 6, if you go for model order 10 what we get this peak that is near 10, that remains there it is not disturbed, but it is a big accentuated it is becoming little more sharp. At the same time in the valley region that is this part earlier there was a very subdued one single peak here.

Now, instead of that we are getting 2 peaks in the valley region we are getting actually rather than peaks we can say more undulations we are getting. So, that is the difference it creates if we use the increase model order. So, the change is not much. So, penalty is not much that big in terms of the spectra, but suddenly it will increase the computation. So, we always prefer to use the that optimal model order which will give us the smooth spectra and we can reduce unnecessary computation which does not add any value for the signal analysis ok.

So, this is about we get that idea that how the model order selection affects and we get the past and also the idea that how the spectrum, computed by the parametric model is superior than the PSD technique, given that we have taken as a appropriate model order. If by mistake we have just stopped at 2 then the situation could have been different it would not it may not resemble actually the signal spectra.

(Refer Slide Time: 27:03)



So, now we will look at one more thing that we have seen for the speech signal the cepstral coefficients are very good as features. So, here we see the relationship between the AR model and the spectral coefficient because with a ACF we could get some relation with the spectral coefficient, we expect the AR model also may have some relationship or the expectation can be increased like the autocorrelation helped us to autocorrelation of the signal helped us to get the cepstral coefficient same way can the AR coefficient help us to compute the cepstral coefficient, because it is helping us to compute the spectrum of the signal ok.

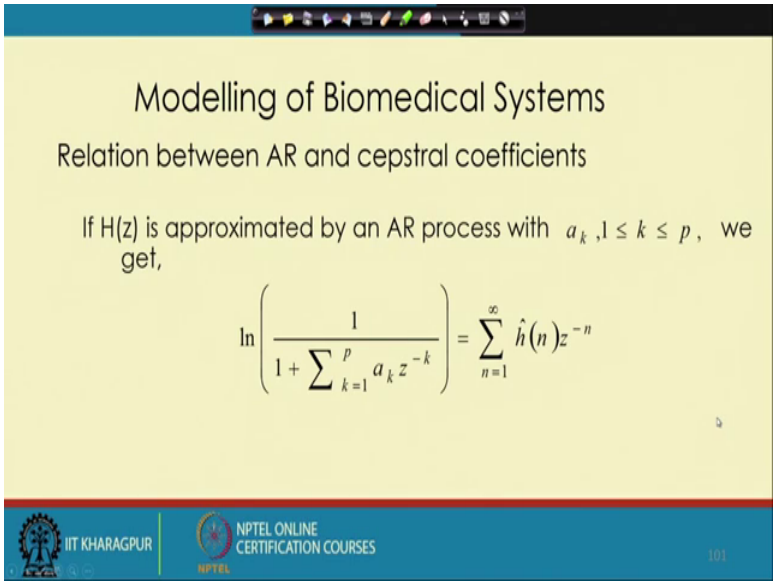
So, from that what we do first we take the that all pole model H z, where the poles are located in the units circle and using the Laurent series we can actually represented as a power series, and if we take log in both the side which is one of the things that is required for the cepstral coefficient or to go to the cepstral domain we need to take the log.

So, if we take that we get in the right hand side sum h hat n ok. As the coefficients of the

different order of z to the power minus n, now using the definition of the complex cepstrum as the inverse z transform of the log of z transform the single. So, what we need to do we need to take actually inverse z transform of the log of H z. So, we get the series the h hat n as the cepstral coefficient of the signal h n ok.

So, we could get actually cepstral coefficient in this way.

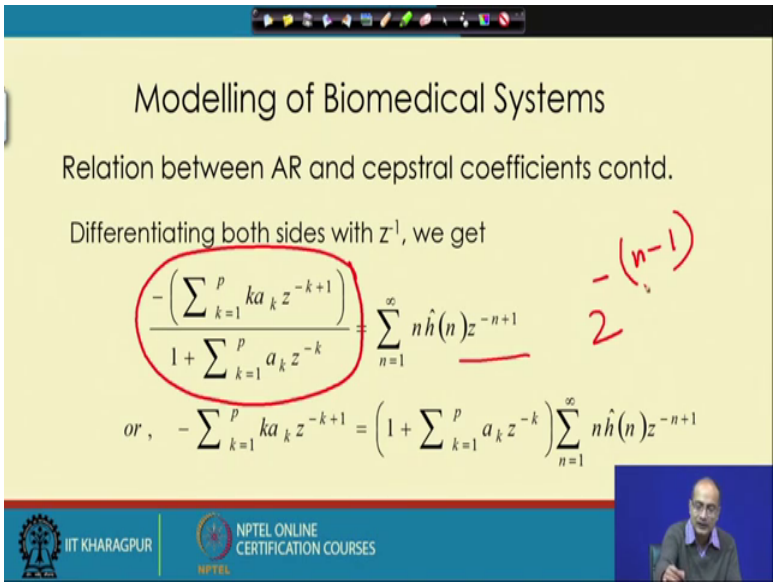
(Refer Slide Time: 29:35)



And let us see that if we take the AR model represented by the model coefficients aks for k equal to 1 to p what is the relationship between them?

So, for that we replace hz in the left hand side that log of hz we have taken we replace hz by 1 by a z and after that we can differentiate both the side by z to the power minus 1.

(Refer Slide Time: 30:10)

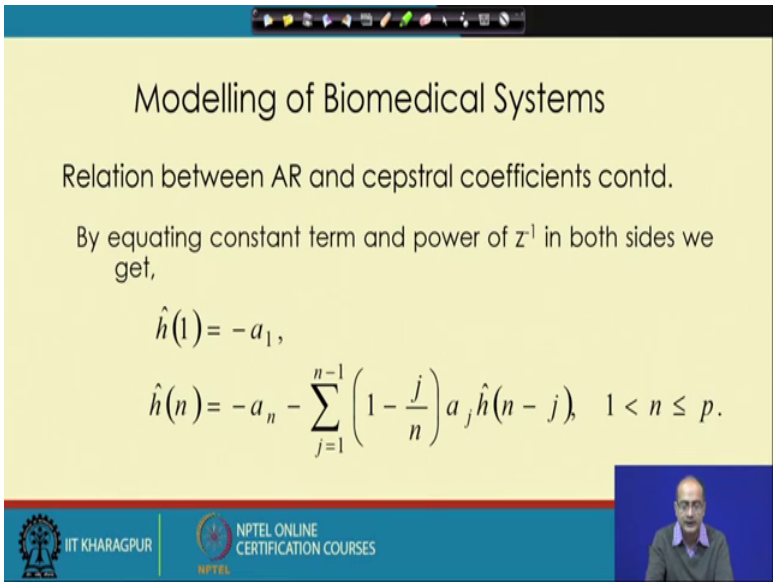


Because the polynomials are we are getting the power of z to the power minus one please keep in mind it is not with respect to z, but z to the power minus 1, because that helps us to do the computation. So, left hand side what we get the log of x is 1 by x and then we get x to the power minus 1. So, we will get x to the power minus 2. So, by cancelling that power we get it that using the chain rule we get the derivative in this form we get here below a z and we get that after taking the partial derivative this is the term we get.

In the right hand side we get n times h hat n as a coefficients of z to the power minus n plus 1 or we can write it in this way z to the power minus within bracket n minus 1 ok. And to make it actually simple to simplify it we can multiply both the side with the denominator of the left hand side. So, we get both the side polynomials and in the left hand side we get the power of z starting from z to the power 0 to z to the power 1 minus p, that is the power of z in the left hand side right hand side.

However, it could be infinite, but because both the sides it need to match with each other. So, the coefficients have to match coefficients of z or z to the power minus one need to match in both the sides.

(Refer Slide Time: 32:30)



So, using that property what we can write that h hat 1 equal to minus a 1 and h hat n for n equal to 1 2 greater than 1 to p; that means, starting from 2 to p we get it is minus an minus some more term. So, as we compute h 1 using that we can compute that h 1 hat with that value we can compute h hat 2 and using h hat 1 and h hat 2 we can compute h hat 3.

So, in that way we can go up to the pth coefficients. So, that the p spectral coefficients we can compute in that way ok. And according to the AR model the model order is p here in this case. So, those many cepstral coefficients are only important and we can compute them using the AR coefficients.

So, if we have AR coefficients we need not have to do the z transform Fourier transform taking the so, many things taking log taking care of the that that phase unwrapping. We can directly use the AR coefficients and get the cepstral coefficients from them. So, that is another important result we could get by learning the AR coefficient.

Thank you.