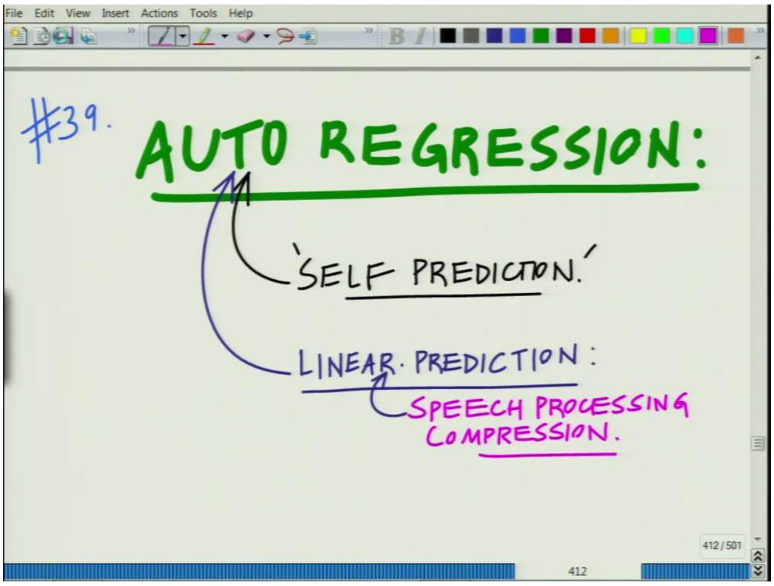
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Lecture 39
Time-series Prediction via auto-regression (AR) model

Hello, welcome to another module in this massive open online course. So, let us continue our discussion on optimal estimation and let us look at another interesting extension of this and this is to the concept of, what we call a self prediction or what essentially auto-regression.

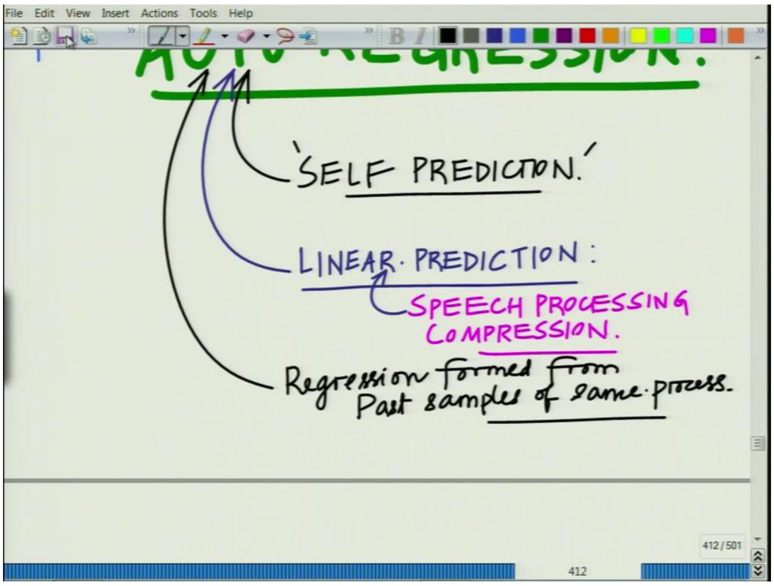
(Refer Slide Time: 00:35)



So, I am going to explain more about this, what we want to discuss or talk about is essentially the concept of auto-regression as you are familiar regression essentially is prediction. So, auto-regression, auto is essentially self prediction using the quantities in fact the past values of the quantities as we are going to see, this auto means the self itself and regression is essentially prediction or approximation and this belongs to an important class of techniques that is essentially your linear prediction so, this essentially is based on the theory of linear prediction or linear estimation.

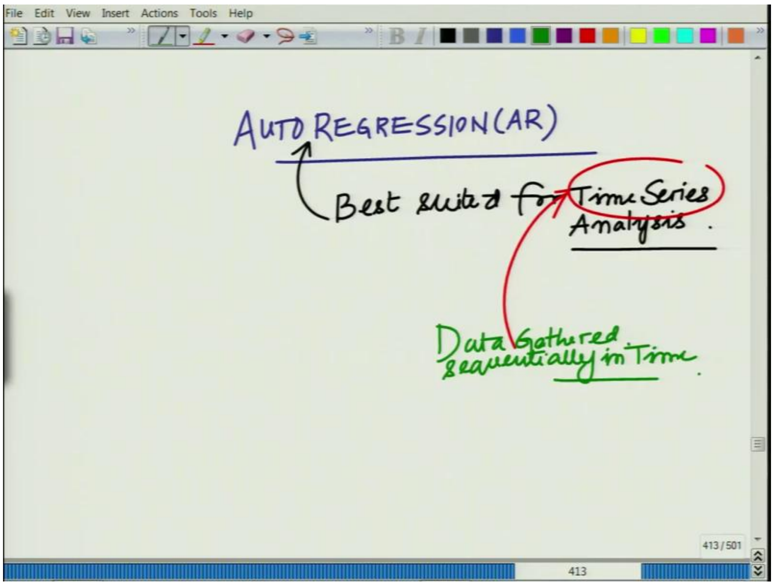
Which is essentially also used in linear predictive coding, which is again used in speech processing and so on so, there are many applications for this which, processing is also used in for instance compression so on and so forth so, what is auto regression?

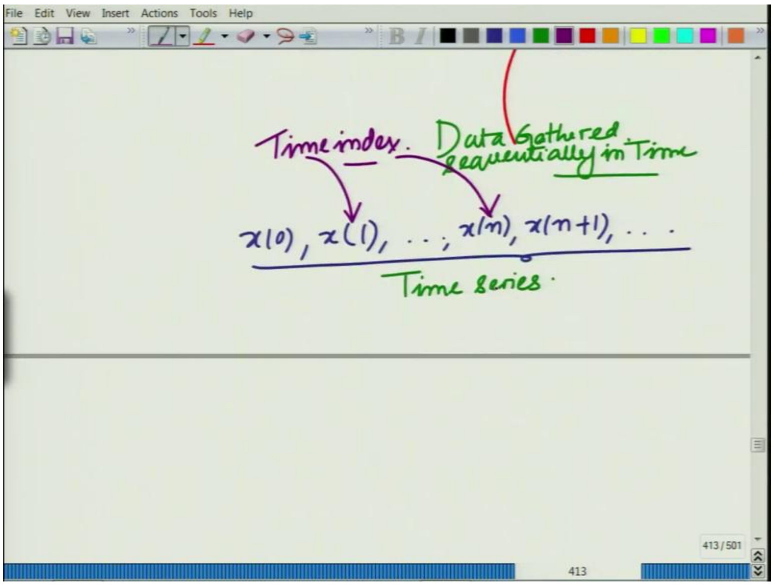
(Refer Slide Time: 02:16)



As I have already said auto-regression is essentially the regression formed from the past sample of same process and where you see this most often is in the context of a time-series that is data gathered sequentially in time those linear prediction is best suited for a time-series or the analysis of time-series.

(Refer Slide Time: 03:05)



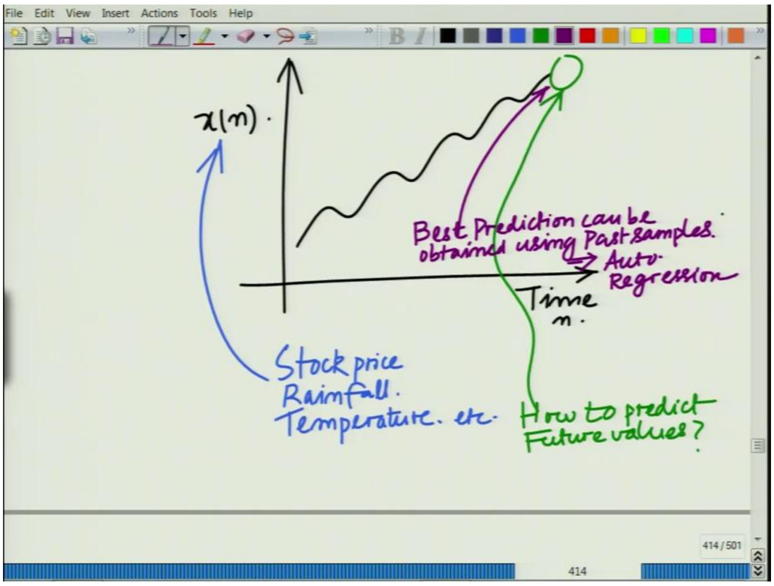


So, auto-regression what, we call as AR auto-regression this is best suited for time-series analysis and what is time-series?

This time-series is nothing but, data gathered sequentially in time, so time-series is basically data that is gathered sequential in time for instance you have the data x0, x1, xn, so on and so forth, so this forms your time-series and this is your time index and this for instance this n, this is essentially the time index.

And, so you have time-series and essentially regression that is formed using the past samples of this time-series, that is essentially auto-regression that is prediction using the past samples where is this useful?

(Refer Slide Time: 05:18)

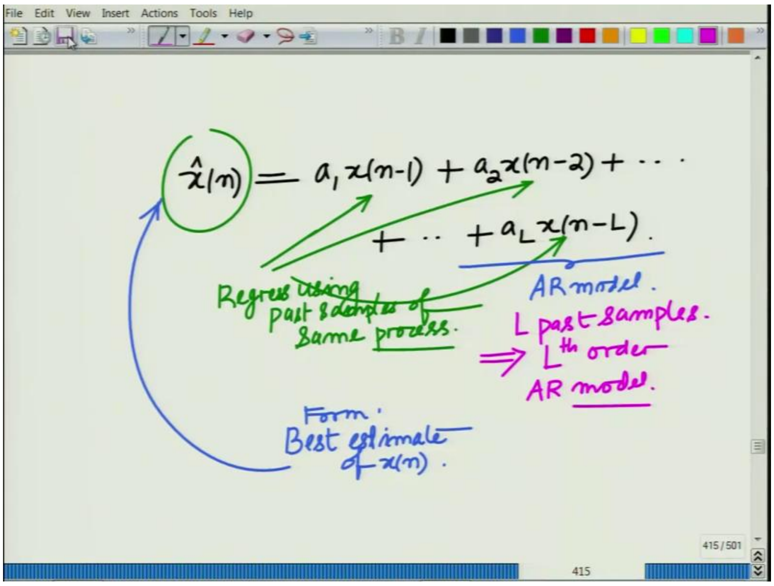


Now, this time-series typically occurs many places such as for instance you have some quantity that is varying with respect to time and this is your x of n and this can arise for instance in your stock price or the amount of rainfall or parameters such as, temperature, etc.

So, these that is when you measure these things, that is the stock price day after day, when you tabulate the stock prices list the stock prices day after day as a series day 0, day 1, day 2 so on so, that becomes a time-series similarly, the amount of rainfall the temperature on a particular day.

So, it occurs very, very frequently that is wherever you are taking samples of the quantity in time. And naturally the question that we are interested in is at this point, if you are at this is the current point, how to predict the future values of the time-series the point is you might have already guessed the best prediction of the future can be obtained using the past samples or the immediate past samples and that is essentially what we would like to regress on that is regress on the past samples and that makes it the auto-regression. So, the best prediction can be obtained using past samples, this implies auto-regression. So, what is auto regression?

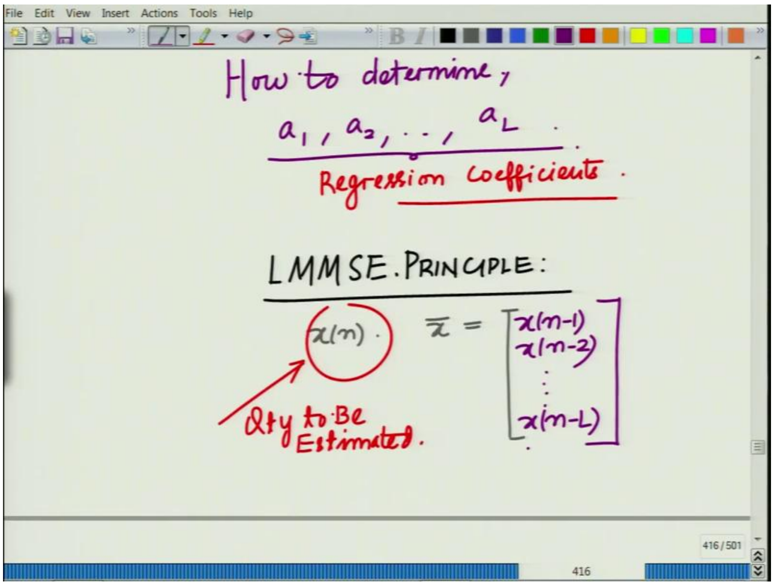
(Refer Slide Time: 07:54)



Auto-regression is essentially you form the prediction that is x hat n using the past samples regress based on past samples of the same process. So, this is if you look at this what we are doing is we are regressing, we regress using the past samples of the same process and we form the best estimate of x n, form best, obtain the best estimate of x n and this is essentially your autoregressive model.

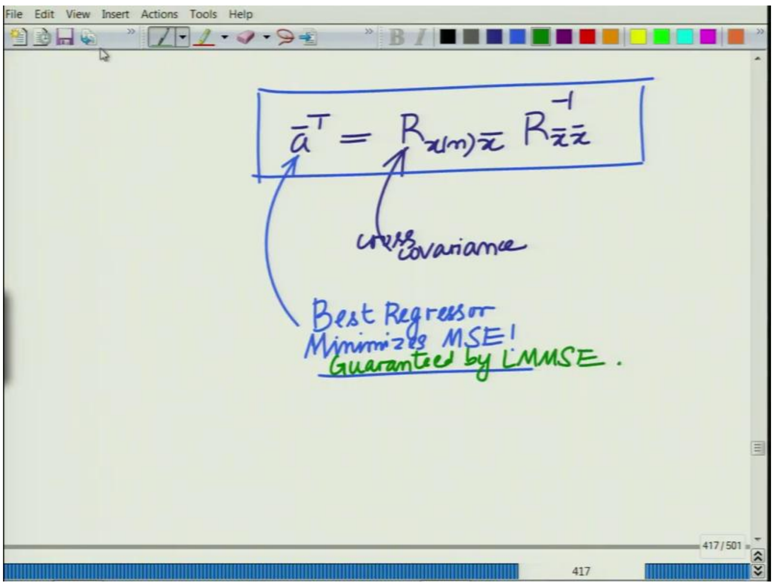
And in fact, you are using L past samples so remember this implies, this becomes L^{th} order, so this becomes an L^{th} order AR model if you are using 1 past sample like xn A1 times xn minus 1 that mean, that is the first order model and A1 xn minus 1 plus A2 xn minus 2 that becomes a second order AR model we are using L past samples. That is the L^{th} order AR model.

(Refer Slide Time: 10:02)



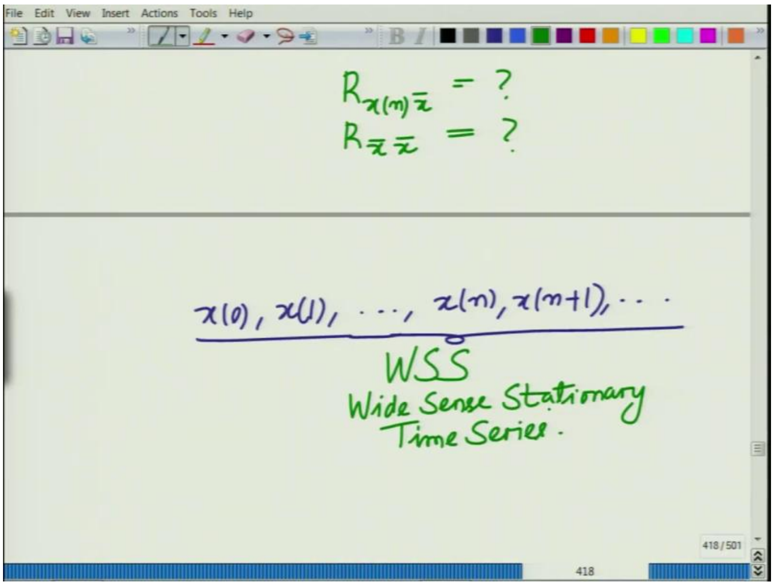
Now, the idea here is what are the regression coefficients, how to determine which are essentially nothing but the, which are nothing but the regression coefficients and these as we know can be determined by using the LMMSE we once again come back to the LMMSE principle that is, if we call the quantity to be predicted as xn and then, we denote by x bar this is the quantity to be estimated and x bar these are the known this is xn minus 1 xn minus 2 so on xn minus L.

(Refer Slide Time: 11:44)



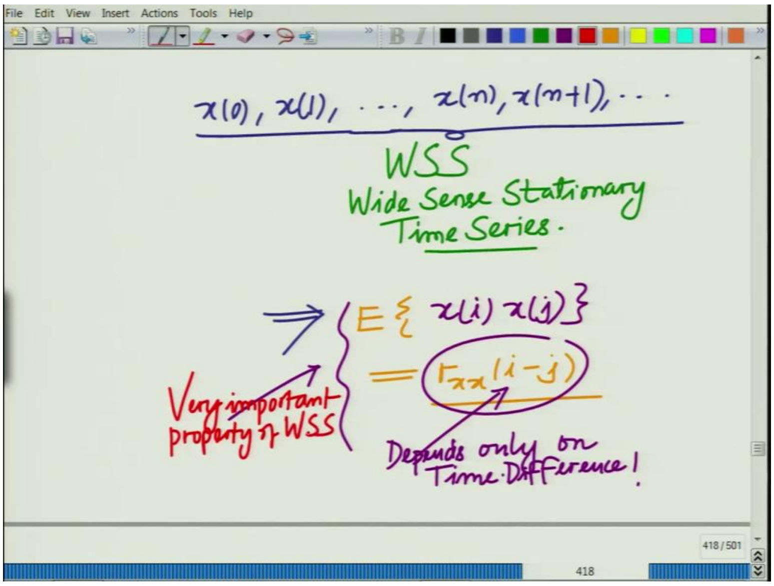
Then, the optimal regressor as we know, is given a bar transpose this is the cross covariance of xn bar x so this is the cross covariance times R x bar x bar inverse this is your optimal regressor, this is the best regressor that minimises the mean squared error, this minimises the MSE and we know that from the LMMSE principle, so that is we want to predict or if you want to estimate x bar then the optimal regressor is Rxy Ryy inverse and this is basically guaranteed by the LMMSE principle.

(Refer Slide Time: 13:13)



Now, let us look at these quantities the question that we want to ask is what are these quantities Rx n x bar and what are these quantities Rx bar and for that we will need some property on this time-series x0, x1, x n plus 1 so on, we will assume that this is what is known as a WSS that is wide, this is a wide sense stationary time series that is we will assume this is a wide sense stationary time series.

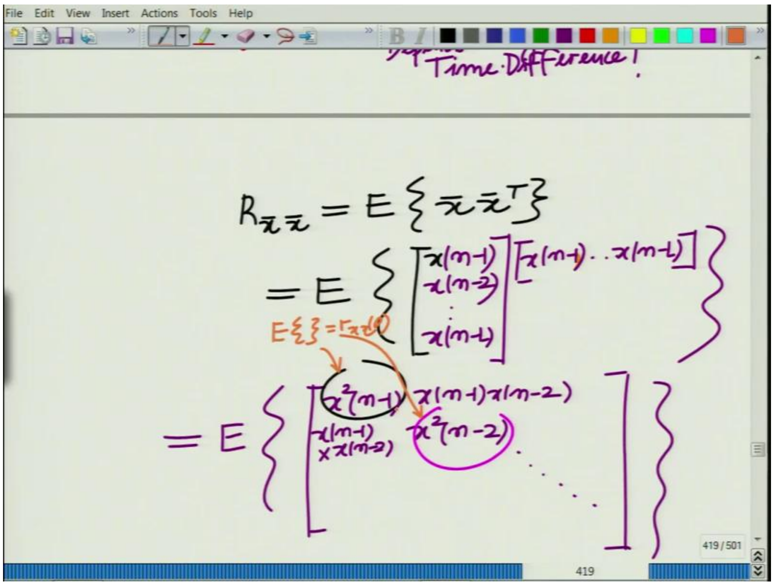
(Refer Slide Time: 14:24)

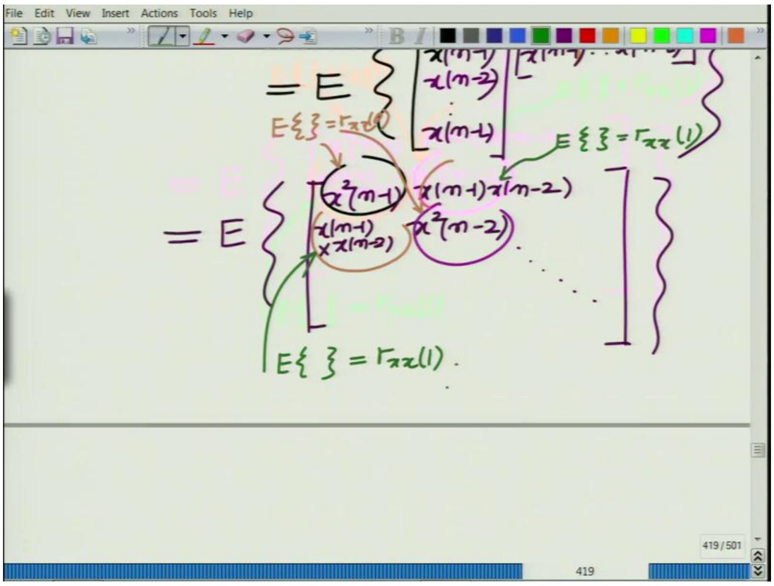


Which essentially implies that if you look at the correlation that is x of n minus i times x of n minus j this only depend on i minus j that is the lag between these two samples this is an important property of wide sense stationary time-series that is if you look at auto correlation between two different samples x of n minus i into x of j or you can also write it as x of i times probably it is better to write it this way x of i times x of j this is simply x of i minus j depends only on the lag, on the time difference, depends only on the time difference between those two is two samples that is we will get two samples x of n and x of m, ask what is the correlation?

The correlation is Rxxn minus m it does not depend on n and m but only the rather that difference n minus m so this is essentially a very important property a characteristic, a very important property of WSS, very important property of wide sense stationary random process.

(Refer Slide Time: 16:19)





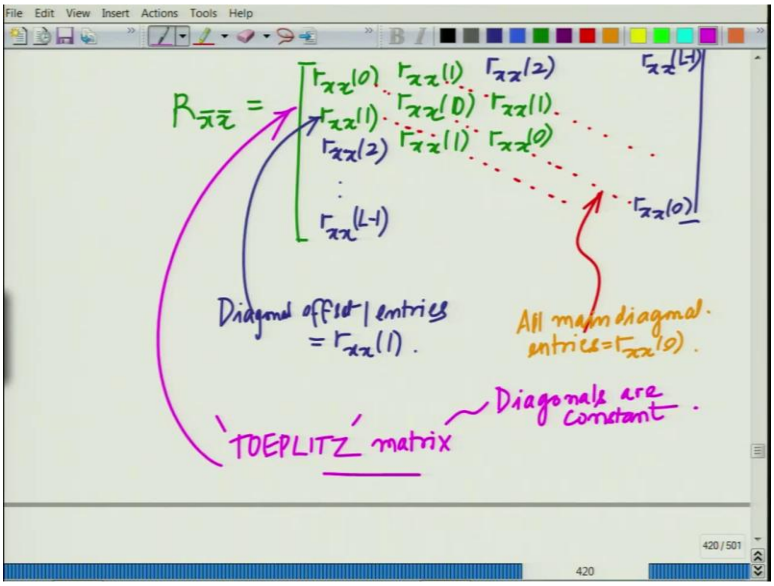
Now, let us ask the question what is Rx bar x bar, this is essentially expected value of x bar x bar transpose, which is expected value of x0, x1, up to xn minus 1 times, I am sorry, this has to be xn minus 1 xn minus 2 xn minus L times the transpose xn minus 1 xn minus L.

Now, if you expand this and write it you will realise that this becomes expected value of for instance, here you will have x square n minus 1 x square n minus 2 xn minus 1 xn minus 2 xn minus 1 into xn minus 2 and so on and interestingly if you look at it you will see all the diagonals expected value of x square n minus 1 x square n minus 2, expected value of x square n minus 1, expected value is Rxx 0 because, the time difference is 0 x n minus 1 into xn minus 1.

Similarly, expected value of x square n minus 2 expected value will once again be Rxx 0 expected value of xn minus 1 into xn minus 2 time difference is 1 therefore, expected value will be Rxx 1.

So, if you look at it all the diagonals will have similar elements, the main diagonal will be Rxx 0, the diagonal with offset of 1 will be Rxx 1, diagonal with offset of 2 will be Rxx 2 and so on. So, for instance expect, if you take the expected operator inside and write this expected value of this and this if you look at the expected value of these quantities for instance, if you look at the expected value of these quantities expected value equals Rxx 1 similarly, here the expected value of this quantity equals Rxx 1.

(Refer Slide Time: 19:50)

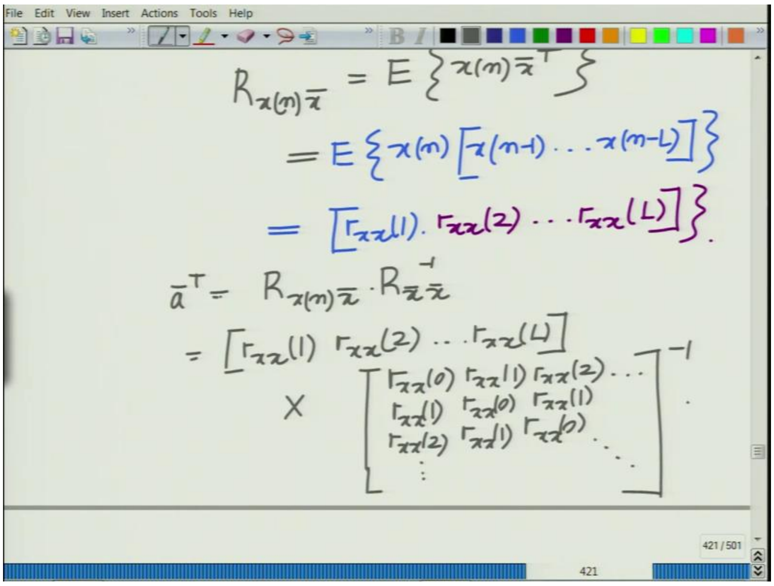


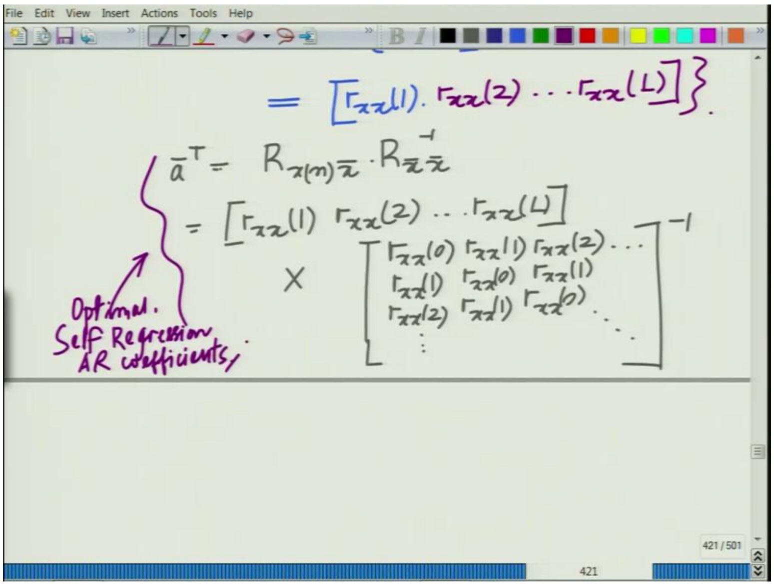
So, if you look at this because, of the wide sense stationary nature of the random process Rx bar x bar will look like all the diagonals will be Rxx 0, Rxx 0 and 1 offset will be Rxx 1, Rxx 1 and this will be Rxx 1 and this will be Rxx 2 so on this will be Rxx of, if I m not mistaken, this will be Rxx of L minus 1 so on, so this will be Rxx of 0 this will be Rxx of L minus 1 and if you look at this, all the diagonal entries if you look at this all, so if you look at this all the diagonal principle diagonal so, all the principle diagonal entries, all at offset of 1 diagonal offset of 1 these are Rxx 1 and 0, so this has a banded structure.

So, you will have the diagonal is Rxx 0 then the diagonals with or offset the sub diagonals you can see which are offset of 1 sub and super diagonals which are offset of 1 they have Rxx 1 the diagonals that are offset of 2 from the main diagonal they Rxx 2 and so on and this matrix is this is a very interesting structure such a matrix is known as a TOEPLITZ matrix so this matrix is known as TOEPLITZ.

So, if you look at x x vector of x bar x bar Hermitian this has a TOEPLITZ matrix each diagonal principal and sub diagonals, diagonals are constant, the diagonals are essentially constant that is if you look at any diagonal all the values on the diagonal are the same.

(Refer Slide Time: 23:30)



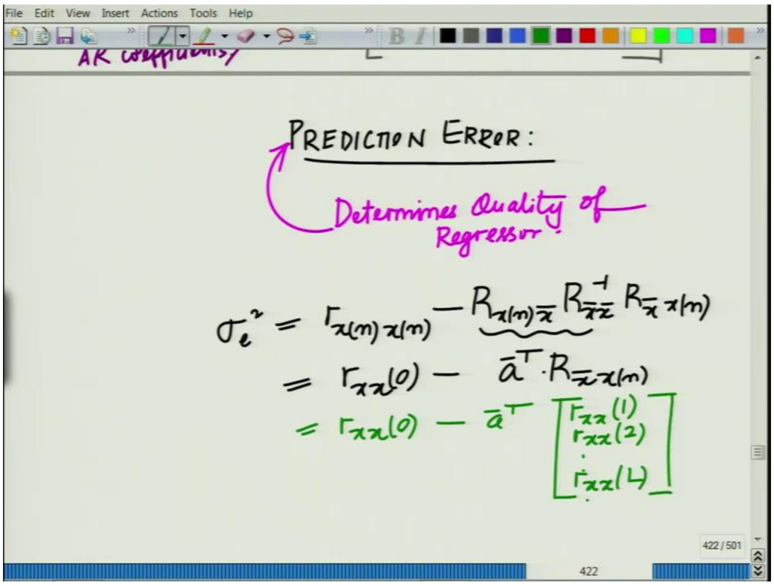


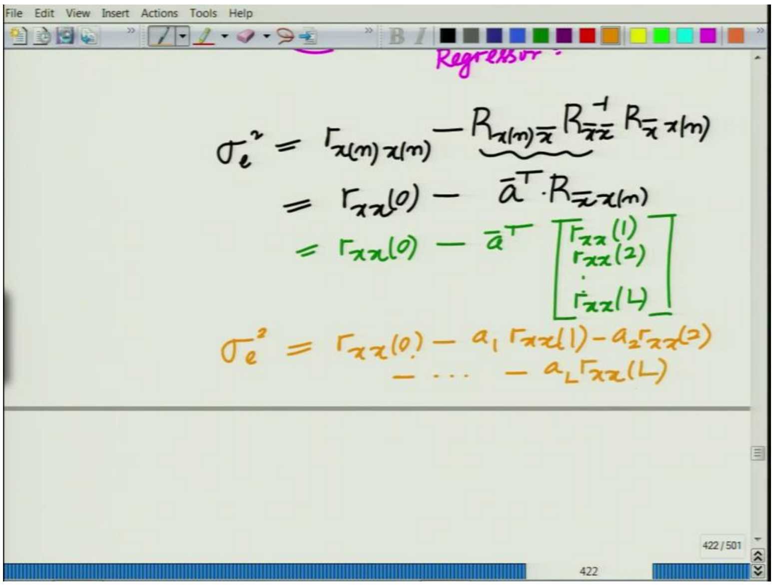
And, now we want to find this, the other matrix that is Rxnx bar that is expected value of xnx bar transpose, which is expected value of xn minus 1 so on up to xn minus L and this is equal to your basically your, you can see this is essentially your Rxx 1 Rxx 2 Rxx L which implies that your a bar transpose, this is your a bar transpose is Rxnx bar times Rx bar x bar inverse, which is essentially Rxx 1 Rxx2 Rxx L times inverse of this auto correlation matrix which is Rxx 0 all the diagonals are Rxx 0 Rxx 1 Rxx 1 Rxx0 Rxx 1 Rxx 2 so on and then you have the inverse of this matrix.

So, this is essentially your optimal self-regressor, optimal self-regression or basically your auto-regression coefficients, which minimise the mean square error. So, these are the best regression coefficients, which minimise the mean squared error hence, can be used for your prediction essentially also your forecasting in the stock market kind of application you would like to forecast the stock price in the next day or the next couple of days and so on and thereby, bet appropriately or thereby investor properly.

So, this is a lot of applicants in fact, even coding compression, if you can predict the next values of the signal then, the remnant what is the difference between the prediction and the actual signal, this is what is known as the innovation and then, one can actually compress only the innovation so, there are many applications actually.

(Refer Slide Time: 27:06)





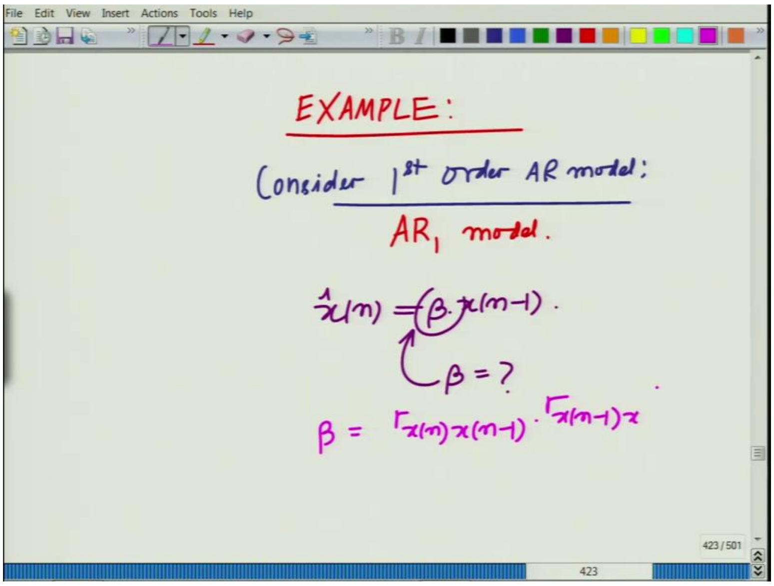
And the prediction error now one can ask the question as a scientist it is not just enough to come up with a prediction but it is also important remember to characterise the prediction error because what determines the quality of the prediction is a prediction error because anyone can come up with a predictor.

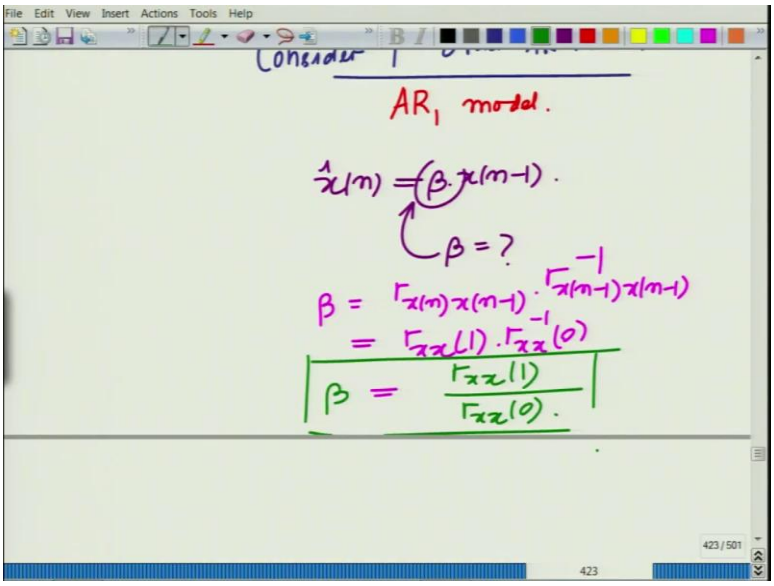
Now, what determines the quality of the predictor, is the prediction error so therefore you cannot just simply give a predictor but also the predictor and that is the regressor and what is the regression error going to be because that determines the quality and hence that will in turn determine whether this quality or this level or this regressor is acceptable or not because without that one is totally in the dark about what is the nature or what is the quality of the prediction.

So, that is an important, so this determines the quality of the regressor and thanks to the element missy principle, we already know that is your error sigma square this is in this case because, this is a scalar quantity so the prediction the covariance itself will be the variance.

So, this will be Rxnxn minus Rxn x bar Rx bar x bar inverse x bar times Rx bar xn, now Rxnxn this is Rxx 0 minus Rxn x bar, now this quantity if you look at this Rxn and x bar Rx bar x bar inverse this is nothing but, a bar transpose so this a bar transpose Rx bar xn that is expected value of x bar this is Rxx 0 minus a bar transpose expected value of x bar into xn this is Rxx 0 Rxx 1 Rxx sorry, this is Rxx 1 because the first value is x minus 1 Rxx 2 so on Rxx L and therefore, this is essentially given sigma e square regression error is Rxx 0 minus a1 Rxx 1 minus a2 Rxx 2 minus 1 minus aL Rxx L. And therefore, that completes our auto-regression.

(Refer Slide Time: 30:30)



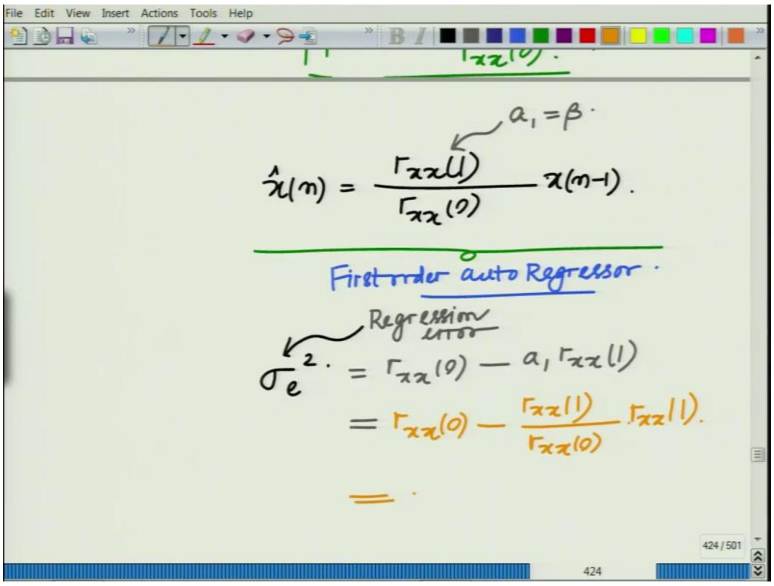


Now, let us look at a simple example to understand this as usual because when we do simple examples, we tend to understand things better so, always, always support or enhance your understanding, we are trying to work out a simple example, like that always helps in significantly improving your understanding.

So, let us look at a first order AR model the simplest, consider a first order AR model although simple, this is used very frequently, this is simply termed as AR 1 that is AR subscript 1 model and of course AR model of order n you can determine denote using ARn so, AR 1 model, which means you are trying to form the prediction x hat of n equals beta times xn minus 1, what is beta?

And the answer to that is beta equals your well, Rxnxn minus 1 times all these quantities are scalars so I am simply reusing scalar quantities, not using matrices to represent these Rxn minus 1 xn minus 1 inverse which is Rxx remember it depends only on lag Rxx 1 times Rxx 0 inverse, which is nothing but, Rxx 1 divided by Rxx 0

(Refer Slide Time: 32:30)



So, therefore your optimal prediction that is x hat of n this is simply Rxx 1 divided by Rxx 0 into this is essentially your first order auto-regressor, this is essentially your first order AR 1 model that is your x hat n equals Rxx 1 divided by Rxx 0 times x n minus 1. And now, one can ask what is the modelling error or what is the regression error?

This is essentially your Rxx 0 minus a1 times Rxx 1 remember this is your a1 equals beta so, this is going to be your Rxx 0 minus Rxx 1 divided by 0 times Rxx 1 which is essentially equal to Rxx 0 R square xx 1divided by Rxx 0, so this is essentially your regression error, this is essentially your regression error for this first order AR model.

So, essentially that completes our discussion of the AR modelling that is auto-regression, which is a special case of regression. Essentially, we are trying to regressed using the past samples of the same process and we are using the property that these samples are wide sense stationary.

So, essentially the correlation that is what we call as the ACF function that depends only on the lag, that is the expected value of xn minus 1 xn minus i into xn minus j is Rxx i minus j. So, that completes this discussion which I have already said is very important as many applications in forecasting, coding, compression, etc. So, please take a look at this again we will come back, we will discuss, continue our discussion of other similar concepts in the subsequent models. Thank you very much.