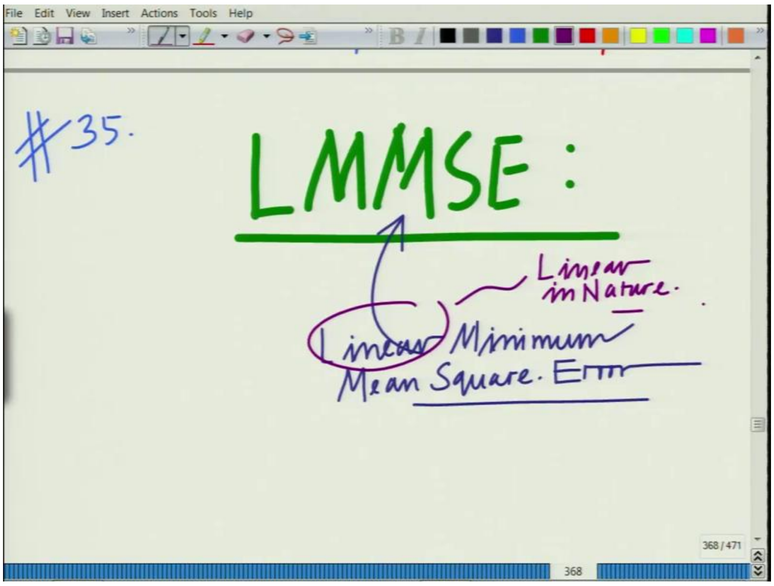
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Lecture – 35
Linear Minimum Mean Square Error (LMMSE) Principle

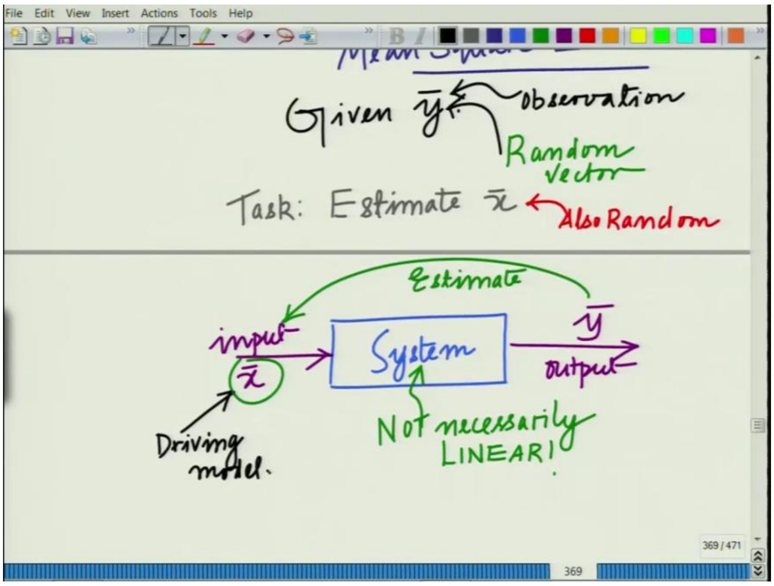
Hello, welcome to another module in this massive open online course on Applied Linear Algebra. Let us start looking at another very important concept which is based on principles of linear algebra, heavily uses of linear algebra and that is the concept of LMMSE estimation which is a specific case of MMSE estimation. But LMMSE estimation is more popular in general owing to its practicability and easy applicability , low complexity low complexity and naturally LMMSE as the name implies Linear Minimum Mean Square Error estimation is used in the analysis of linear systems and essentially linear models .

(Refer Slide Time: 01:06)



So, let us look at the concept of LMMSE which is essentially stands for, I am going to explain the name in detail, linear minimum means square means square or you can also say mean squared error, this is the principle and the key word here is linear, this is linear in nature al. So, this estimator is linear in nature .

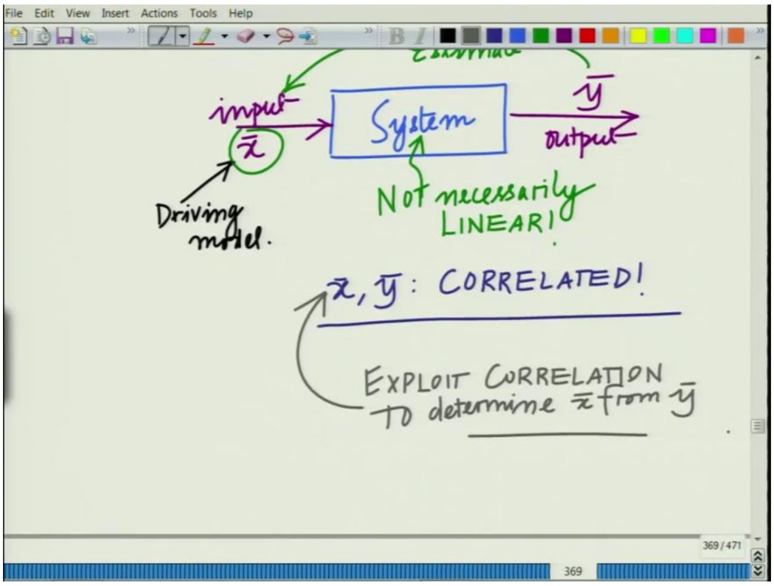
(Refer Slide Time: 2:37)



And the idea here is the follows, so given vector y bar, this you can think of this as the observation , given y bar which is the observation which is essentially a random vector , which is y bar which is observation which is a random vector. The task is to estimate another vector x bar, which is also basically a random vector, this is also a random that is x bar essentially is an underlying vector, which drives the model, you can think of this as an input to a model and the output or the observation is y bar and therefore, naturally we would like to use y bar which is essentially, we would like to use y bar and determine what is the input x bar that drives this model .

So, essentially if we look at this as an input output system . For example, this is your system which by the way need not necessarily be linear , not necessarily linear, the system itself is not necessarily. What is linear, is the estimator that is the interesting. So even though the system is not necessarily linear, 1 can construct a linear estimator, it might be suboptimal, but we will be interested in finding the best linear estimator, so, that is the point. So here you have the input, which is x bar which is driving the system and this is your output y bar.

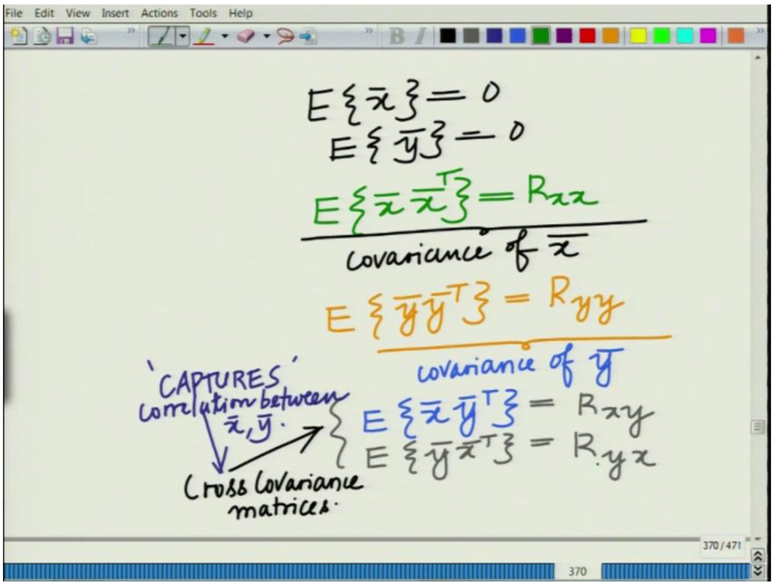
(Refer Slide Time: 5:23)



And now, you want to use y bar to estimate use y bar to essentially estimate what is the input x bar which is driving the model. This is essentially driving your IO model. This is the input to the model . And the point here is what we want to exploit is the fact that x bar comma y bar these are correlated, this is a very important point, x bar influences y bar. So, naturally x bar and y bar are correlated, we want to exploit this correlation , we want to exploit the correlation between x bar and y bar to determine x bar from y bar that is the point.

So, exploit the correlation to estimate that is the important point here, exploit is a very important aspect of MMSE estimation is exploit this correlation could determine exploit this correlation to determine x bar from y bar.

(Refer Slide Time: 6:31)



Now, let us look at the properties of x bar and y Bar, for simplicity, let us assume that expected value of x bar equal to 0, expected value of y bar equal to 0. This is may not be the case, but this principle can be readily extended to scenarios where they are not 0, for simplicity of analysis let us assume that these are zero mean.

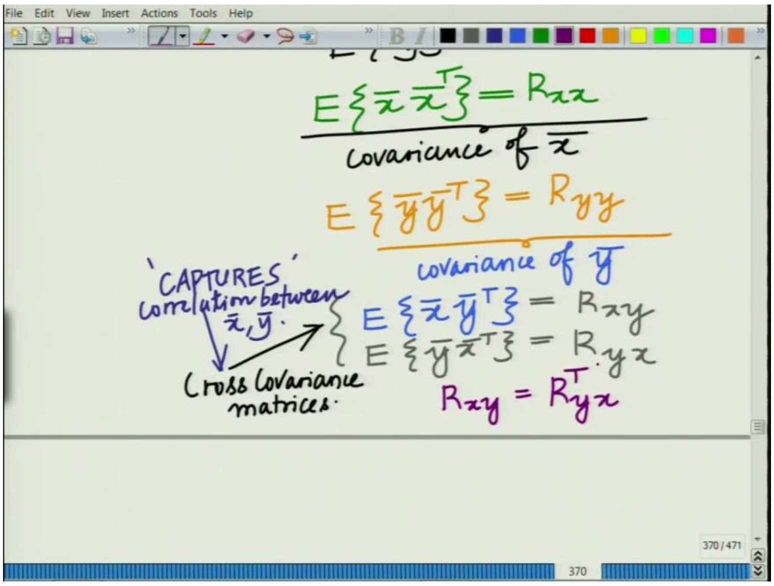
I will talk to you about what to do when these are not necessarily zero mean. And then what happens here is your expected value of, now we are interested in what are known as the second order statistical properties when we talk about that, we have expected value of x bar x bar transpose, which is R x x, this is the covariance of x.

Then we have expected value of y bar y bar transpose which is R y y which is essentially the covariance, this is also something that you must be familiar with, this is the covariance and expected, the interesting quantities are expected x bar y bar transpose which we denote by R x y and expected y bar x bar transpose which is R y x.

And these are essentially the interesting quantities, these are what we call as the cross covariance. In fact, these cross covariance matrices, these are what are capturing the correlation between x and y. If these cross covariances are zero, then naturally it means that x bar and y bar are uncorrelated.

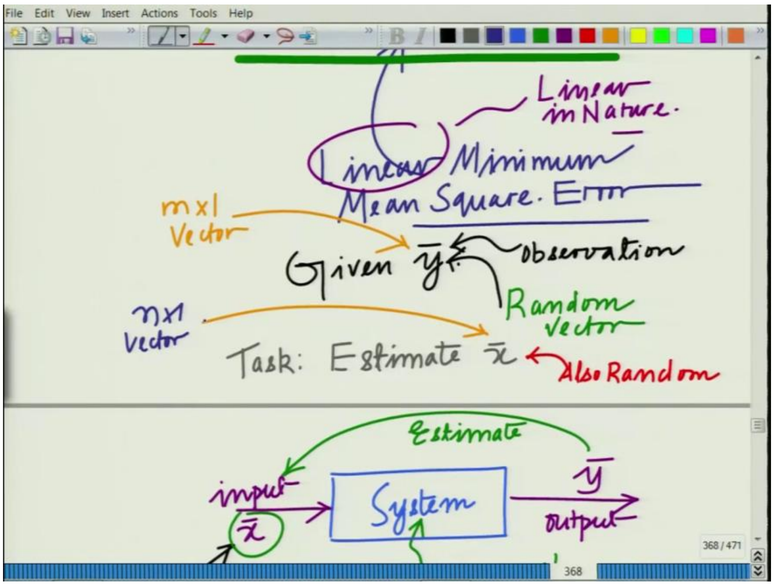
If they are uncorrelated, then there is not much you can do about it, obviously makes sense if x bar and y bar are not correlated, then it is reasonably difficult. And you can see if they are not [co] there is no correlation, 1 cannot expect to determine or get an idea of x bar from y bar, the better the correlation, the better is your estimate of x bar based on y bar going to be. So that is a very intuitive idea. So, cross covariance and this basically captures the correlation, this captures the correlation between x bar and y bar that is the interesting aspect. And we can also see that our x y is nothing but these are the transpose of each other.

(Refer Slide Time: 9:57)



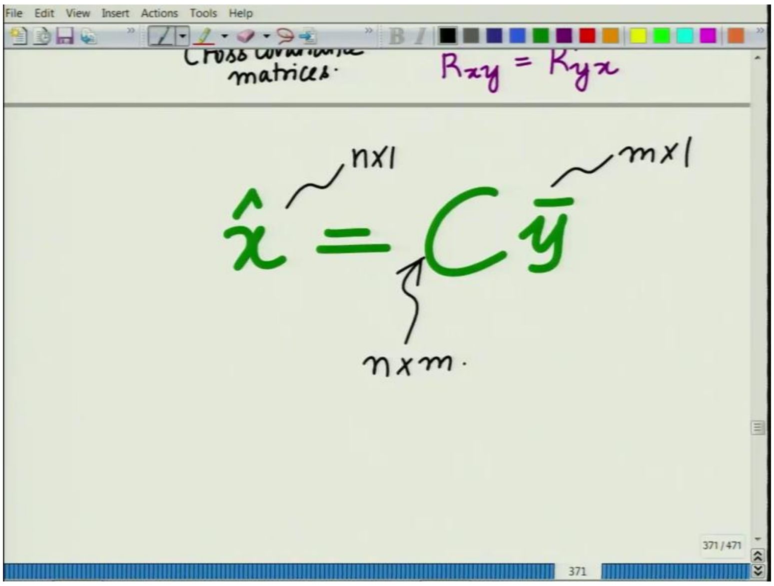
So, this is essentially what you can see is that R x y equals, this satisfies the property R x y equals R y x. So that is these are basically the different quantities that we are going to use, and now we are going to estimate x hat as the following thing.

(Refer Slide Time: 10:10)



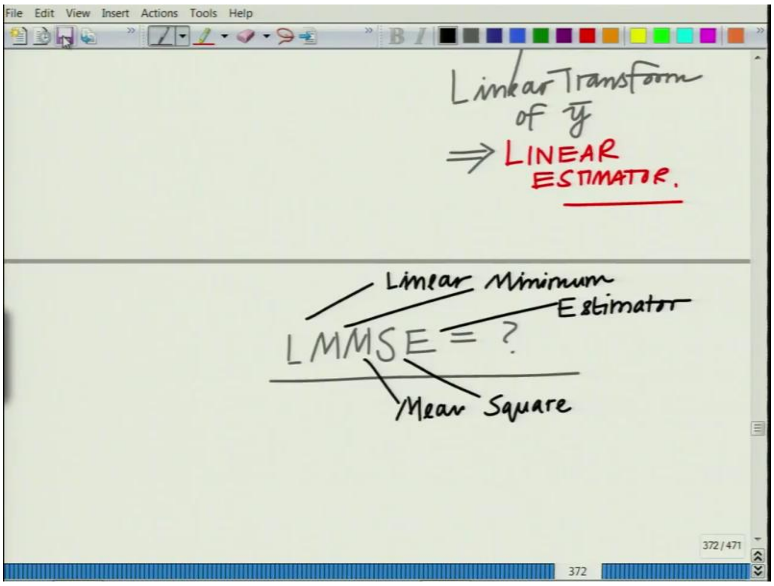
x hat, let us further assume that just to give these dimensions let us assume that for instance y bar this is an m cross 1 vector this is an m cross 1 vector m cross 1 and x bar this is an n cross 1 vector , x bar this is this is an n cross 1 vector .

(Refer Slide Time: 11:01)



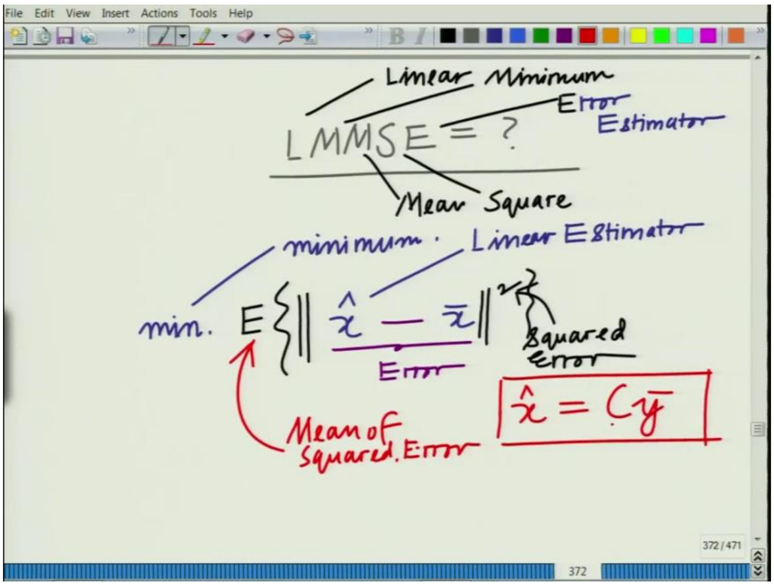
Now, the idea here is to determine the estimate x hat equals C times y bar, where naturally this matrix C, this is going to be this is going to be an n cross m matrix because this is m cross 1 this is m cross 1 and this is your n cross 1. So, this is going to be n, this is going to be this is going to be an n cross m matrix and essentially you can see this is a linear transform.

(Refer Slide Time: 11:59)



So, this is a linear transform of y bar, y bar implies this is a linear estimator . So, that is an important point implies this is implies this is a linear estimator this is the linear transform of y bar implies this is a linear estimator and hence the LMMSE. Now, let us look at what is the LMMSE? What do we mean by the LMMSE? So, there are a couple of terms, one is Linear Minimum Mean Square and this is Estimator. Now, let us understand so this is basically the anatomy of this term. Let us dissect this term and understand the meaning of each component and this is a loaded term LMMSE .

(Refer Slide Time: 13:38)



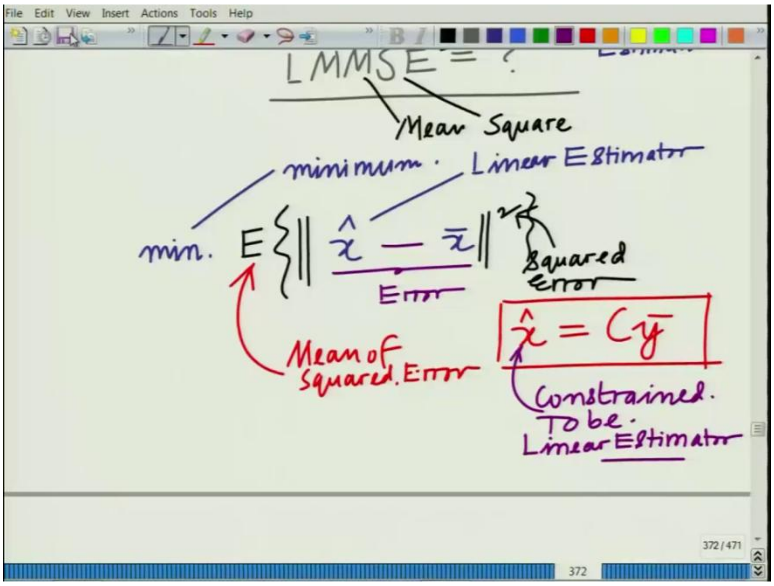
The first thing to perform LMMSE estimation is to essentially understand what this represents, all . So, we have the vector x bar and we have the estimate x which is essentially a linear estimate that is so, that explains part of this, so that explains the linear and estimator, minimum mean square error, I am sorry this has to be error , linear minimum mean square error estimator that is your LMMSE estimator.

Now, so now we need to find the error between these, the estimator and actual quantity so this basically is the error . And now we take the square of this error, you take the norm so this is where you get the squared error. And now you take the mean of this squared error, now you take the mean of this squared error .

So we are performing the mean of the squared error, the mean of the squared error and then we want to find the estimate x hat size that we want to minimize, so this gives you the minimum .

So, this is essentially your Linear Minimum Mean Squared Error that is basically find the linear estimator x hat, x hat equals C y bar such that if you look at the square of the error norm x hat minus x bar square and you look at the mean that is the expected value of norm of x hat minus x bar square that should be minimized that is what your linear minimum mean squared error estimator is doing.

(Refer Slide Time: 16:24)

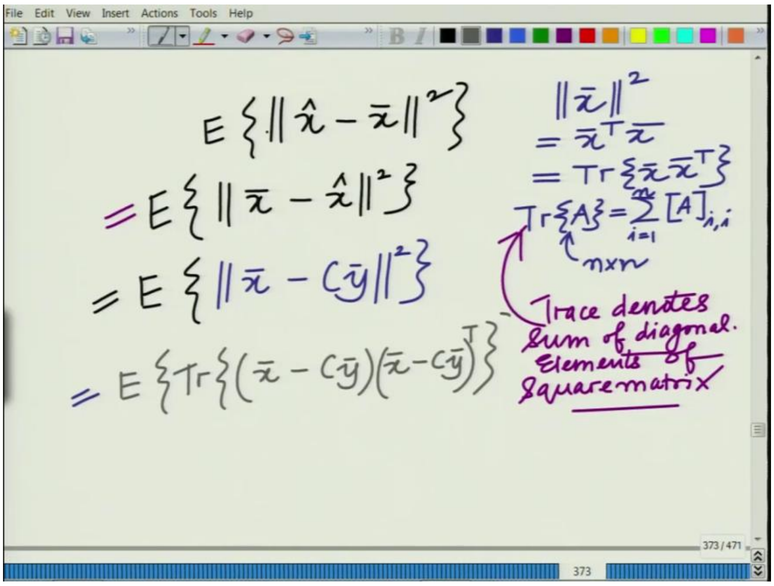


Now, the condition here is that x hat important condition here is that x hat equals C y bar that is x hat this is essentially a linear estimator , x hat is constrained to be x hat is constrained to be a linear estimator, if x hat is not necessarily linear, if x hat is an arbitrary function of y bar can be annoying if it is a nonlinear, it can be linear or nonlinear then it becomes the MMSE estimator, simply the minimum mean squared error estimate that is the best estimate x hat such that expected value of norm x hat minus x bar square is minimized.

If x hat is the general is the best estimate amongst all estimators that it becomes the MMSE estimate, if it is best only among the class of linear estimators, it becomes the LMMSE estimator, the linear minimum mean squared error estimate.

And as I already said, the linear MMSE estimator is frequently used in practice because it can be determined relatively easily, especially when the model when the input-output model is linear, as we are going to see shortly, as we are going to see as we go through these modules, we are going to see how to determine the linear minimum mean squared error estimate.

(Refer Slide Time: 17:45)

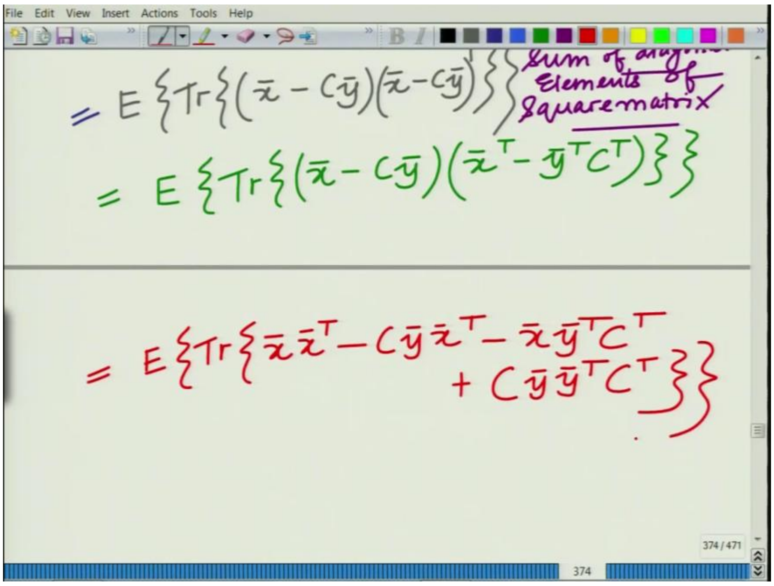


Now, the general principle of linear minimum mean squared [esti] or the general linear minimum mean squared error estimate can be found as follows. So we want so consider this quantity, expected value of norm of x minus x bar square so we have this quantity, we have this quantity and I am going to use the property here that which we well know norm x bar square for real vectors this is simply x bar transpose x bar which can also be written as trace of x bar x bar transpose will trace of any matrix A is simply the sum of the diagonal elements for an n cross n matrix A trace is denotes the sum of the diagonal elements . So, this I think all of you should be familiar trace denotes the sum of the diagonal. So, trace denotes the sum of the diagonal elements of a square matrix.

So, I can write norm of x bar square as x bar transpose x bar which is also equal to trace of x bar x bar transpose. So this is an interesting property, if you are not familiar with it, you can just quickly check it because if you look at x bar x bar transpose the diagonal elements will be x 1 square x 2 square so on x n square. And this is what a real vector and you can similarly extend it to a complex vector. In that case we will simply write x bar hermitian x bar all trace of x bar x bar hermitian.

And this is essentially equal to therefore I can write this as expected value of, let me just write this as first x bar minus x hat norm square, which now substitute the expression for x hat, remember it is a linear estimator so this becomes x bar minus C y bar norm square, which is expected value of trace of now let us bring in the trace . So, so trace of x bar minus C y bar times x bar minus C y bar transpose, and basically close the brackets.

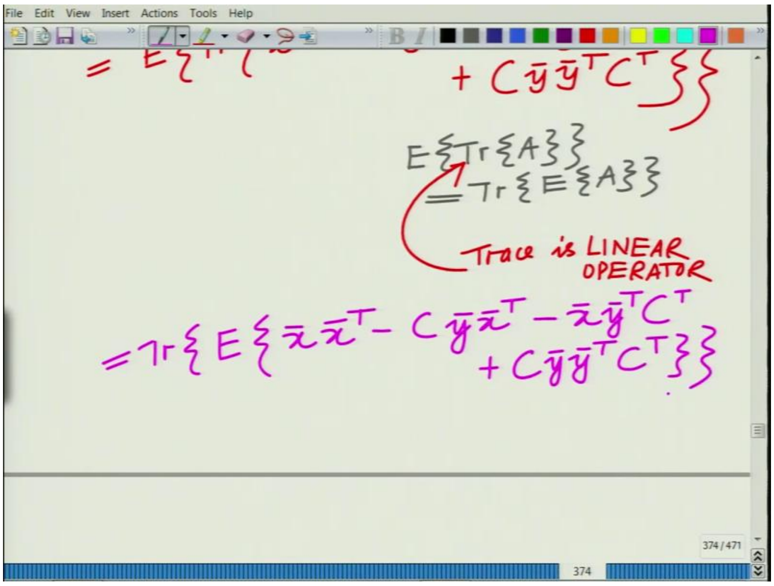
(Refer Slide Time: 20:55)



And we are using C y bar because, remember we hat x that is equal to C y bar this is a linear estimator and from here, we can expand this thing which is essentially you make this as follows, expected value of trace x bar minus C y bar x bar minus C y bar times x bar transpose minus y bar transpose C transpose which if you expand this term by term this is going to be equal to the expected value of the trace or x bar x bar transpose minus C y bar x bar transpose minus x bar y bar transpose C transpose plus y bar transpose plus C y bar y bar transpose C transpose and close the brackets.

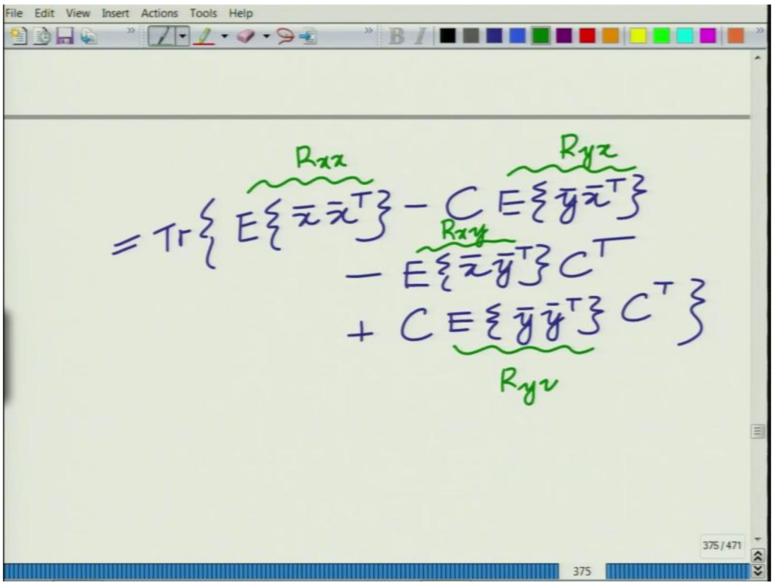
And now, we are going to do a standard trick which is essentially to interchange whenever we have the second order statistics, we can interchange the trace and the expectation. So trace of expectation equal to expected value of the trace .

(Refer Slide Time: 22:21)



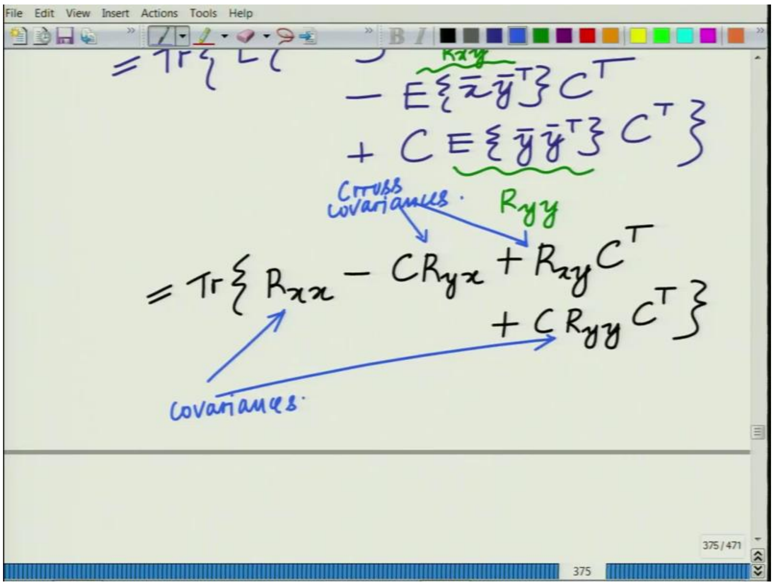
So, I am going to use the property expected value of the trace of a matrix because trace is a linear at the end of the day, equals trace of expected value of A, this is possible because trace is a linear operator , trace of A plus B equals trace A plus trace B and so on because trace is a linear and therefore now this becomes very interesting now, when you say take the trace inside, so this becomes trace of expected value of writing exactly the same thing, x bar x bar transpose minus C y bar x bar transpose minus x bar y bar transpose C transpose plus C y bar y bar transpose C transpose.

(Refer Slide Time: 23:44)



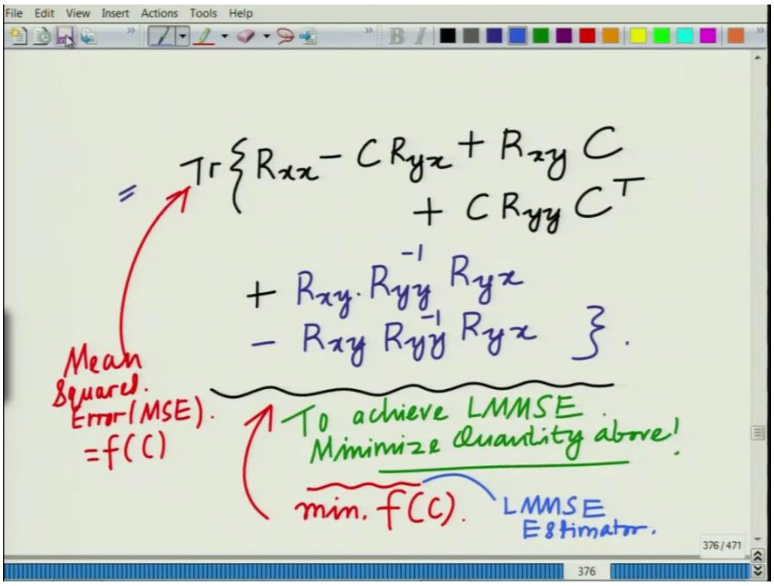
Which now, if you take the expected value of the individual terms, this becomes trace of expected value or x bar x bar transpose minus C expected value of y bar x bar transpose because C is a constant matrix minus x bar expected value of C transpose plus C expected value of y bar y bar transpose C transpose. Now, this is essential if you look at this you can see this is equal to R x x, expected value y bar x bar transpose this is R y x. This is essentially R x y and this is essentially R y y.

(Refer Slide Time: 24:55)



So, essentially basically you can simplify this as trace or substituting these quantities, this becomes trace of R x x minus C times R y x plus R x y into C transpose plus C R y y into C transpose. This is essentially the succinct expression that you get in terms of the various covariances that is you have R x x, R y y and also importantly note the cross covariances R x y R y x . So, these are the covariances so this expression involves the covariances, all the second order statistics and the cross covariances. Implies, this involves the covariances and the cross covariance .

(Refer Slide Time: 26:35)



And therefore, now I can further write this as the following I can simplify this as, so trace R x x minus C R y x plus R x y C plus C R y y C transpose, and I will add and subtract terms I will add and subtract terms, we will see the need for this later. R x y R y y inverse R y x and subtract this term R x y R y y inverse. So I will add and subtract these terms and now I have to minimize this whole quantity to achieve the LMMSE estimator. To achieve LMMSE one has to minimize one has to minimize the above quantity, so let us say essentially we have now set down the quantity that is we can call this as the mean squared error.

So, essentially what we have simplified so far is expected value of norm x hat minus x bar square where x hat is a linear estimator. So, this is basically the mean squared error as a function of C. So, this is the what you call as the mean squared error MSE. This is basically a function of C. So this MSE is equal to this is basically a function of the matrix C.

Now, if I minimize this function of C, then that gives me the LMMSE estimate, if I minimize this as a function of the C that gives me the Linear Minimum Mean Square Error Estimator. So, let us carry out the simplification because this is going to take a little bit of work. So, let us carry out the simplification in the next module and then we will wrap up and we will look at the intuition behind this principle of LLMSE estimation. Thank you very much.