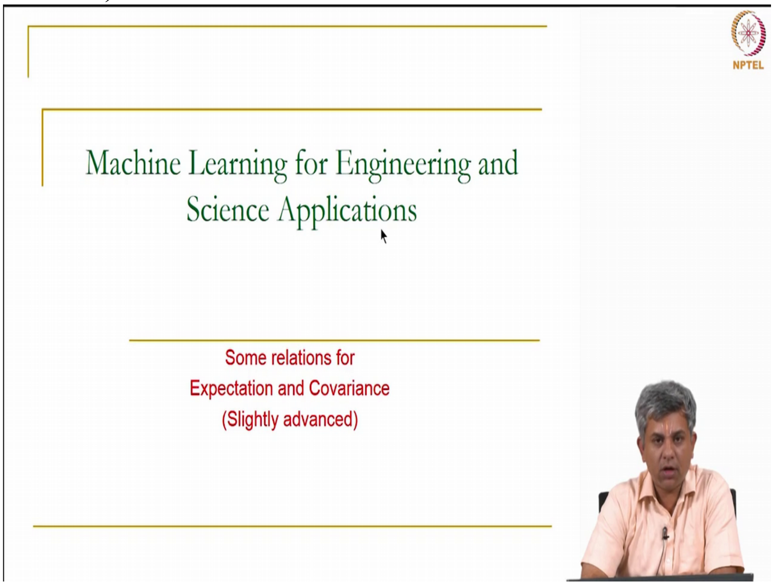
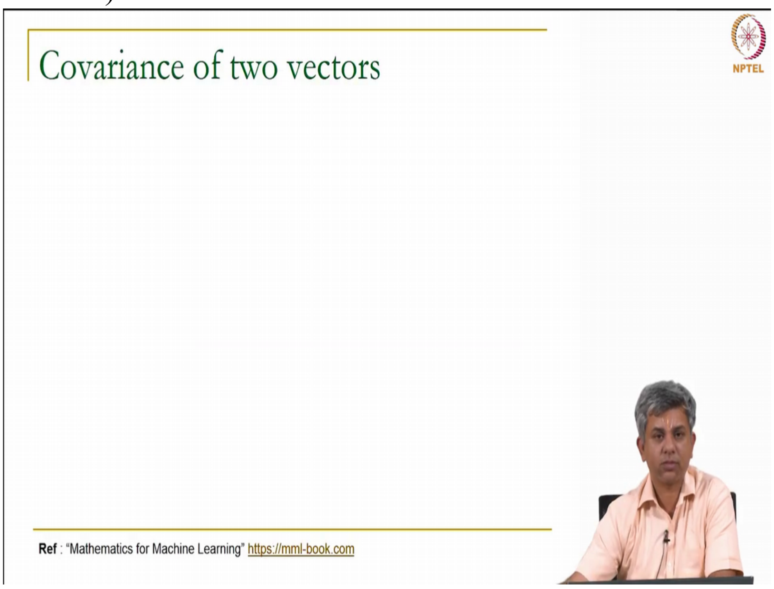
Machine Learning for Engineering and Science Applications
Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology Madras
Some Relations for Expectation and Covariance (Slightly Advanced)

(Refer Slide Time: 00:14)



In this video we will be looking at some more relations for expectation and covariance. These are slightly advanced relations. We will be using these only rarely in the course. So in case you do not understand this portion that is Ok, you will get to understand it a little better as the course progresses.

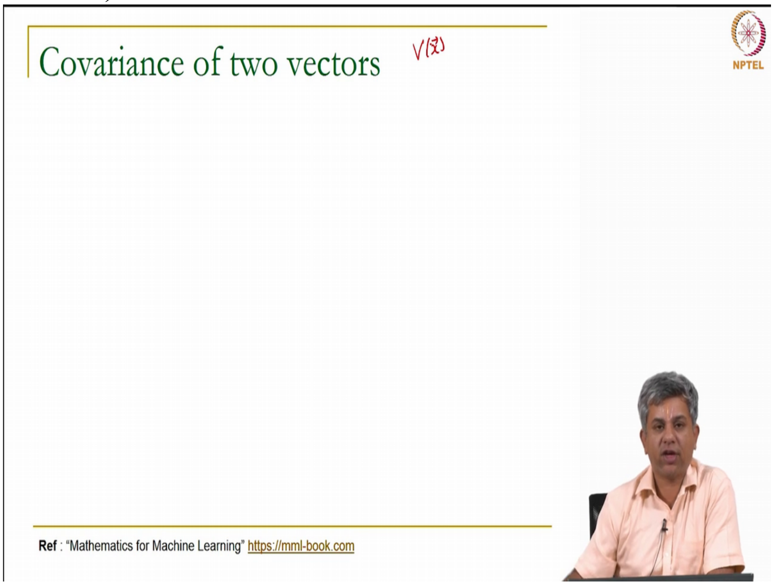
(Refer Slide Time: 00:33)



So please do not panic in case it looks a little bit unfamiliar to you.

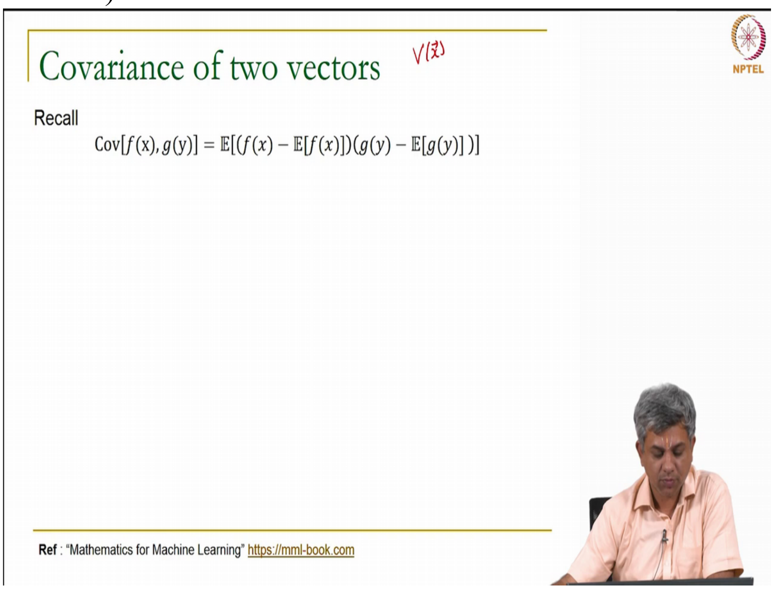
So we have looked at the variance for a single vector, variance of x vector.

(Refer Slide Time: 00:46)



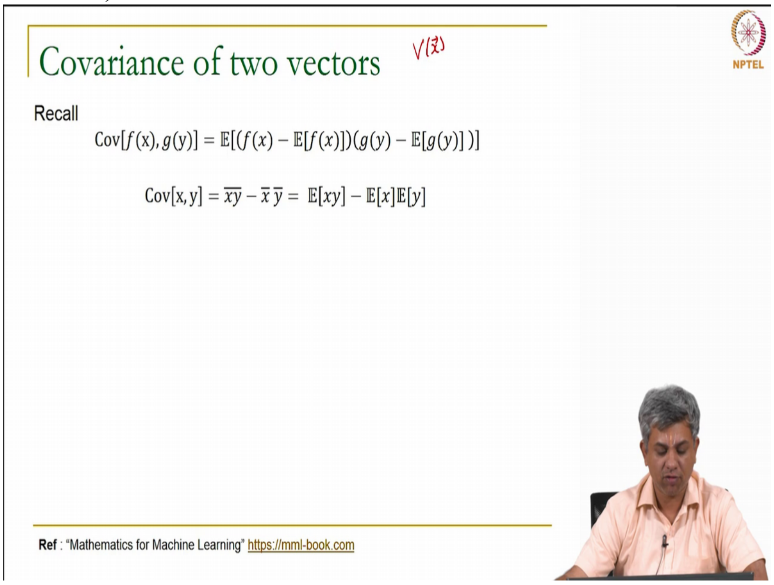
Now we are going to look at the covariance of two vectors.

(Refer Slide Time: 00:50)



So remember that for scalar functions f of x and g of y we had defined the covariance as the deviation from the expectation of f of x multiplied by the deviation from the expectation of g of y and expectation

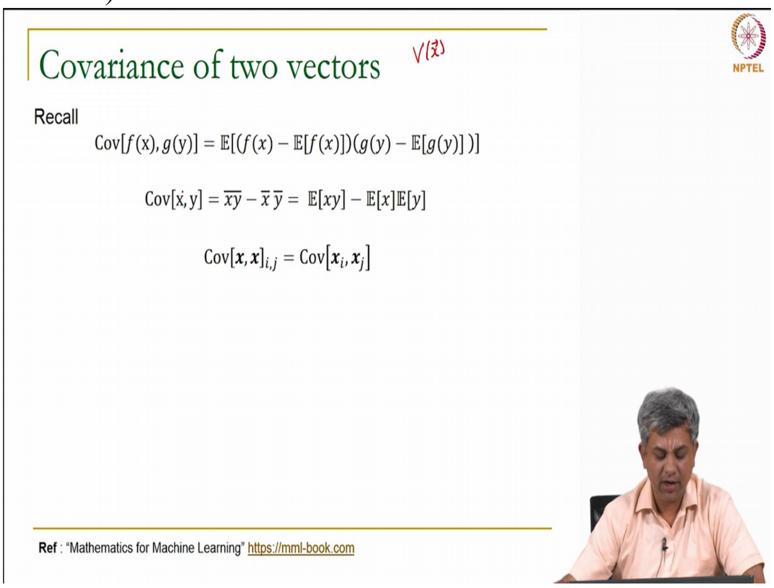
(Refer Slide Time: 01:08)



of the whole thing, Ok.

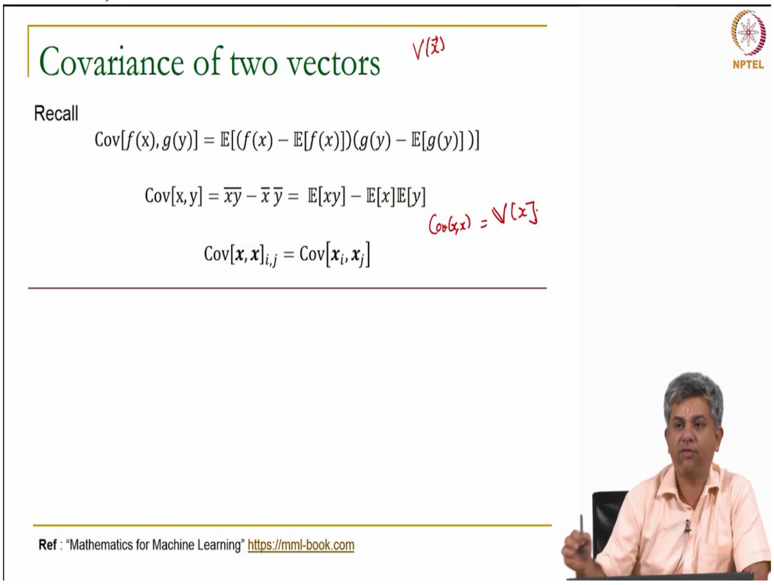
We also, you might also remember that the simpler definition was that of single x and y and that

(Refer Slide Time: 01:16)



definition is given here. Finally we had looked at covariance of two vectors, Ok. This can be called as the variance. Remember just like covariance of x x is called

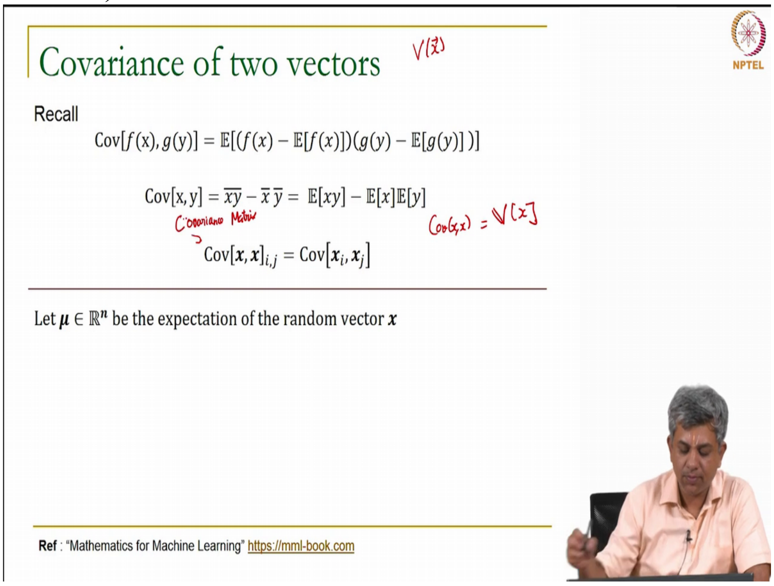
(Refer Slide Time: 01:33)



variance of x which is the square of the standard deviation, Ok.

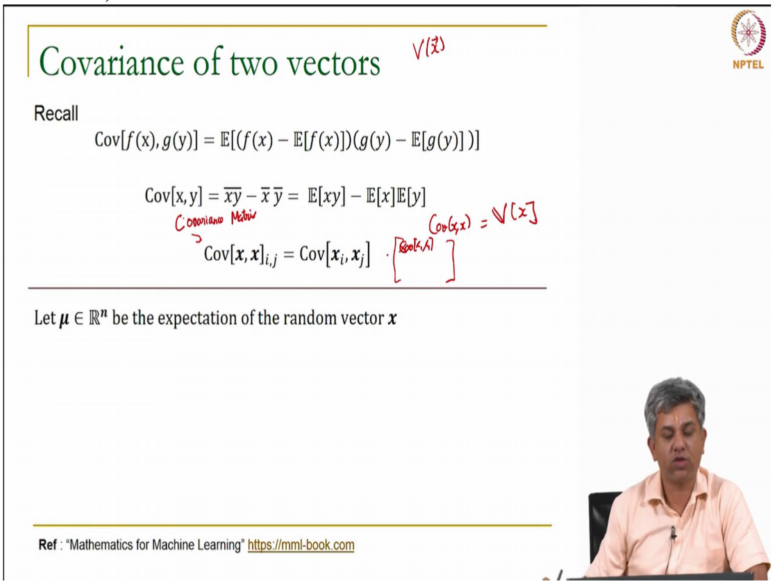
Now we had defined something called the covariance matrix.

(Refer Slide Time: 01:52)



This was a matrix of all covariances, you can recollect this from the previous video. This is simply a matrix of covariance of, let us say the first element will be covariance of x 1 and x 1, so on

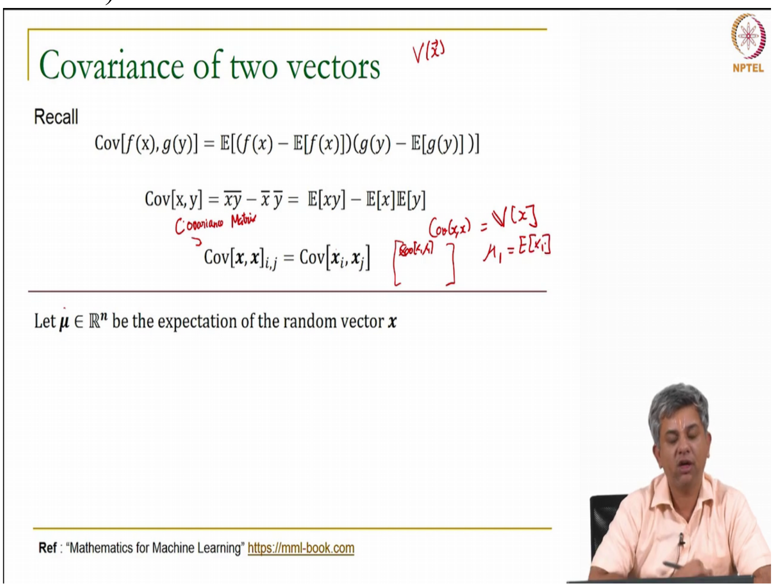
(Refer Slide Time: 02:09)



and so forth. Remember x is a vector and x i is the ith component of the vector x.

Now let us say we denote by mu, this is standard notation, mean or expectation is denoted by mu. Let us say mu is the expectation of the vector x and since it is a vector, mu is also going to be a vector. Again if you look at the previous videos you will see that the first element of mu 1 is the first element

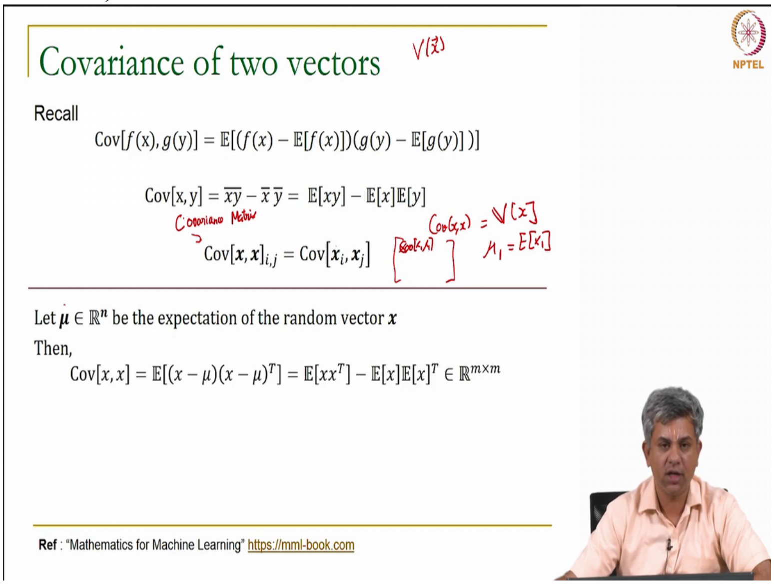
(Refer Slide Time: 02:37)



of the expectation so on and so forth. That is a full vector.

So let us say mu is the expectation of the random vector

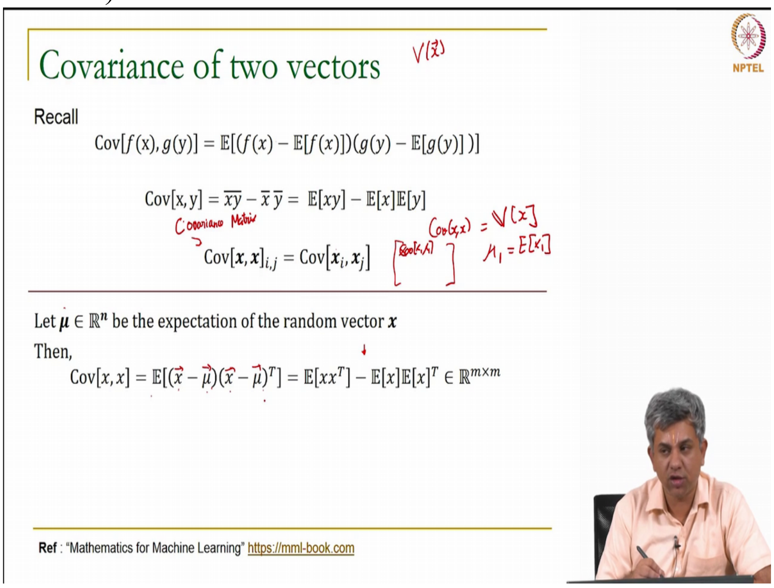
(Refer Slide Time: 02:44)



x then covariance of x comma x which we have seen before, Ok can be written as x minus mu times x minus mu transpose. This is the way that we had defined it before, Ok.

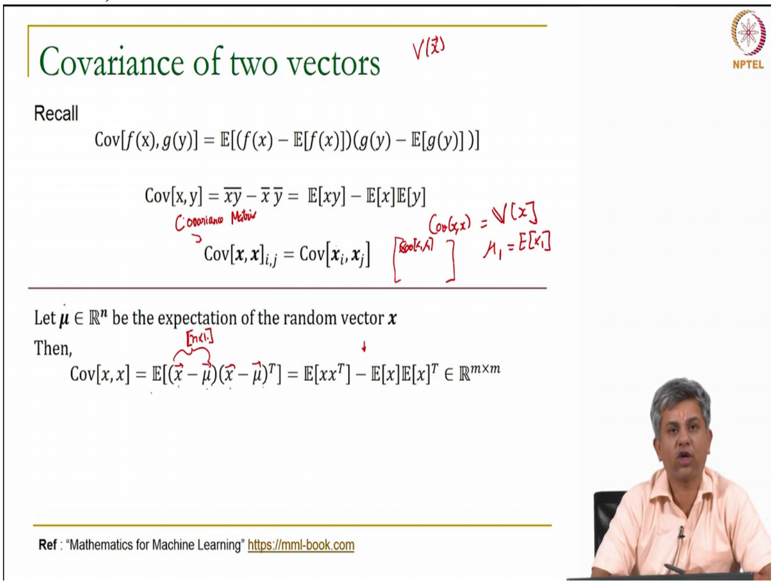
So covariance of x simply by extending this is x minus mu. Remember now, x is a vector, x minus mu is also a vector

(Refer Slide Time: 03:09)



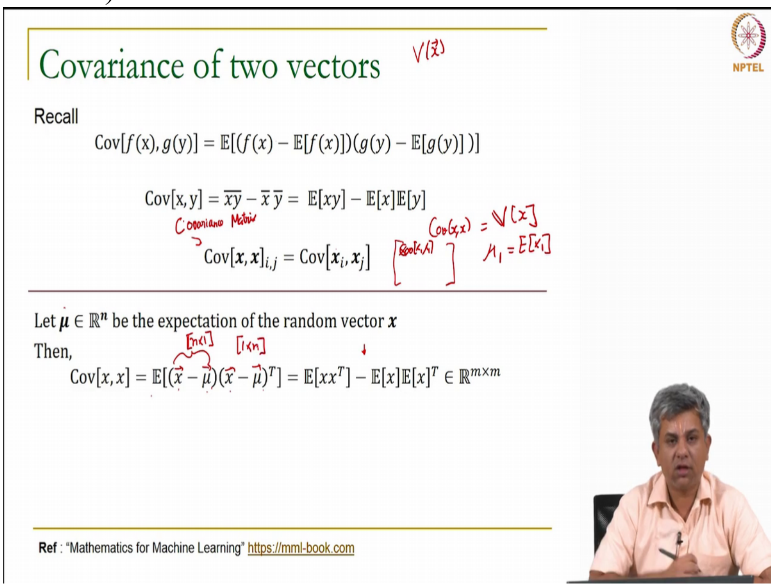
and you are taking a transpose, Ok. Why are we taking a transpose? What is the size of this? This is a n cross 1

(Refer Slide Time: 03:21)



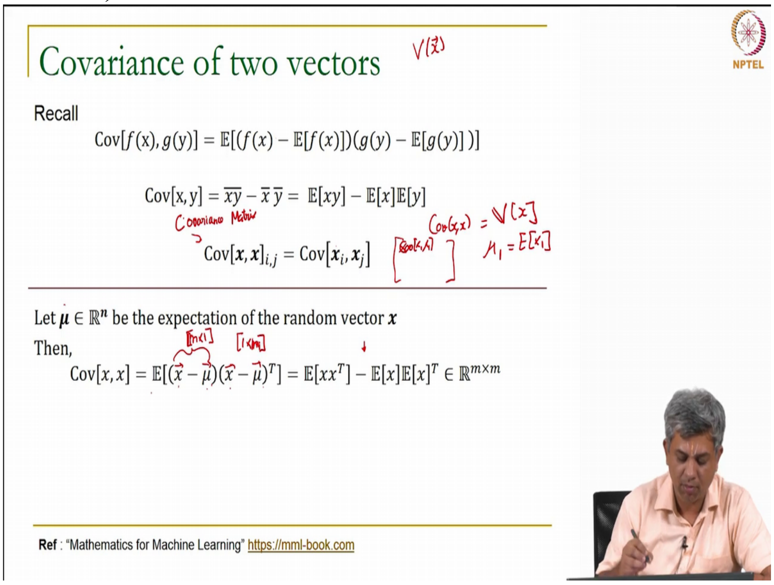
vector. Suppose I take a transpose,

(Refer Slide Time: 03:25)



this is 1 cross n. Therefore what you will get is I have written m cross m, so this should be m. And this can be m

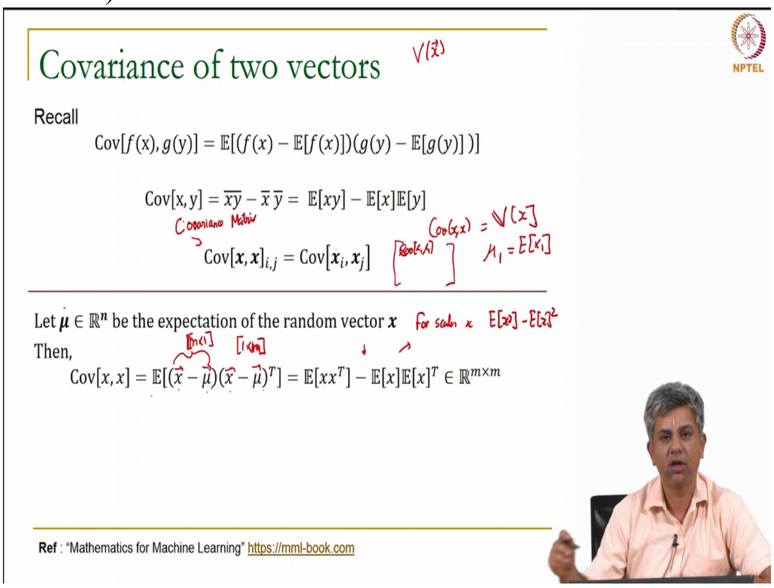
(Refer Slide Time: 03:33)



also, Ok. So covariance of x comma x can now be defined in this way. Expectation of x x transpose minus expectation of x multiplied by expectation of x transpose, Ok.

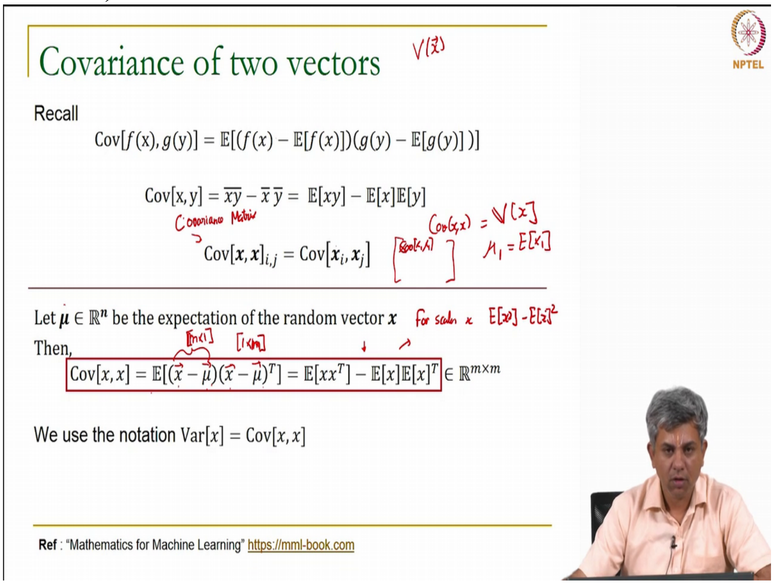
Remember this, if x is a scalar this simply comes to expectation of x square minus expectation of x the whole square.

(Refer Slide Time: 04:00)



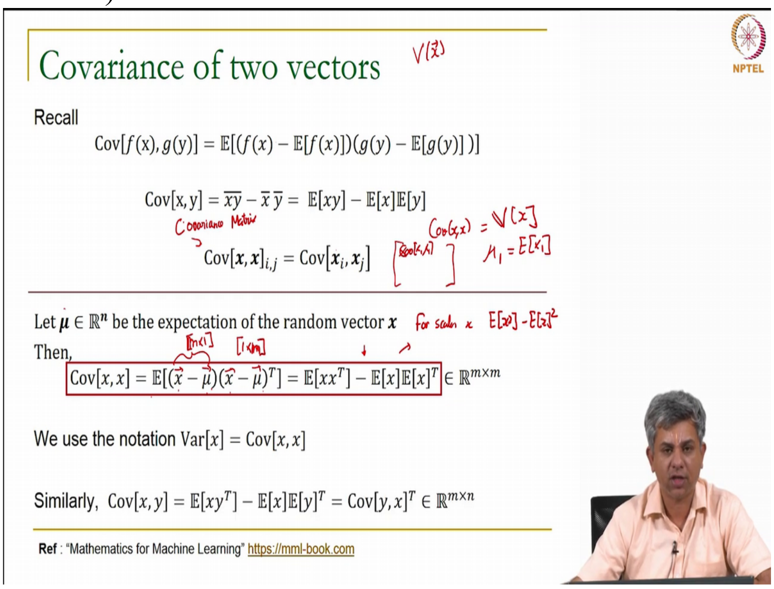
This we have seen before also. As I

(Refer Slide Time: 04:03)



said before we used the notation that variance of x is equal to covariance of

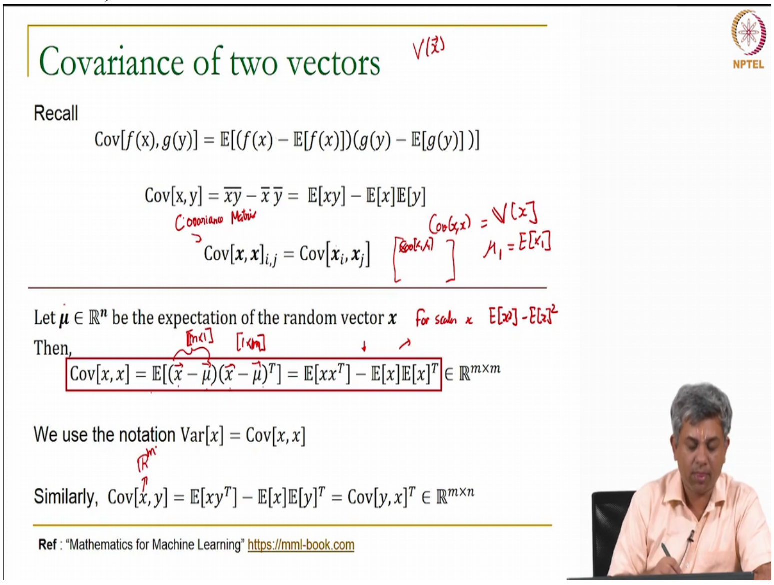
(Refer Slide Time: 04:07)



x comma x.

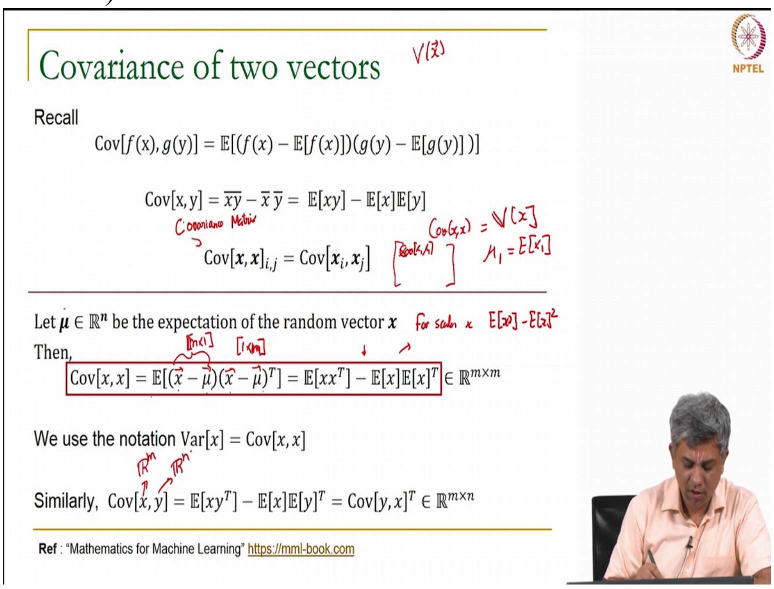
Now similarly, just very, very similar to this idea, we can now define covariance of x comma y. Now let us say x is in R

(Refer Slide Time: 04:19)



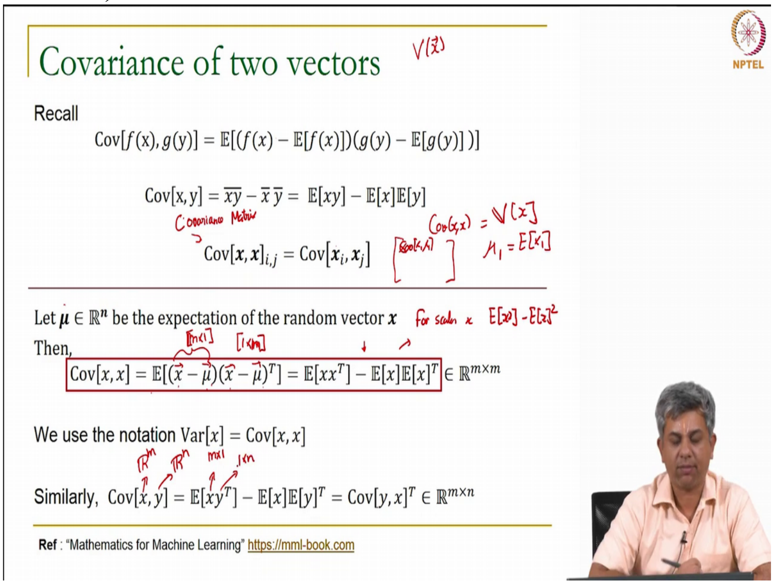
m, Ok that is, and y belongs to R n

(Refer Slide Time: 04:23)



which means that x is now a m cross 1 vector, y is a n cross 1 vector. So y transpose is now

(Refer Slide Time: 04:34)

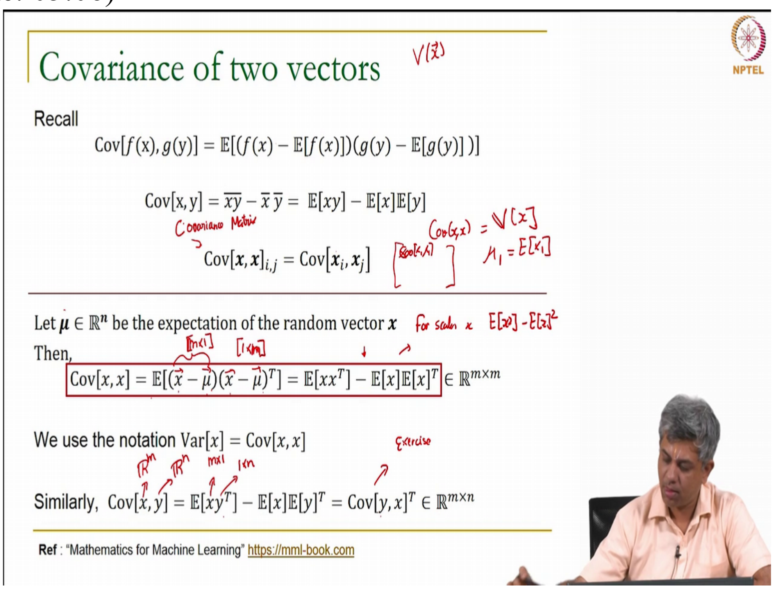


1 cross n, Ok.

So we would write instead of x x transpose, we would write expectation of x y transpose minus expectation of x multiplied by expectation of y transpose, this is simply a definition. This is just a way we define covariance of x comma y.

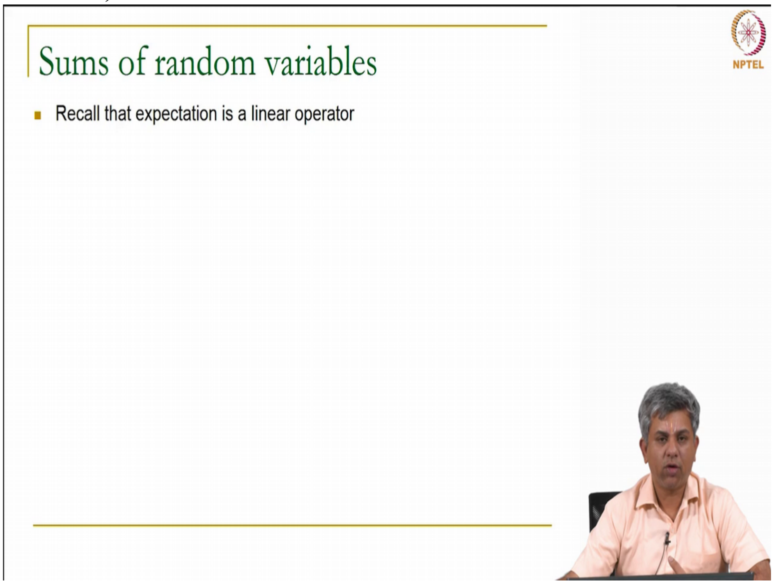
You can show easily that covariance of x comma y is the same as covariance of y comma x the whole transpose. I would suggest that you try this as a exercise.

(Refer Slide Time: 05:06)



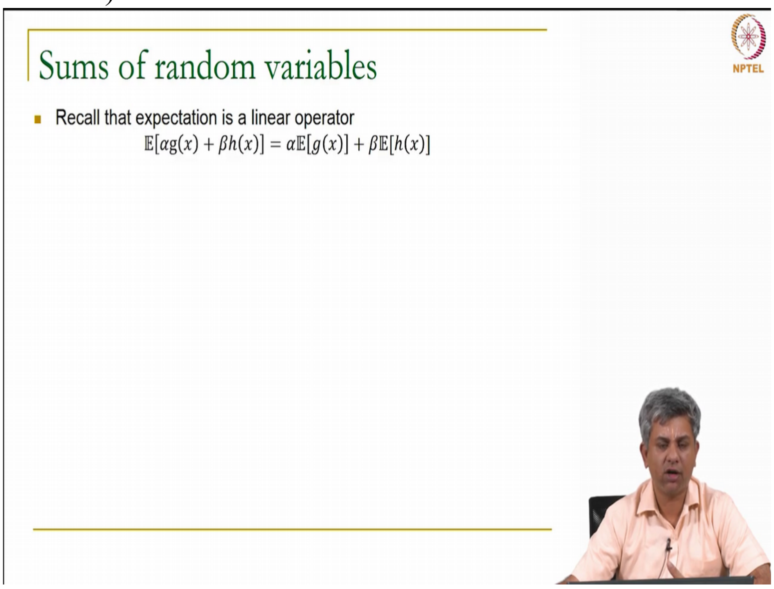
Now another idea that we are going

(Refer Slide Time: 05:14)



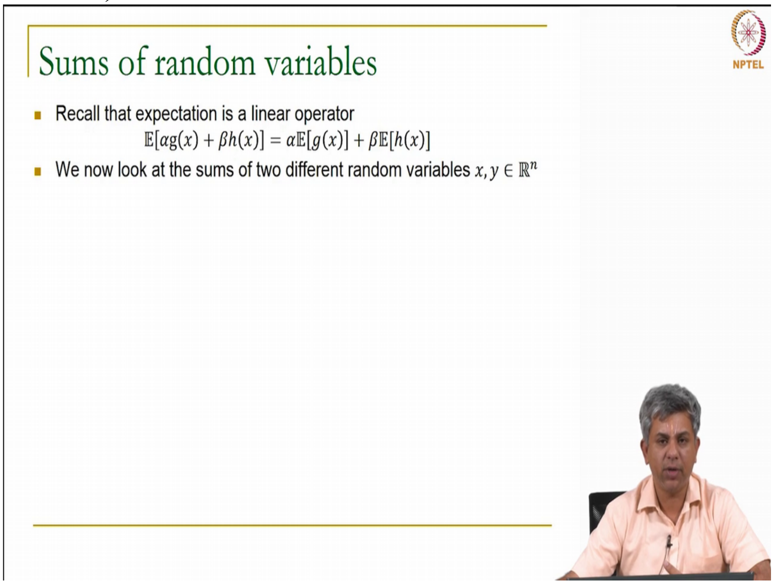
to look at is the sums of two random variables. Remember that we had

(Refer Slide Time: 05:18)



already discussed that expectation is a linear operator so that if you have expectation of alpha g plus beta h, it is going to be alpha times expectation g

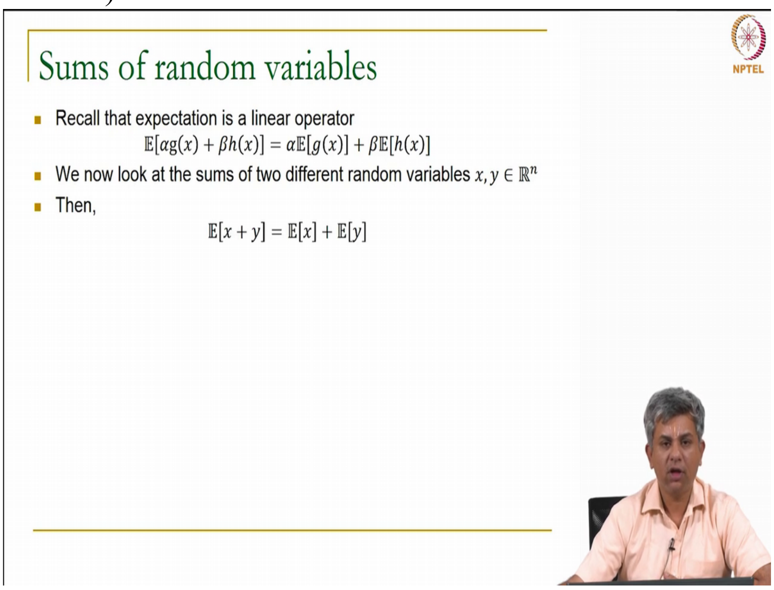
(Refer Slide Time: 05:27)



plus beta times expectation h.

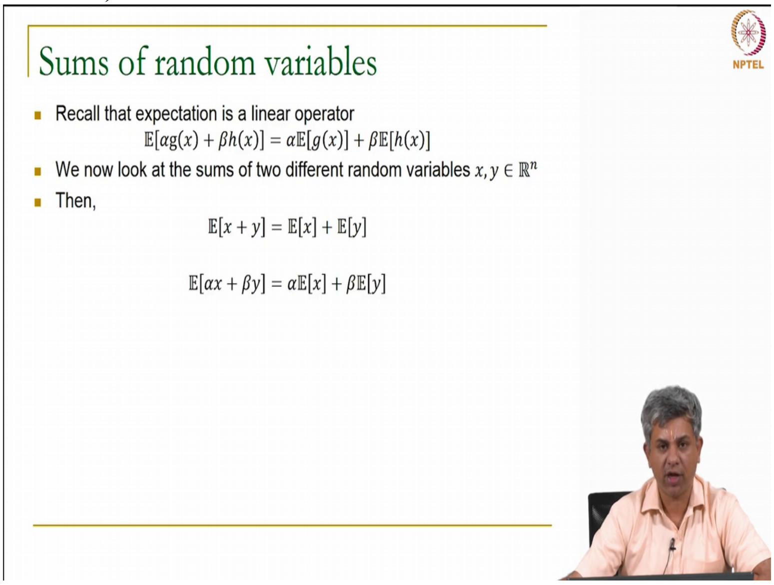
Now we are going to extend this idea to that of two random variables, two different random variables let us say x and y both are now of the same size n cross 1,

(Refer Slide Time: 05:41)



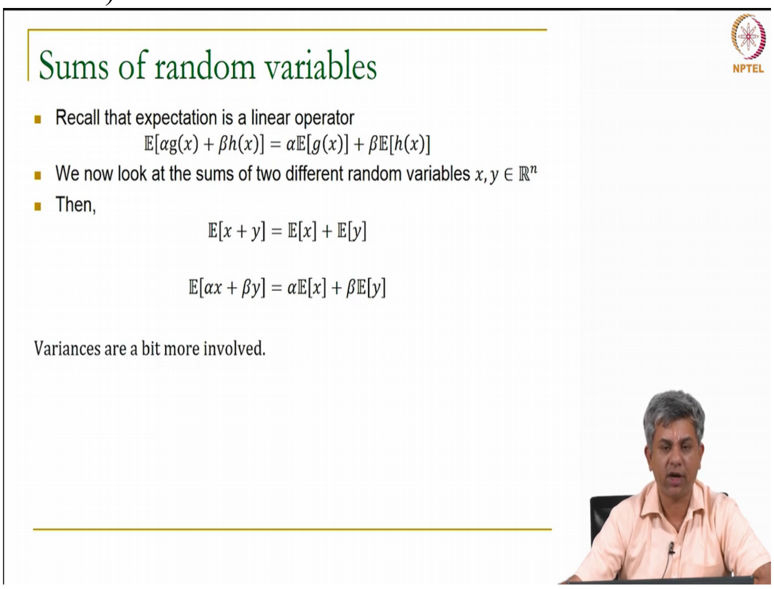
then we know that since expectation is linear, expectation of x plus y is simply going to be expectation of x plus expectation of y.

(Refer Slide Time: 05:48)



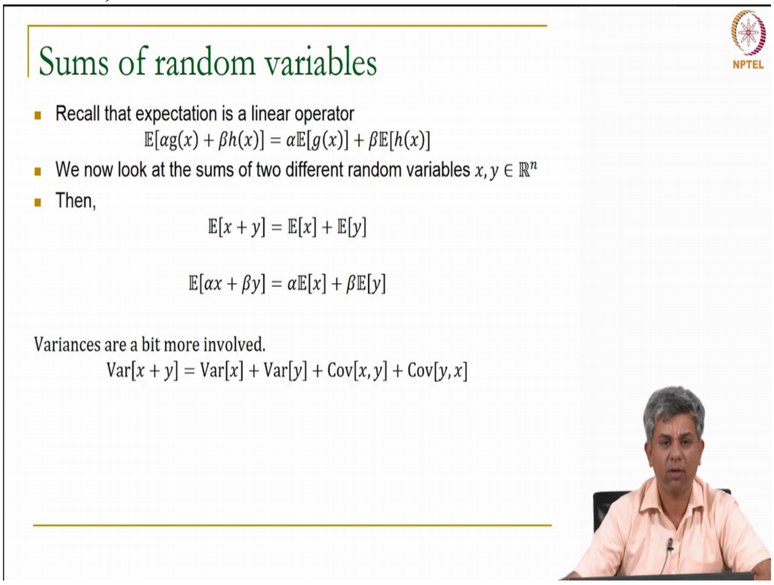
Similarly, expectation of alpha x plus beta y is going to be alpha times expectation x plus beta

(Refer Slide Time: 05:55)



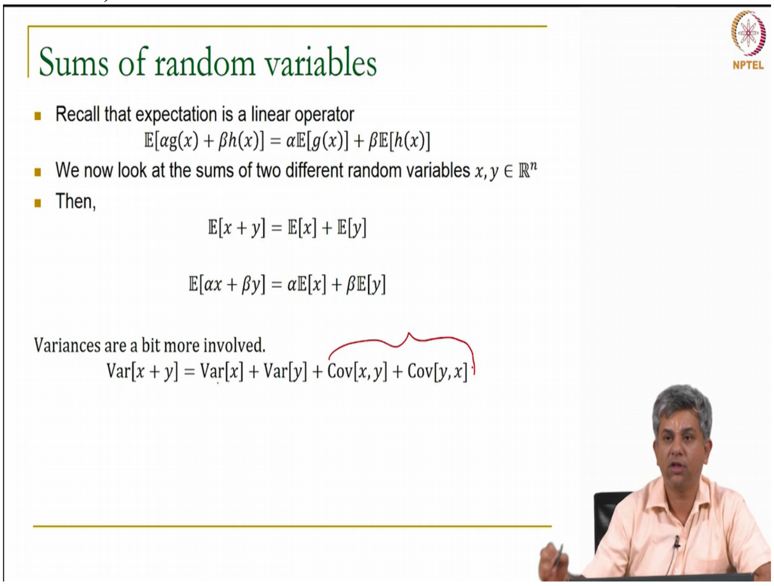
times expectation y. Variances however behave a slightly more differently,

(Refer Slide Time: 06:01)



Ok. So variation of x plus y, variance of x plus y is not simply variance of x plus variance of y. You have

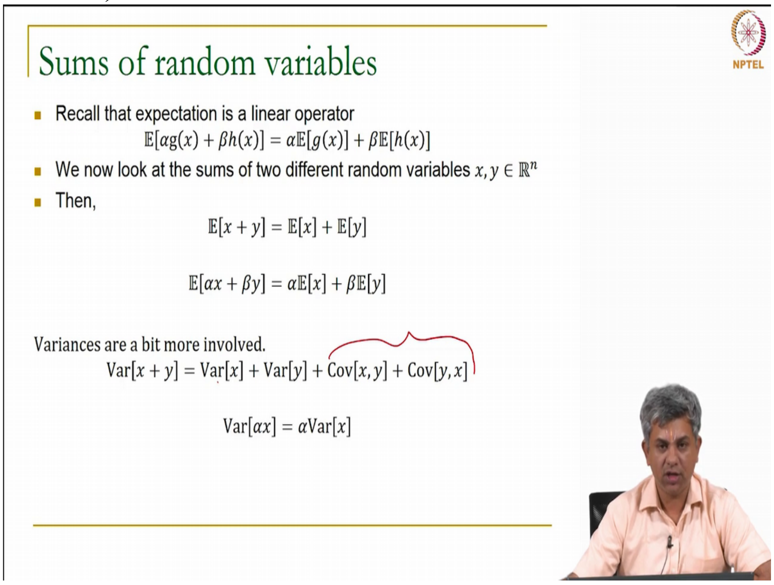
(Refer Slide Time: 06:11)



these two additional terms as well, Ok. You have these covariance terms, covariance of x comma y and covariance of y comma x.

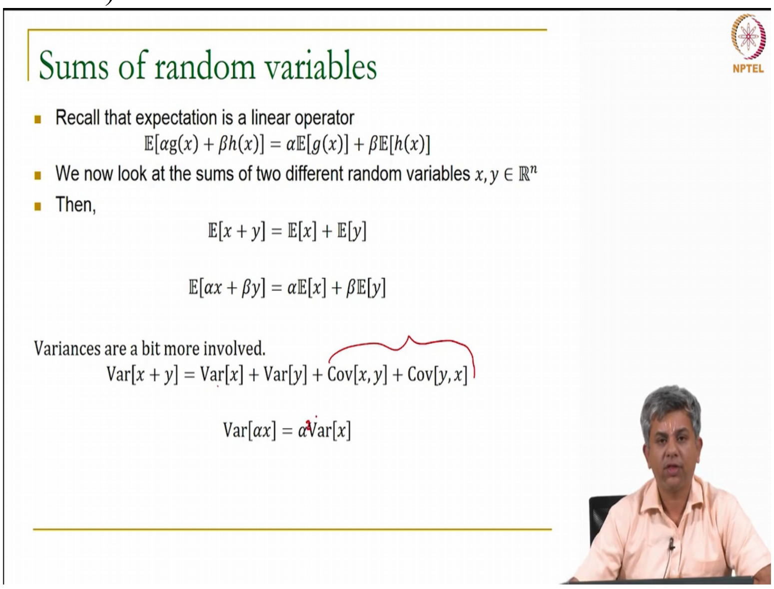
These two are not equal. They are the transposes of

(Refer Slide Time: 06:22)



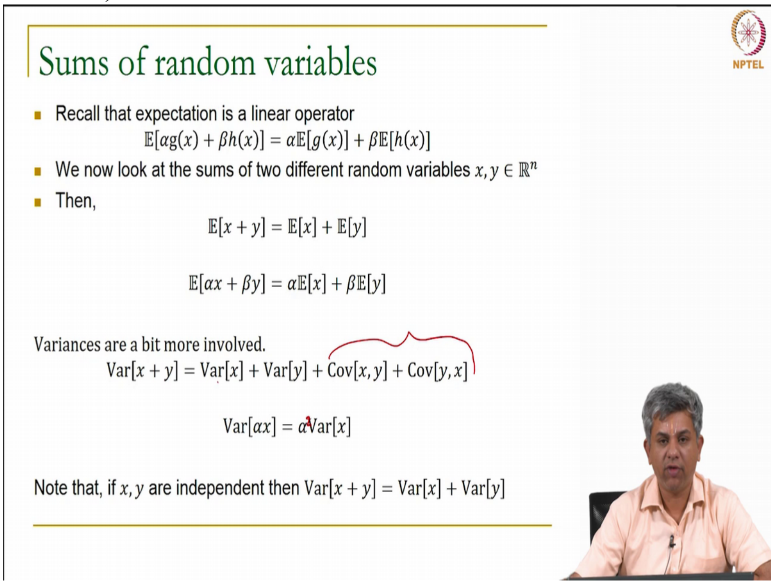
each other. If you take variance of alpha x, it should be alpha square, sorry,

(Refer Slide Time: 06:30)



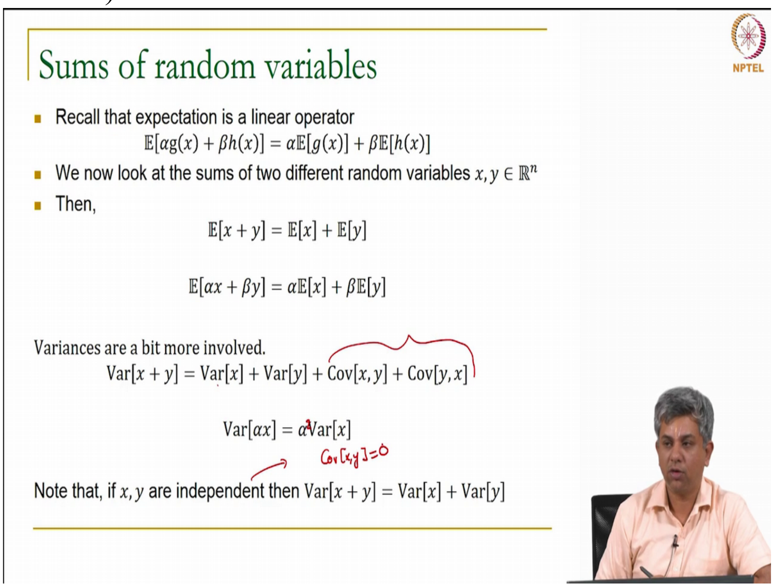
it is alpha square times

(Refer Slide Time: 06:33)



variance of x. Now notice that if x and y are independent, Ok if x and y are independent we had this discussion in an earlier slide, that if x and y are independent then covariance of x y is 0,

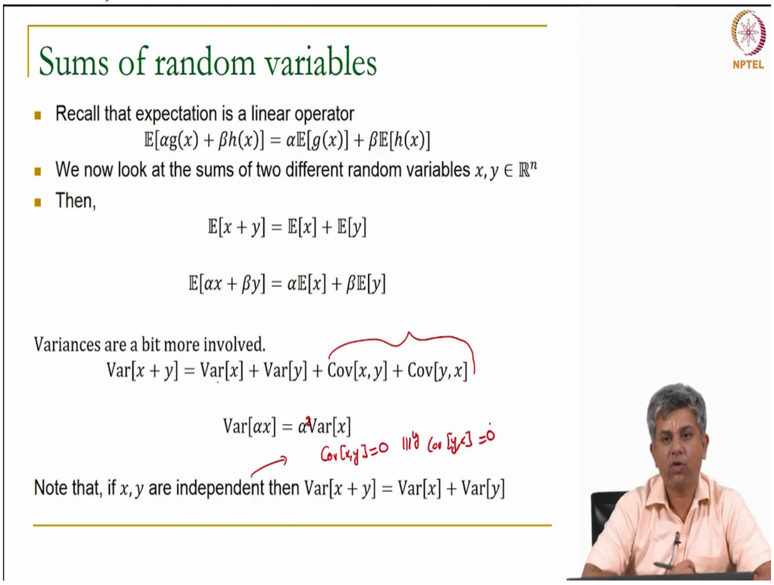
(Refer Slide Time: 06:53)



Ok.

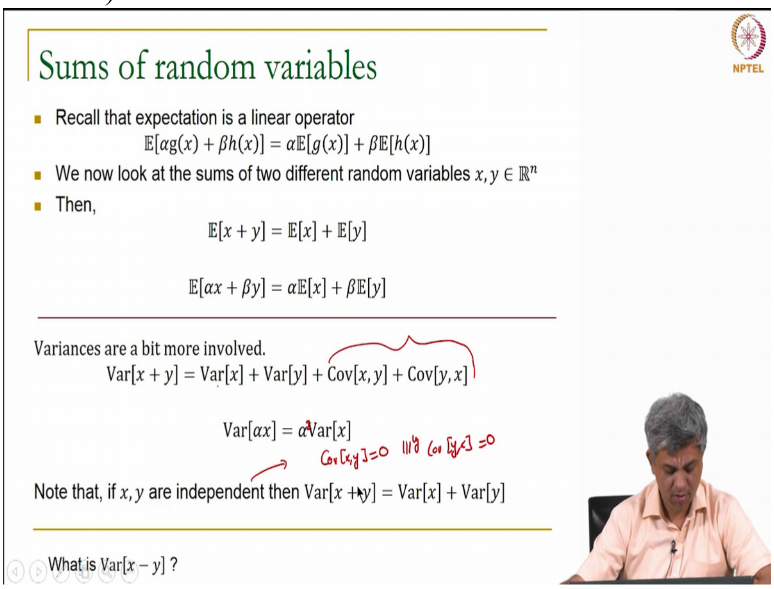
Similarly, covariance of

(Refer Slide Time: 07:00)



y x is also 0, Ok. So this gives only under the condition that x and y are independent, do we get that variance of x plus y is variance of x plus variance of y, Ok.

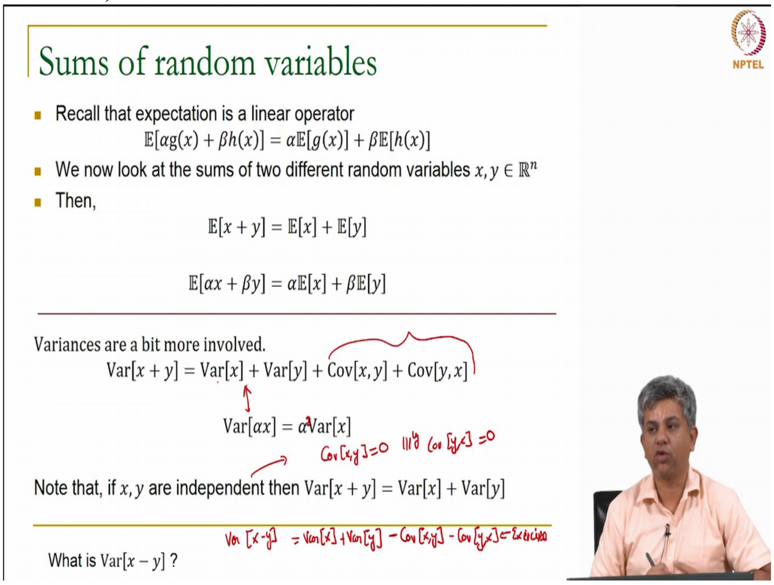
(Refer Slide Time: 07:15)



Now as an exercise please think about what happens to variance of x minus y.

So if combine these two relationships, I would again suggest this as an exercise. I will give you the final expression. So variance of x minus y will be variance of x plus variance of y, notice the plus, it is not minus, minus covariance of x y minus covariance of y x. Again I would suggest this as a quick exercise for you to try

(Refer Slide Time: 07:55)

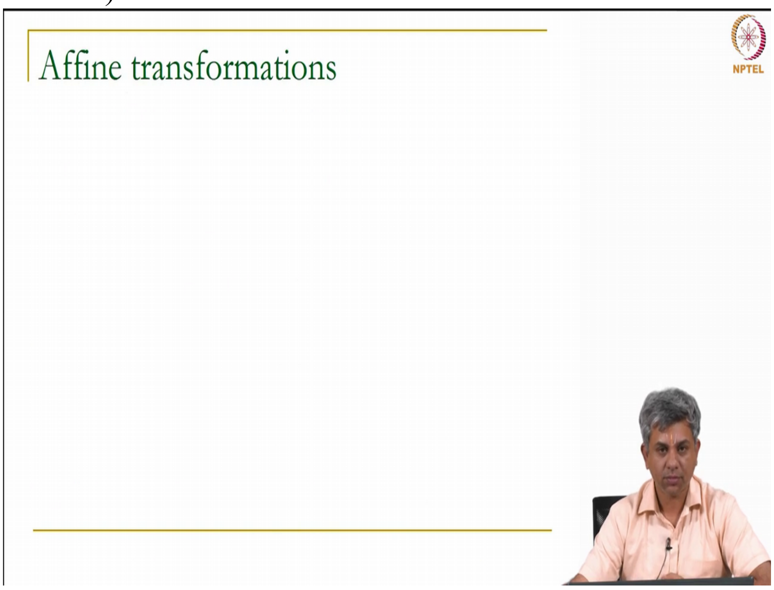


out, Ok.

If you ignore the minus parts, minus covariance x y etc, notice that even if you have the difference of two variables, the errors will still add, Ok. Variance is like the error, Ok. It is like the variation.

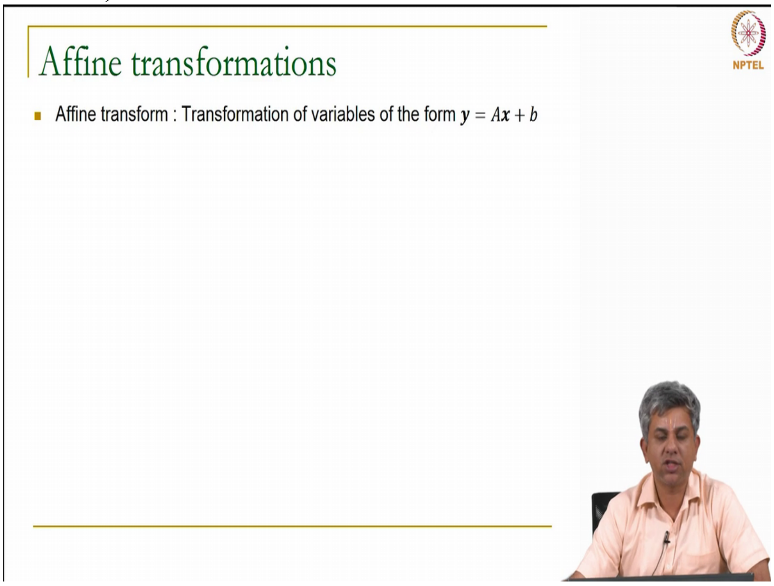
You would have seen this in simple, you know experimental measurements perhaps a little bit before that if you have two variables, you know I make one length minus another length it does not mean the error subtract, the error still add because the errors go as the variance, Ok. So therefore the errors add there.

(Refer Slide Time: 08:26)



The final idea that I would like

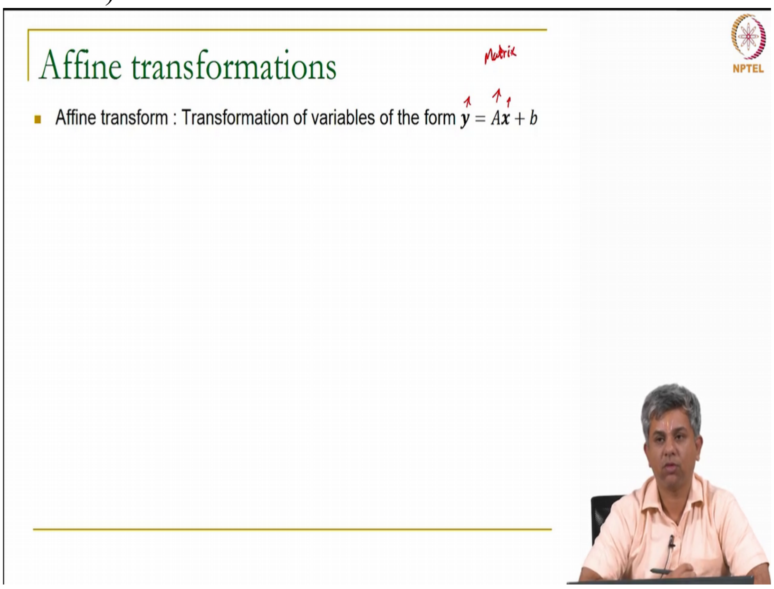
(Refer Slide Time: 08:29)



to discuss in this video is that of an affine transformation. An affine transformation is nothing but another name for a linear transformation.

So let us say you have two random variables or two variables x and y. x is a vector, y is a vector. A is a matrix.

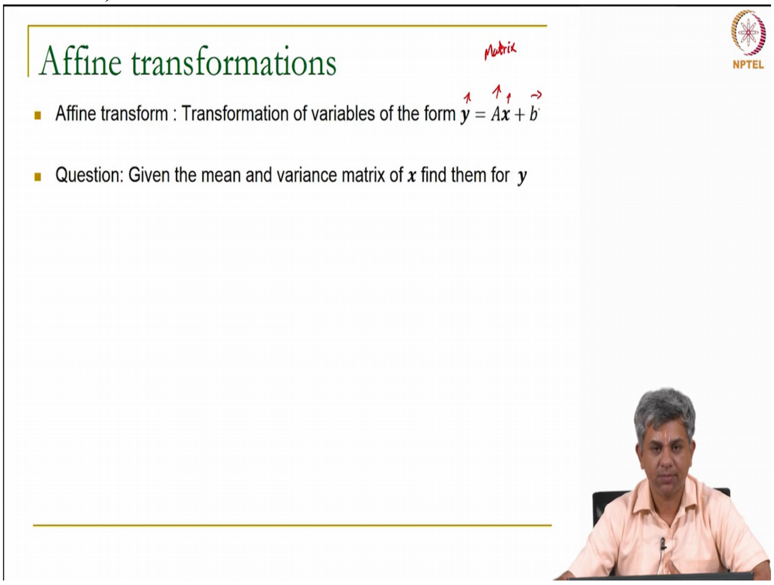
(Refer Slide Time: 08:48)



You can transform, y can be seen as x with a linear transform, b is also a vector here of course in order for our dimensions to match, Ok.

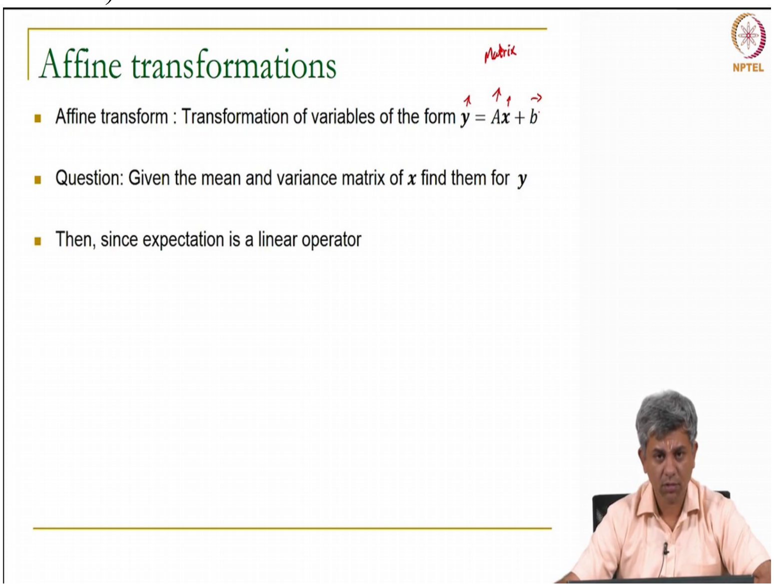
So x is a vector, b is a vector, A is a matrix and y is another vector. So y is A x plus b. Now if I have this,

(Refer Slide Time: 09:08)



suppose I know the mean of x, or suppose I know the variance of x, can I find out the mean and variance of y? It turns out that

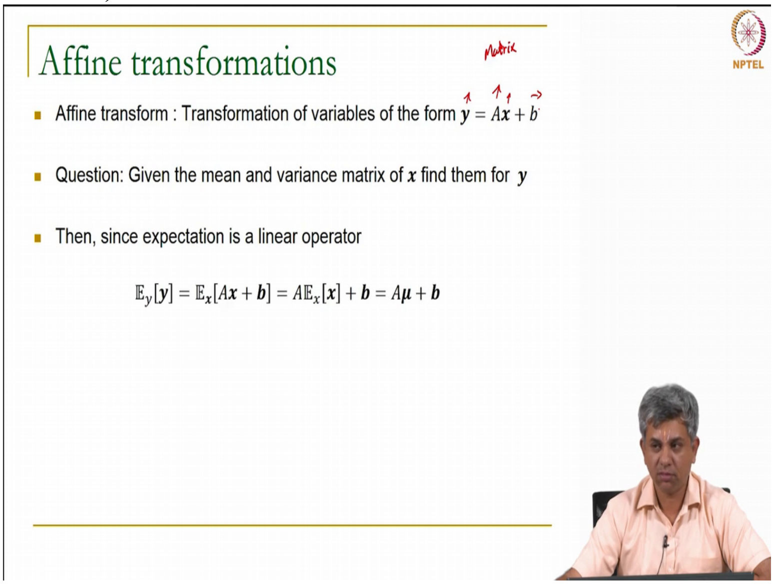
(Refer Slide Time: 09:17)



there are mathematical relationships. Expectation works out as we expect, Ok.

So

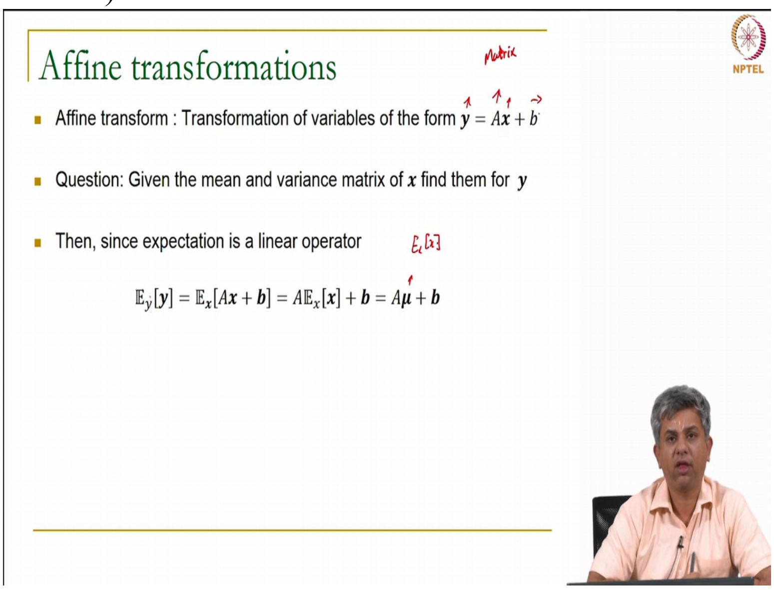
(Refer Slide Time: 09:23)



expectation of y, we will write the expressions here. Please notice this, the subscript. This expectation is over the variable y. This expectation is over the variable x.

So expectation of y is expectation of A x plus b which since, expectation is linear we simply take out A which is a constant matrix, outside. A times expectation x plus b, mu of course is the notation for expectation of

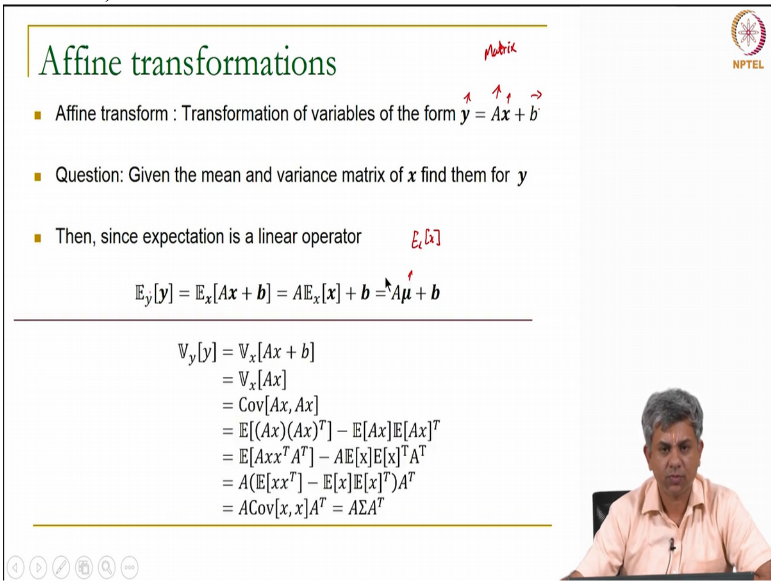
(Refer Slide Time: 09:55)



x, Ok.

So this is quite simple as far the expectation is concerned.

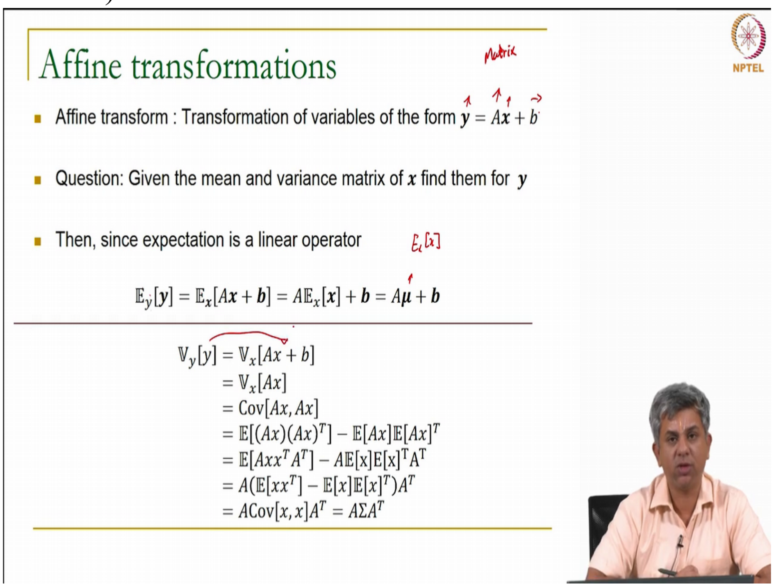
(Refer Slide Time: 10:01)



Now variance, we will go through this slowly is once again slightly more involved, Ok. Let us look at it step by step.

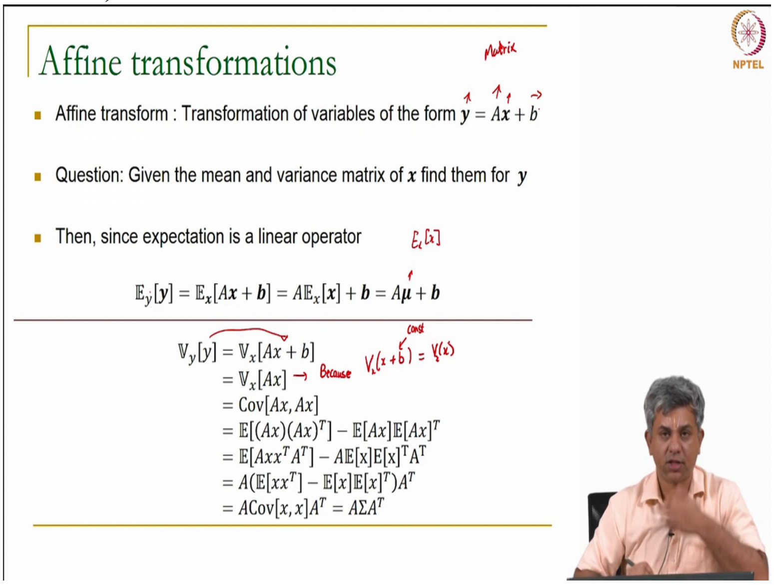
So variance of y is, of course y is defined as A x plus b so

(Refer Slide Time: 10:15)



this is fine. Variance of A x plus b is same as variance of x because variance of something x plus a constant is the same as

(Refer Slide Time: 10:34)



variance of x.

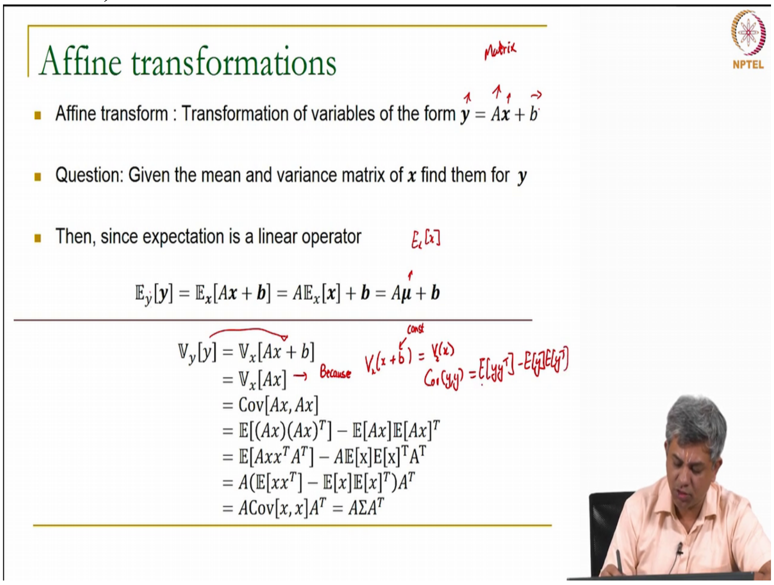
Why is that? You can prove this very easily mathematically but let us just look at it very simply from a physical point of view.

Variance measures the difference from the mean, Ok or the distance from the mean. If I change the variable by a constant the mean will also go up by a constant and that constant subtracts out.

We are only looking at the difference from the mean and we are not looking at the actual value of the variable, Ok. Since that is the case so whether I add b or not, I am going to get the same variance, Ok.

Now variance of A x was defined as covariance of the variable with itself, A x comma A x. Now we can write it out based on the definition that we had earlier. Remember that covariance of the variable with itself was expectation of y y transpose minus expectation of y expectation of y transpose. That is what is

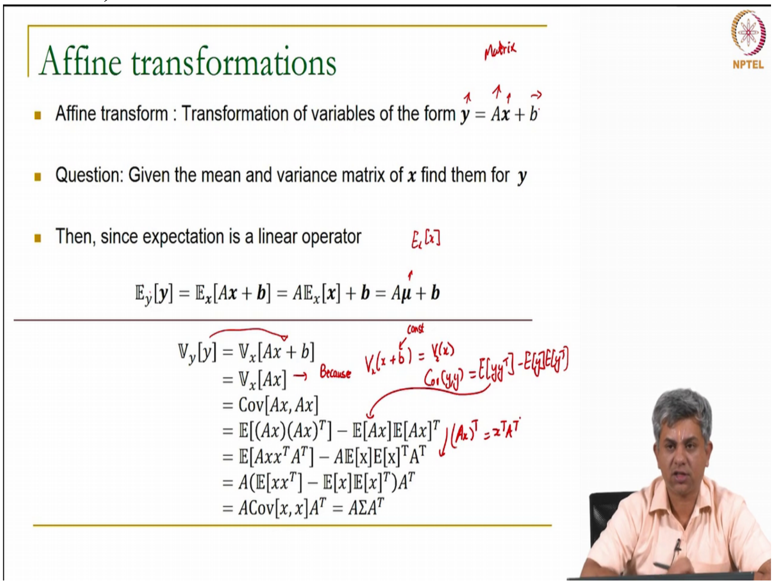
(Refer Slide Time: 11:32)



used here, Ok.

Instead of y I have A x. A x A x transpose, A x expectation of A x transpose. Ok. Now we can open this up noting the fact that A x transpose is the same as

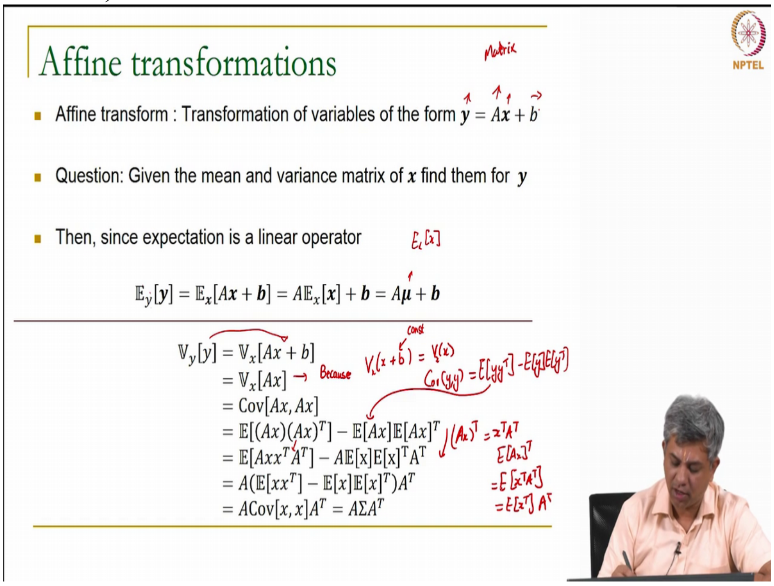
(Refer Slide Time: 11:49)



x transpose A transpose.

So that is opened up here. Now also I have used the fact that expectation of A x is A times expectation of x. Similarly expectation of A x transpose is the same as expectation of x transpose A transpose which is the same as expectation of x transpose multiplied by A transpose.

(Refer Slide Time: 12:13)

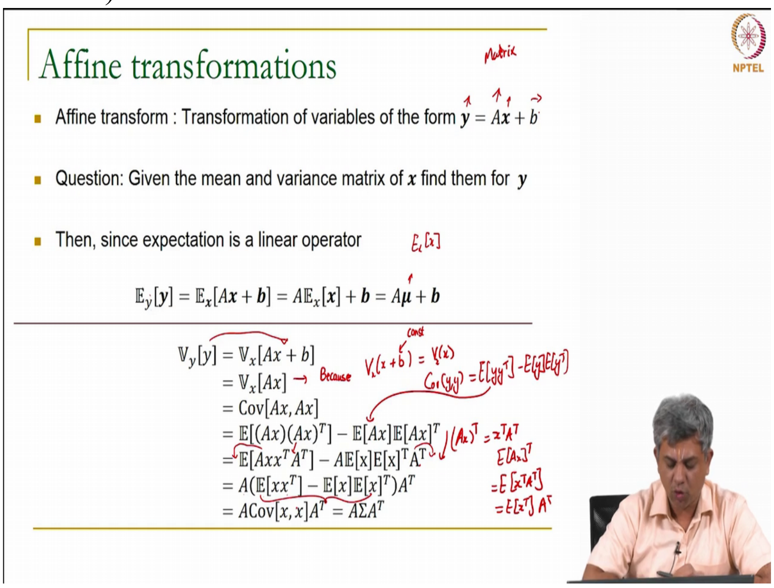


That is what is used here. Now I can bring out A from the front end and bring out A transpose from the back end, Ok. So we can do that.

So A from the front end, A transpose from the back end and we can take that out common and we get A times expectation of x x transpose minus expectation of x, expectation of x transpose. And A transpose comes out.

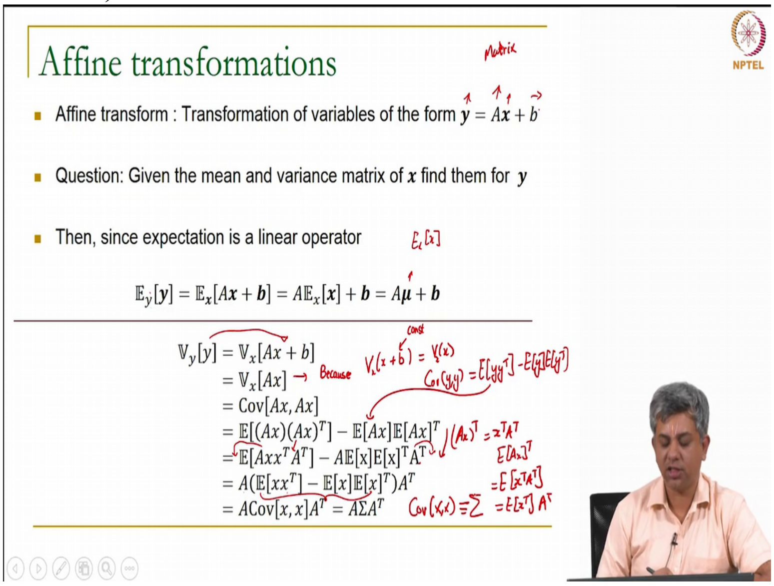
So you can now rewrite it. This is nothing but

(Refer Slide Time: 12:45)



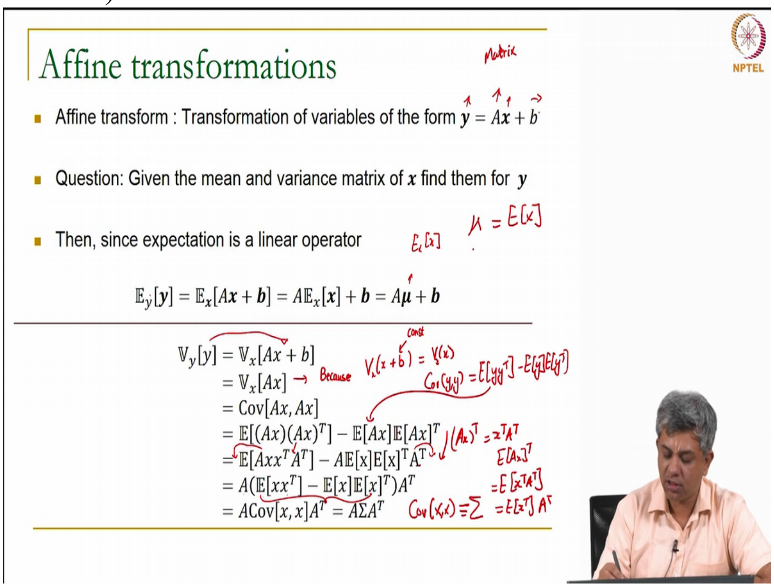
covariance of x comma x which we discussed in the previous slide. So A times covariance of x x, x comma x A transpose. Typically the covariance matrix is denoted by

(Refer Slide Time: 13:00)



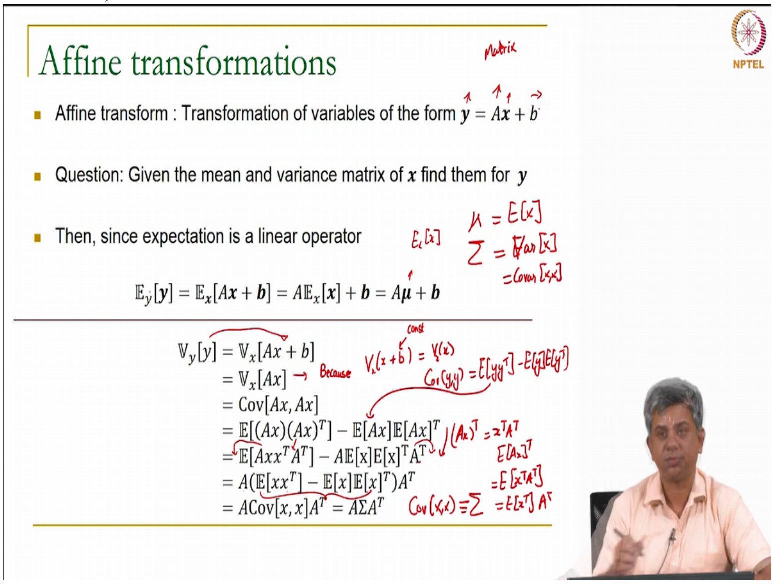
sigma just like, use the expectation of x,

(Refer Slide Time: 13:06)



capital sigma is the variance of x which is the same as covariance of

(Refer Slide Time: 13:18)



x on x.

So we have looked at a lot of mathematical relations in this video. As I said earlier you might or might not be comfortable with it. Even if you are not comfortable with it, we do not use at least this set of expressions too often. But please get comfortable by watching this a few times in case you found it unfamiliar. We will use it a little bit towards the latter half of the course, thank you.