Applied Multivariate Analysis

Prof. Amit Mitra

Prof. Shramishtha Mitra

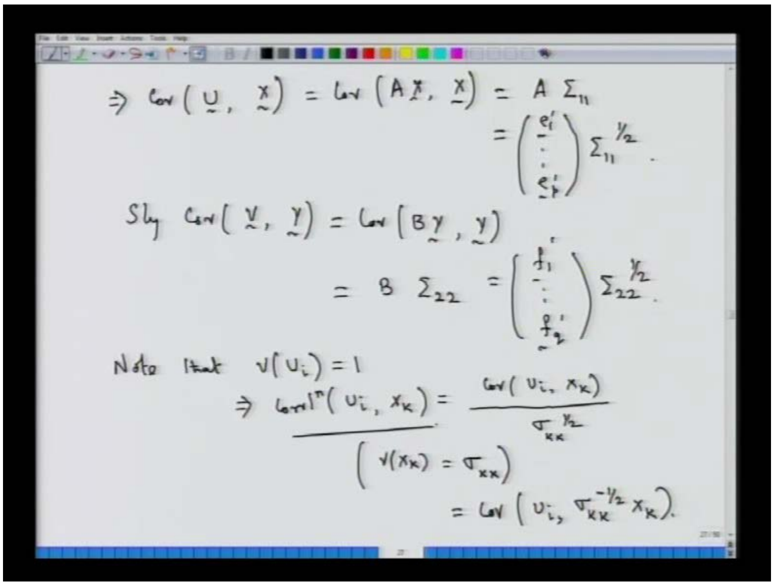
Department of Mathematics and Statistics

Indian Institute of Technology, Kanpur

Lecture No. # 42

Canonical Correlation Analysis

(Refer Slide Time: 00:35)



In this lecture we continue our discussion on canonical correlation analysis. Now, in the last lecture, at the end, what we had looked at was the following concept, that we were looking at the correlation coefficient between the canonical variables and the original variables.

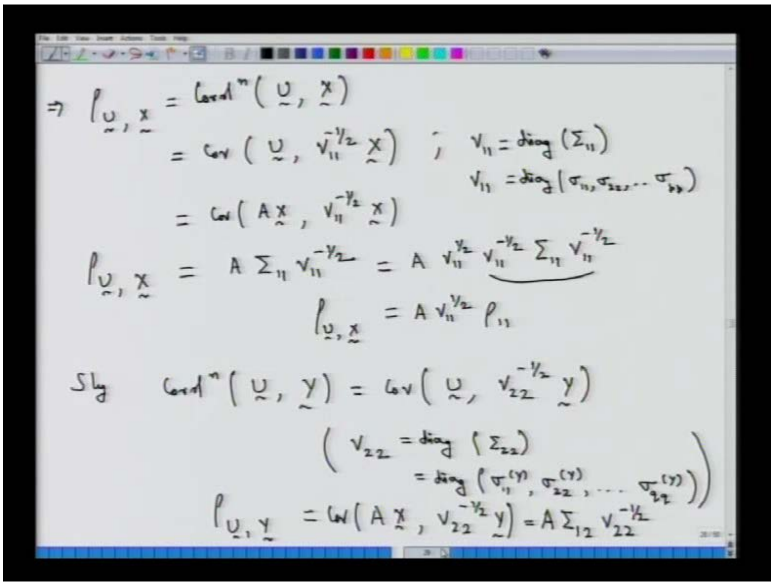
We had come up to the following point, that we had noted, that if we denote by u vector the p dimensional vector containing the canonical variables, the first component of the canonical variable pairs and x to be the vector of original random variables, then what we are having is, that the covariance between u vector and x vector is covariance between Ax and x vector, which is equal to A times sigma 1 1, which is given by this and similarly, the covariance between v vector and y vector can be given by this. And other combinations, like the covariance between u and y or v and x can be, similarly be obtained.

Now, at this particular point what we note is that variance of, we have not yet got to the correlation coefficient between the two vectors, we are, we have at the moment looked at the covariance structure. We note, that this variance of u i, because they are canonical variables, that is equal to 1. So, this would imply, that the correlation between u i and say, x k ith component of this u vector and kth component of x vector, that is going to be given by the covariance between u i and x k, which is going to be the i kth element of this particular matrix because this is the covariance matrix of u and x. And what we are talking about is between the ith component of u and the kth component of x.

Now, this would be divided by the standard deviation of the respective components. So, this will be 1 times, let us write this as sigma k k to the power half, wherein this variance of x k, we are denoting by sigma k. So, the standard deviation is this quantity.

Now, if we look at this term here, this is nothing, but the covariance between u i and sigma k k to the power minus half times x k because if we look at observing this particular constant as in here, sigma k k to the power half into this variable, here we can write this statement as covariance between u i and sigma k k to the power minus half times x k. So, correlation between these two variables is just equal to this.

(Refer Slide Time: 03:16)



So, using this particular fact what we can say is, the correlation matrix between u vector and x vector, which is the correlation matrix between u vector and the x vector, would be the same as the covariance matrix of u vector, that multiplied by v 1 1 to the power minus half times x, where v 1 1 is basically, what we have defined earlier, this diagonal matrix containing the diagonal elements of sigma 1 1, that is, this is the diagonal matrix containing sigma 1 1, sigma 2 2, sigma k k and sigma p p as the diagonal entries of that particular matrix. So, if we look at this v 1 1 to the power minus half, it is basically what is serving the purpose of sigma k k to the power minus half for these two univariate random variables, and hence this one can easily be computed.

Now, by the previous calculations what we had shown was, that u we had denoted by A times A is that p by p matrix times the x vector, where A is e 1 prime e 2 prime e p prime sigma 1 1 to the power minus half. So, we will use that particular fact and write this u as A x. So, this is the covariance matrix of A x and v 1 1 to the power minus half times x and thus, this is nothing, but equal to A times the covariance matrix of x, which is sigma 1 1 times v 1 1 to the power minus half. So, this is the final form as such of the correlation matrix between the cannonical variables and xs where u is are the cannonical variables using the linear combination of the elements of xs only.

Now, if we look at this form, one can alternatively write it in terms of the correlation matrix. In what way we can write this expression here? As A times v 1 1 to the power plus half into v 1 1 to the power minus half times sigma 1 1 into v 1 1 to the power minus half. Now, note that this particular matrix is nothing, but the correlation matrix of the x random vector. So, this is v 1 1 to the power half times this rho 1 1, it is an alternate equivalent form of this correlation matrix between u and x.

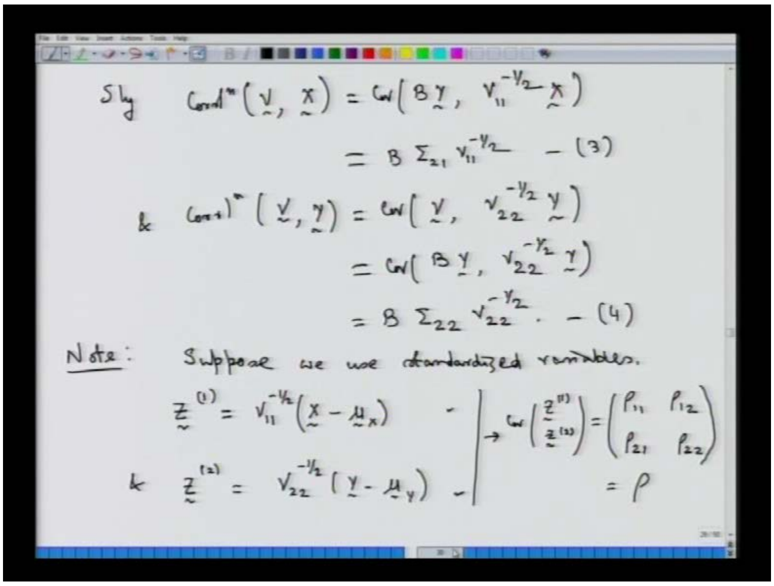
So, in a similar way we can look at, actually we can compute the correlations between the other pairs, so we derived this one explicitly. So, if one is looking at, similarly correlation between u and y, note that here we had written this v vector as b times y and u as A times x, so when we are looking at this correlation matrix between u and y, the other vector, so this would be similar to this one.

The covariance matrix between u and by in y, what we are going to do is to take v 2 2 to the power minus half times y. So, this correlation matrix will be equal to this covariance matrix, wherein as before we take this v 2 2 to be the matrix with diagonal entries as the diagonal elements of this sigma 2 2 matrix, that is, this is sigma 1 1 corresponding to the y variable sigma 2 2 to the power, not to the power, sigma 2 2 y.

So, this is the 2 2th element in this sigma 2 2 matrix, which denotes the variance of y 2 in this y vector and this is sigma q q because y is q dimensional and we will have this as this and this can be easily be computed because this is the covariance in place of u. Once again what we write as A x and this is v 2 2 the power minus half times y vector. So, this covariance matrix is going to be A times the covariance matrix of x and y. So, that is equal to sigma 1 2 times this v 2 2 to the power minus half.

Once again if we want to write this, this is what, this is correlation matrix between u and y. So, if we want to write this in terms of the correlation matrix, one can do. So, by introducing the appropriate diagonal matrix in order to have write this rho sigma 1 1 in terms of rho 1, I am sorry, sigma 1 2 in terms of rho 1 2.

(Refer Slide Time: 08:06)



So, in a similar way, what once again, what we can write is covariance. This correlation between the, correlation between the other components, which is correlation between v and x, this would be the covariance between, now v is B times y and this x we will write as v 1 1 to the power minus half. This is the covariance now times this x and thus, this is going to be B times covariance between y and x, that would be sigma 2 1 times v 1 1 to the power minus half. And the last component, that what one can write is correlation matrix between v and y that we have not looked at. So, this is going to be the covariance matrix between v and y, would be standardized to have unit variance.

So, this we require v 2 2 to the power minus half times y and thus, this is going to be same as covariance between B y and v 2 2 to the power minus half times y, this is going to be equal to B times sigma 2 2 into v 2 2 to the power minus half. So, we have computed all the four possible correlations that may exist. So, one is this u and x, which is the final expression of that, is given by this. So, this is one, this is the 2nd expression where we are looking at u and y; this is the 3rd expression and this is the 4th expression

So, this 1, 2, 3 and 4 gives all the correlation matrices between the original and the derived or the canonical variables. Now, we put it as a note here, an important thing. Suppose we are using standardized variables, that is, instead of the covariance matrix we are going to work with correlation matrix. Then, if we are looking at the correlation matrix and then, the canonical variables derived from the correlation matrix, if we compute the correlation matrices of the canonical variables derived from such a matrix and the original standardized variables, what would be the form of those correlation matrices and how does those compare with the original correlation matrices, which we have just now obtained as 1, 2, 3 and 4?

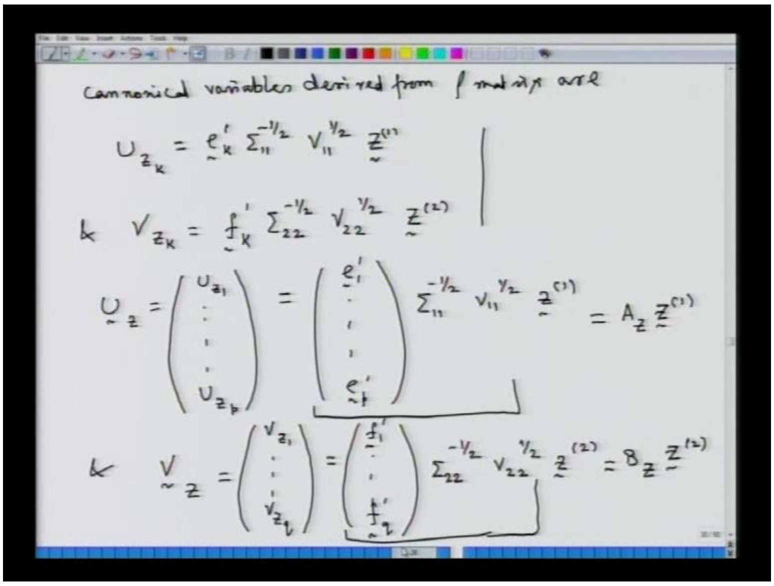
So, what we are now going to see? We had seen earlier was a thing, that we had noted was, that if we are looking at the covariance matrix between x and y given by sigma 1 1, sigma 1 2, sigma 2 1 and sigma 2 2, then if we are deriving the canonical correlation matrix from that matrix, the canonical correlation coefficients from that matrix are going to be same as that of the canonical correlation coefficient derived from the rho matrix. Now, rho matrix is of course, the covariance matrix between the standardized variables.

Now, we are going to see the similar result here. Now, suppose we use standardized variables here, standardized variables, say the standardized variables, I denote by Z 1, which is the standardized form of x vector. So, how we are going to standardize that? Suppose we have mu x as the mean vector of x and then this has to be multiplied by v 1 1 to the power minus half and the 2nd component, the 2nd component vector, which is y is standardized as, standardized as y minus mu y times v 2 2 to the power minus half. So, these two are the standardized variables.

Now, the covariance matrix of this standardized set, here Z 1 and Z 2, this is going to be the covariance matrix of Z 1 here. What is the covariance matrix of this element here? It is just equal to rho 1 1, the covariance matrix of this element here would just be rho 2 2, which is the correlation matrix corresponding to the y vector and this is going to be the cross rho 1 2, which is the correlation covariance matrix between Z 1 and Z 2 and similarly, this is rho 2 1, which is the transpose of rho 1 2.

Now, if we look at the canonical correlation coefficients to be derived from this, this is the rho matrix canonical correlation coefficient derived from rho matrix or the original sigma matrix is the same.

(Refer Slide Time: 12:58)



The canonical variables are given by the following. So, canonical variables derived from rho matrix are given by the following, let us denote that by u Z k. To denote, that this is the kth cannonical variable computed using the standardized variable, which we have denoted by Z. So, this is given by e k prime, then sigma 1 1 to the power minus half times this v 1 1 to the power plus half times z 1. So, this is what is the form of canonical variables derived from the rho matrices. Canonical correlation coefficients remain the same, as we have said, and similarly, the v component.

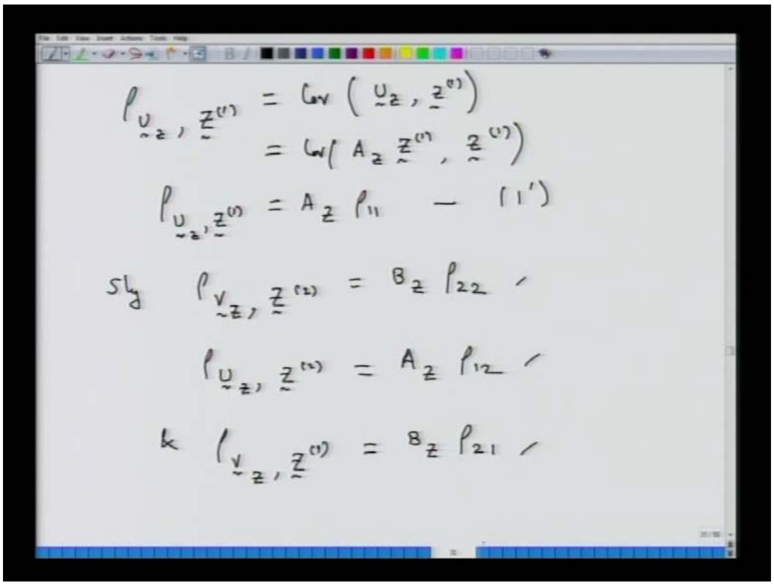
If we are denoting that by v z k the kth canonical variable, which is going to be derived from Z 2, this is going to be given by f k prime sigma 2 2 to the power minus half. Then, this is v 2 2 to the power plus half times. This is Z 1 vector and this is Z 2 vector, we have denoted by upper 1 entry. So, I will just stick to that, this is Z 1 and this is Z 2, so these are the two canonical variables derived from the rho matrix if we write this u z vector containing all these entries. So, this is going to be u Z 1, u Z 2, u Z p because the starting order, there was p dimensional.

So, this is going to be given by in a similar way to what we had done for the canonical variables derived from the sigma matrix. We will write this as e 1 prime e 2 prime e p prime times this sigma 1 1 to the power minus half times v 1 1 to the power half Z 1. So, we just denote this term by the quantity, which we are saying, that it is just equal to A Z matrix, where A Z matrix is this entire matrix up to this times this Z 1 vector. And similarly, we will denote this v Z vector to contain these q canonical variables; this is v Z1 v Z 2 up to v Z q. So, this vector is going to be given by this f 1 prime, f 2 prime, f q prime times sigma 2 2 to the power minus half v 2 2 to the power plus half times this z 2 vector.

So, we will denote once again, starting from this matrix up to this matrix here by a matrix, which we denote by v z. So, this v z matrix times this z 2 vector, so this is what we are going to get if we are looking at the standardized variables.

Now, similar to what we were calculating when we were looking at the canonical variables computed from the covariance matrix, this rho u x and things like that. The four expressions 1, 2, 3 and 4, we can compute the similar things when we are looking at the canonical variables derived from the rho matrix.

(Refer Slide Time: 16:50)



What we are going to get is the following, that this rho u z, this is the vector of the canonical variable derived from the rho matrix and Z 1 vector, which is the vector of standardized variables. This is going to be, now when we are looking at standardized variables these are going to have the matrix, the covariance matrix, which is going to have identity elements in its diagonals. So, we can just write, that this correlation between u Z and Z 1 is just equal to the covariance between u Z and Z 1. So, that is as simple as that.

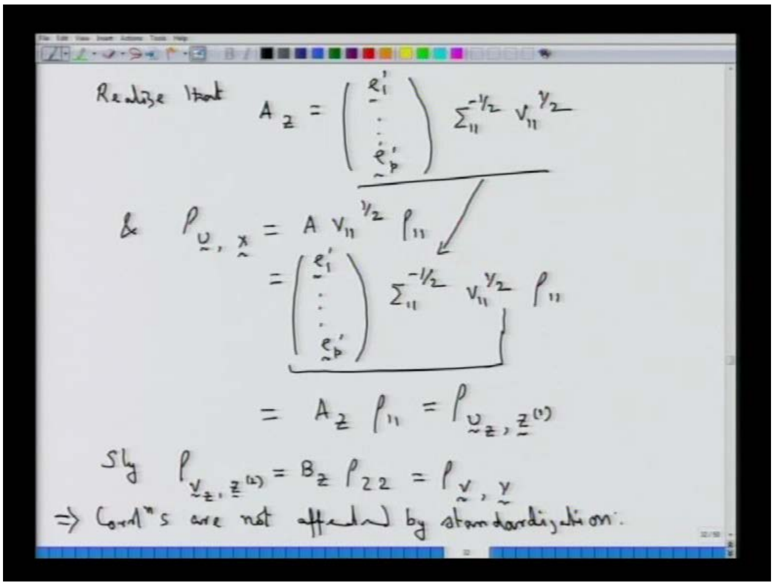
And what we have written for u Z is A Z matrix times this Z 1 vector, so this is the covariance between these two terms and which is very simple. So, this write that as A Z times the covariance matrix between Z 1 and Z 1, what is that, that is equal to rho 1 1. So, this is say, 1 prime, this is the corresponding term to the correlation matrix between u, the vector of canonical variables and the original variables. Now, the original variables are standardized and hence, we have got this as the form.

Now, similarly we can write, that this correlation matrix between v z and Z 2, this is going to be given by B Z times just rho 2 2. So, this is corresponding to that, this is v and this Z 2, we will also be able to write, what is going to be the correlation matrix between u Z and Z 2, what would that be? u Z is given by A Z times Z 1 and Z 2 is as it is. So, what we will have is A Z times the correlation, the covariance matrix between Z 1 and Z 2, which is going to be given by rho 1 2.

And lastly, what we are not yet consider is the correlation matrix between v Z and Z 1, what is that going to be equal to? That is the covariance between v Z, which we are denoting by v Z times the Z 2. So, this is just going to be B Z times the covariance matrix of Z 2 and Z 1, that is equal to rho 2 1.

So, these are actually the corresponding forms of 1, 2, 3 and 4 in appropriates orders when we are looking at the canonical variables, which are computed from the rho matrix.

(Refer Slide Time: 19:34)



Now, an interesting thing is to see whether there is any change in this correlation matrices realize, that when we have denoted this A Z matrix by the following matrix, which is e 1 prime, e 2 prime, e p prime, that times sigma 1 1 to the power minus half times this v 1 1 to the power plus half. So, this is that A Z matrix.

And what we have earlier obtained was that rho between u vector and x vector was equal to A times v 1 1 to the power half times this. What we had derived earlier in this one expression here was, that this is the correlation between u and x, we had denoted that by A times sigma 1, actually derived it to be equal to be A times sigma 1 1 times v 1 1 to the power minus half and we had seen, that if we want to write it in terms of the correlation matrix, then this correlation matrix between u vector and the x vector is given by A v 1 1 to the power plus half times rho 1 1, that is what we are using here. So, this is A times v 1 1 to the power half times this rho 1 1 matrix.

Now, what we are going to do is we are going to write this A matrix. Now, what was A matrix? A matrix was e 1 prime, e 2 prime and e p prime, this times sigma 1 1 to the power minus half, this was what we had as the A matrix in the previous non-standardized variable setup. And what we have now is that multiplied by v 1 1 to the power plus half times this rho 1 1 matrix.

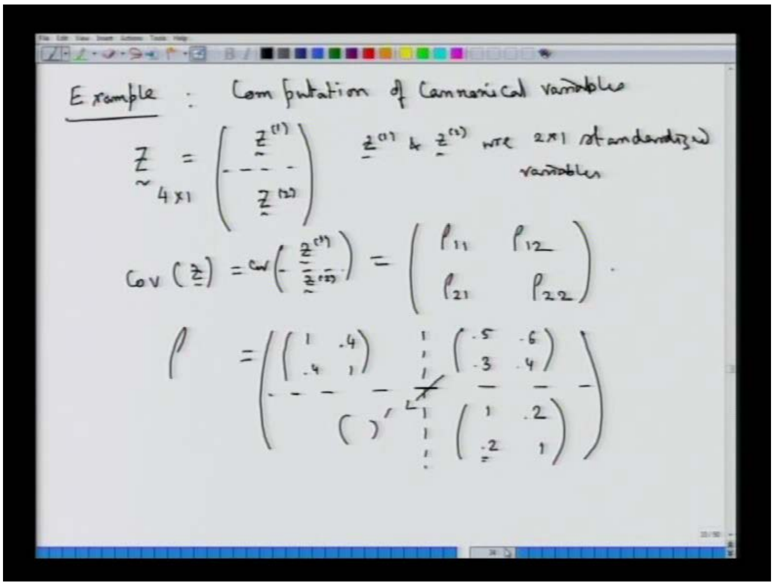
Now, what is this in terms of A Z matrix? This is e 1 1, e 1 prime, e 2 prime, e p prime multiplied by sigma 1 1. So, this basically is this matrix itself, starting from this up to this particular point, this is this A Z matrix itself. So, this is A Z times rho 1 1, which we have just obtained to be the correlation matrix between u z vector, which is the vector of canonical variables obtained from standardized variables and the standardized variables.

So, what is the observation? The observation is, that the correlation matrix is unchanged under this particular change in the starting point, that instead of looking of the covariance matrix, we are now looking at the correlation matrix and doing the calculations, the derivation for the canonical variables straightaway from the correlation matrix.

Now, similar is the situation for the other expressions. We can also prove, that the variance between the correlation matrix between v and y is what we have just now derived as B Z times rho 2 2, which would just be equal to what this is, v Z times this would be just equal to v times, sorry, this is, this is v Z and the Z 2 component, this is just going to be equal to the correlation matrix between v and the original set of variables, that is, y.

So, these observations imply that the correlations are not affected, are not affected by standardization of the variable. So, whether we work with non-standardized variables, that is, x 1, x 2, x p and y 1, y 2, y q, all we work with the corresponding standardized variables, which we are going to have the same canonical correlation coefficients. We are also going to have these correlation matrices to remain unchanged. What is going to change little bit is the vector, which is combining it because anyway, we are using the standardized variables. So, the canonical variables forms look a little bit change, although it is not that much.

(Refer Slide Time: 24:03)



So, what we are now going to see is to look at an example of the computational steps involved in this canonical correlation coefficient analysis; computation of canonical variables and the corresponding correlations, starting from correlation or covariance matrix.

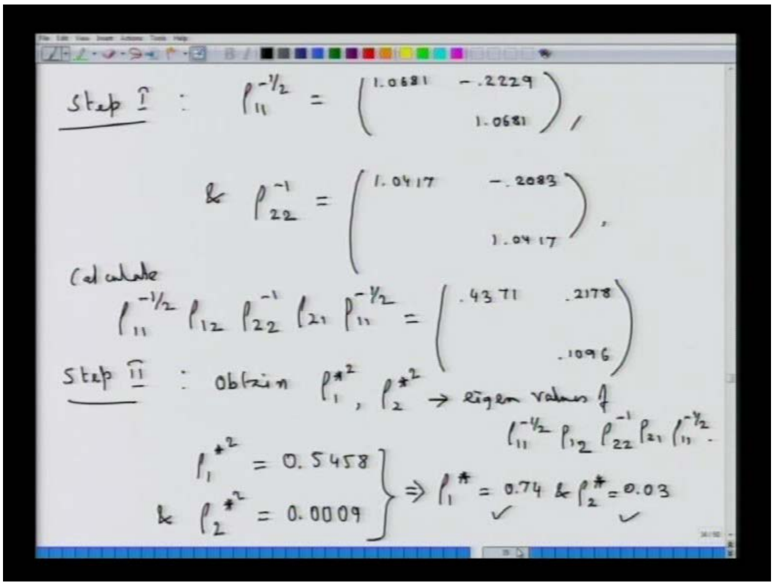
Now, in this example we consider a Z vector, which is a standardized variable, which is the vector of standardized variables, which is partitioned into two parts, Z 1 and Z 2. Now, this Z 1 and Z 2 are 2 by 1 standardized variables. So, we have this starting point and this covariance matrix of Z vector. The covariance matrix of Z vector is the covariance matrix of Z 1 component and Z 2 component in that partition form.

Now, this is going to be given by, since these are standardized variables in our earlier notation, this we write as rho 1 1, rho 2 2, rho 1 2 and the transpose as rho 2 1. Now, suppose these values are the following, we take the following numerical values of these four block matrices, 2 by 2 matrices, each of them. So, this suppose is given by 1, 1, 0.4, 0.4, so this one, the second element we take these numerical values as 0.5, 0.6, 0.3 and a 0.4 and this block here 1, 1, 0.2. Well, this is the symmetric matrix, however this one is not a symmetric matrix that is why, it is required to write all the elements here.

We can ignore to write this particular element here because that is what is element of that symmetric matrix and what we are going to have here is the transpose of this matrix here.

So, starting from this, this is the rho matrix, starting from this matrix we will look at what theory we learned and how to go stepwise in calculation of the canonical variables.

(Refer Slide Time: 26:36)



So, what we will first do is, that let me write that as step one, step one, that is the first thing, that we will be requiring is to look at rho 1 1 to the power minus half. Why are we going to be require that, it will be evident very quickly.

Now, the numerical value of this rho 1 1 to the power minus half turns out in this particular example as 1.0681 minus 0.2229 and this entry is 1.0681. This is this element only, it is the symmetric matrix and rho 2 2 inverse, that is also necessary. So, from the 2 by 2 rho 2 2 matrix we can easily find out what is rho 2 2 to the power minus 1. So, this turns out to be 1.0417 minus 0.2083 and this also is 1.0417. So, using these two rho 1 1 to the power minus half and rho 2 2 the power minus 1 and the forms of rho 1 2 and rho 2 1 we can calculate then this important matrix, which is rho 1 1 to the power half rho 1 2 rho 2 2 to the power minus half rho 2 1 rho 1 1 to the power minus half.

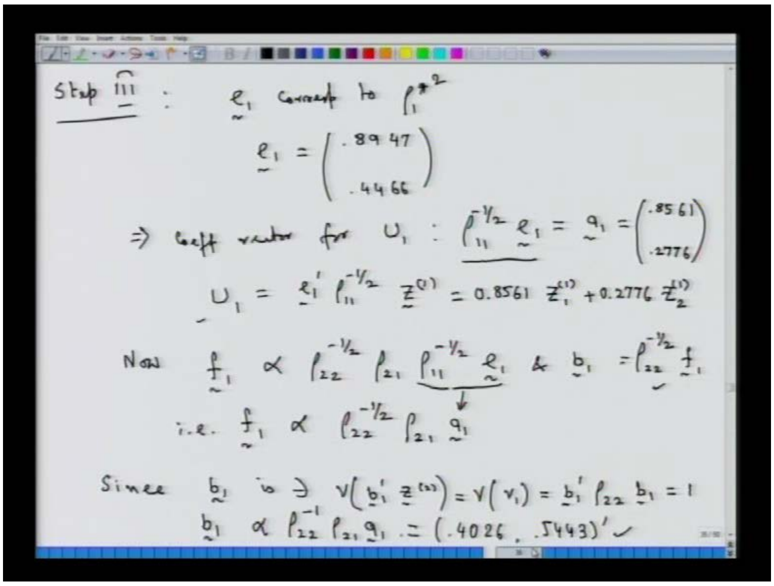
So, this is the trivial calculation with these two matrices and rho 1 1 and rho, rho 1 2 and rho 2 1, as given earlier, this turns out to be, this is 2 by 2 matrix, which turns out to be 0.4371, 0.2178, this is 0.1096, this also is a symmetric matrix. So, we have calculated up to this particular point.

Now, when we have this, then all the canonical correlation coefficients, the canonical variables, all the derived calculations, like correlation between original set of random variables and the canonical variables, all the things can actually be performed from this particular matrix.

So, first, what we, we are going to look at is the canonical correlation coefficients. So, step two is to look at this and we obtain rho 1 star square, rho 2 star square, these are what Eigen values of this matrix, which we have computed, which is rho 1 1 to the power minus half, rho 1 2, rho 2 2 to the power minus 1, rho 2 1, rho 1 1 to the power minus half. Now, it turns out, that this rho 1 star square, the Eigen value, the largest Eigen value of this matrix here is 0.5458 and this rho 2 star square is small, this is equal to 0.0009.

Now, these two are the two Eigen values of this crucial matrix. Now, from here what we are going to get is the canonical correlation coefficients. This rho 1 star is the first canonical correlation coefficient, which is square root of this particular term and this square root is approximately equal to 0.74 and this rho 2 star is a square root of this particular term, which is 0.03. So, the 1st canonical correlation coefficient is 0.74, the 2nd canonical correlation coefficient is 0.03. Now, these are the canonical correlation coefficients.

(Refer Slide Time: 30:34)



What are the canonical variables, which is important? Now, in order to calculate the canonical variables, we move on to this 3rd step of this computational term. What we need to find out is the orthonormalized Eigen vector of this matrix corresponding to the Eigen value, which is rho 1 star square. So, we need to calculate this e 1 corresponding to this rho 1 star square. It turns out, that the orthonormalized Eigen vector has this particular form, this numbers are 0.8947, 0.4466. So, this is the e 1 vector.

Now, this would imply, that the coefficient vector for u 1, the first canonical variable u 1 is going to be given by the following. What is that coefficient vector? That is rho 1 1 to the power minus half times e 1. Because we are working with the correlation matrix, we are considering this matrix as rho 1 1. If we had the starting point as the sigma matrix, the covariance matrix we would have taken here as sigma 1 1. So, whatever is the starting correlation or covariance matrix, we will stick with that.

Now, we will have this term coefficient vector to be given by this a 1 and this calculation of this one turns out, that this a 1 vector is going to be given by 0.8561. So, these are the numerical values, which are the elements of the first canonical variables coefficients. So, if this is a 1, then u 1, which is the first component of the first canonical variable derived from xs, this is going to be given by this e 1 prime rho 1 1 to the power minus half this times x vector. Now, x is not there, we have Z 1 vector because that is the standardized variables. So, what we are going to have? This first canonical variable as 0.8561 that multiplied by the first element of this Z 1 vector, this plus 0.2776 times the 2nd component of the 1st vector, first standardized vector. So, this is as far as u 1 is concerned.

Now, we move on to look at how to compute the v 2 vector. v 2, the canonical variable corresponding to the Z 2 variables, now in order to get to that we would require the f 1 vector. Now, we realize, that this f 1 vector, as we have seen in the theoretical lectures, that this f 1 is promotional to rho 2 2 to the power minus half times rho 2 1, that multiplied by rho 1 1 to the power minus half times e 1. So, we can explicitly compute this particular term here, but before computing that, we see, that this particular term is just equal to a 1.

So, and from here, of course what we will able to write is, that this b 1, that vector is going to be given by this rho 2 2 to the power minus half times this f 1 vector, whatever we are going to obtain.

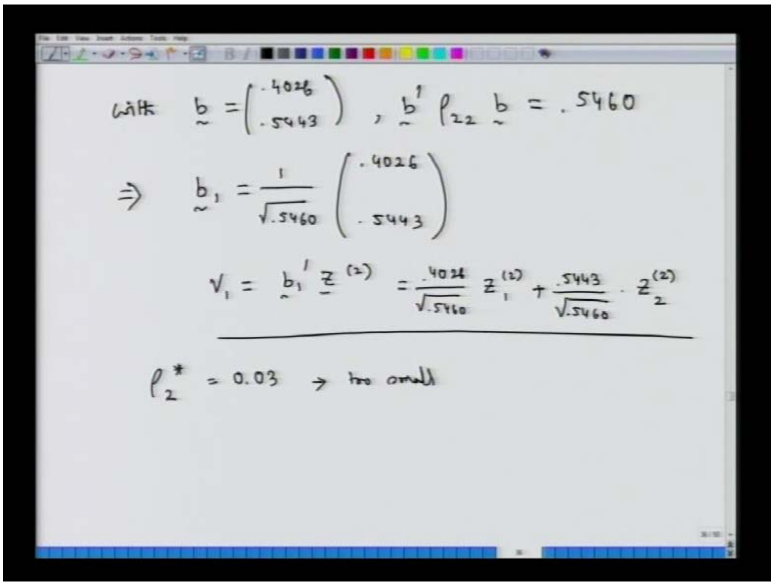
Now, this proportionality constant here is going to be obtained in the following way. Let us move on one more step, that is, from here we will write this f 1 as being proportional to rho 2 2 to the power minus half times rho 2 1 times this vector, what we have already obtained. And it turns out, that the numerical values of this components of this particular vector with a 1 given by this rho 2 1 is a known thing, rho 2 2 inverse is a known thing. So, what we are going to have is to write the numerical value, which is 0.4026 after computations, so this is, .433, .5443. So, this f 1 is going to be proportional to this.

Now, how do we obtain the proportionality constant? Now, note, that when we are going to obtain this b 1, now b 1 is the coefficient vector, which is going to be multiplied, b 1 prime times z 2 vector. Now, that is going to lead us to v 1. Now, v 1 is the canonical variable. So, the variance of v 1 is going to be unity and that is where we are going to obtain the proportionality constant in this particular vector out here.

Now, since we will have this b 1 vector defined as in here is such, that variance of b 1 prime times this z 2 vector, which is of course nothing, but variance of v 1, the canonical correlation variable, canonical variable. So, this is going to be given by b 1 prime, the covariance matrix of z 2, which is nothing, but rho 2 2, this times b 1 vector, that should be equal to 1. So, when we have obtained this term here and the corresponding b can be obtained so as to satisfy such a condition, that whatever the coefficient vector we are going to get by, because b 1 is equal to this and f 1 is promotional to this, f 1 is proportional to rho 2 2 to the power minus half rho 2 1 times a 1.

We see, that our b 1, what we have written there? b 1 is going to be proportional to rho 2 2 to the power minus half times this particular term here. So, this is going to be this, b 1 is proportional to rho 2 2 to the power minus 1 rho 2 1 times a 1, where we have this term as, I am sorry, this is not this particular term, this is just this term out here, f 1 is proportional to this, which is coming by looking at this term to be just equal to a 1. And since b 1 is equal to rho 2 2 to the power minus half times f 1 we will have this b 1 proportional to rho 2 2 inverse times this and this vector, actually the final vector turns out to be previous vector, that I had written earlier. So, this is 0.4026 and the 2nd entry is 0.5443 transpose. So, this b 1 turns out to be this.

(Refer Slide Time: 37:35)



Now, we will normalize this with b as equal to this previous vector 0.4026, 0.5433. We will have this b prime rho 2 2 b, this numerical value turns out, that this is equal to 0.5460. So, this would imply, we now get the proportionality constant and the b 1 vector as 1 upon root over of this particular norm, so that we will have 5460 under root here and then, this is 0.4026, 0.5443. So, this b 1 vector is the combining vector and this b 1 vector given by this and hence, v 1, the first component, there is b 1 transpose times the z 2 vector, that is, this is going to be equal to 0.4026, that divided by 0.5460, that multiplied by the first component of the second set Z 2, this plus 0.5443, that divided by under root of, 0.4, 0.5460, that multiplied by the 2nd component of the Z 2 component, Z 2 component here.

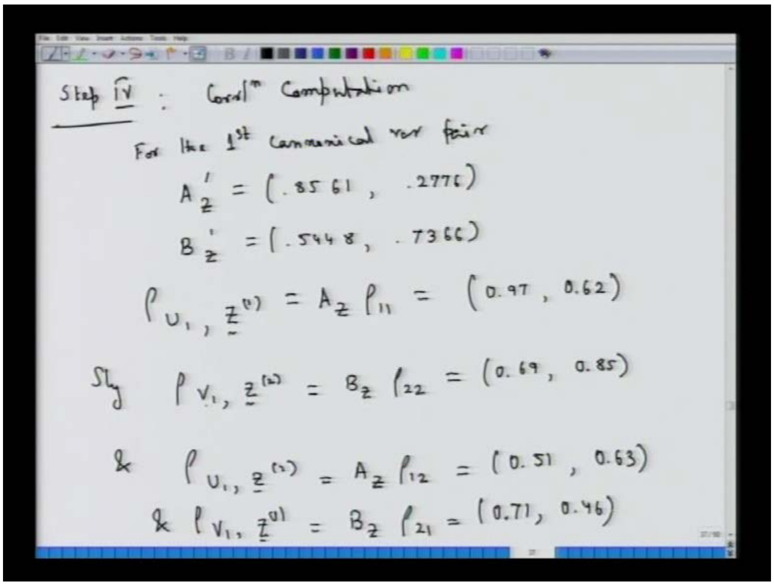
So, we have this as the v 1 and we have previously computed u 1 as this. So, u 1 and v 1 now comprise the first pair of the canonical variables with the canonical correlation coefficient as we have computed earlier. The canonical correlation coefficient corresponding to that first pair, that is, the correlation between u 1 and v 1 is going to be 0.74.

So, once we have obtained the pair of canonical variables u 1 and v 1 starting from this matrix what we have computed crucial matrix where is that this rho 1 1 to the power minus half rho 1 2 rho 2 2 the power inverse wherein we had this as largest Eigen value and this as the smallest Eigen value because there are two Eigen value only

Now, the first canonical variables, u 1 and v 1 were computed from the Eigen vectors straightaway, as we have seen from e 1, and all the computations, basically on that f 1 is also derived from e 1 in this way. So, if we further want to calculate the 2nd pair of canonical variables, the u 2 and v 2, what we will have to do is to get back to this rho 1 star and rho 2 star and then compute e 2, which is going to be the orthonormalized Eigen vector corresponding to this rho 2 star of this matrix here and then replacing e 1 by e 2, which would be the Eigen orthonormalized Eigen vector corresponding to, rho 1, rho 2 star square. We will do all these steps and we will get to u 2 and v 2.

However, in this particular situation we see that this rho 2 star is equal to 0.03, which is too small actually. So, if we are at all going to compute the 2nd pair of canonical variables, that is, u 2 and v 2, the correlation, that we are going to get between u 2 and v 2 is going to be as small as 0.03, which is negligible and hence, we are not going to compute u 2 and v 2. We will see later on actually, after this particular example that how these values of the canonical correlation coefficients decide, can decide as to the order of approximation, that one can obtain through these canonical variables.

(Refer Slide Time: 41:38)



Now, at the, as the last step we will look at this correlation calculations. So, this will comprise the last step of this particular computational example. So, what we will look at is the correlation calculation or computation correlation between correlation matrices, between the original and the canonical variables.

Now, what we have is for the first canonical variable pair actually, because we will look at in pairs, what we have is A Z, it is a vector. The transpose of that vector is the coefficient vector 0.8561 is the numerical value, the second value was this term and this b Z transpose, that is equal to 0.5448, 0.7366. So, these are the two combining vectors, which is playing role of A Z and b Z matrix, as in the theoretical part.

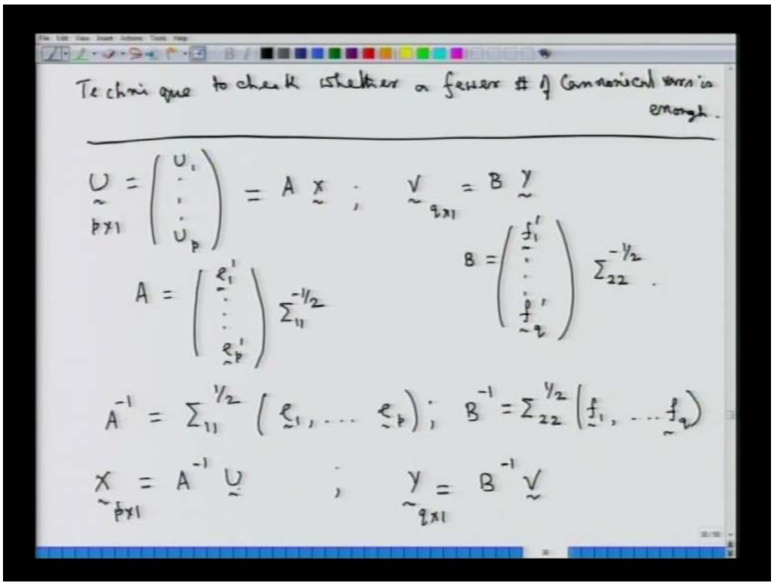
So, if we want to calculate the correlation between the first canonical variable u 1 and the first set of original variables, this is as we had seen in the theory, that this is given by A Z times rho 1 1. So, for the given example, this with rho 1 1 value and this A Z value, this is going to be A 1 by 2 vector, which is going to be comprising of these two elements, which is 0.97 and 0.62. What does this indicate? This is coming from the theory part, but what it indicates is important, that the first canonical variable u 1 is highly correlated with the first component of this z 1 vector, having a correlation of the order as close to 1 as possible, 0.97. And the 2nd component is of z 1, is also fairly highly correlated with the first component of the first canonical variable pair.

Similarly, one can obtain the other terms, like correlation between v 1 and the 2nd component of that Z vector from where, actually v 1 is derived. So, this, as per the theoretical notations, this was given by v Z times rho 2 2. This turns out to be, that the coefficients of this vector is 0.69 and 0.85.

So, if we look at the 2nd component of the 1st canonical variable pair, which is v 1, that is, more closely associated to the 2nd component of this Z 2 vector than to the 1st component with the two correlations reasonably high. And we can also look at the cross term, that is, u 1 with z 2, this is, was given by A Z times, this is A Z times rho 1 2, the numerical values of this is 0.51 and 0.63 and the other last one, which is v 1 and Z 1, this is going to be given by this b Z times rho 2 1, the numerical values of this is given by 0.71 and 0.46.

So, this concludes all the calculations that we have done in the theory part to compute the canonical correlation coefficients, to compute the canonical variables one or two, whatever is the desired number of pairs that one wants to calculate. And also, for interpretation of those canonical variables, one can look at this correlation coefficient calculations and all of them can be done starting from that crucial matrix rho 1 1 to the power minus half rho 1 2, rho 2 to the power minus 1 rho 2 1 rho 1 1 to the power minus half.

(Refer Slide Time: 45:56)



Now, we move on to the next concept in which we are going to look at the following technique. This is the technique to check whether a fewer, fewer than the original variable dimension, whether a fewer number of canonical variables is enough. So, this is what the concept is.

And we are going to now look at how to check what we are actually trying to convey by this particular or achieve by this particular technique is the following, that we have, say starting x vector and a y vector, its covariance structure given by the block sigma 1 1, sigma 2 2, sigma 1 2 and sigma 2 2 and the correlation matrix derived from that covariance matrix.

Now, we are trying to replace these by pairs of canonical variables u 1 v 1, u 2 v 2 up to the point where we can form the pairs that is the last pair. If we assume, that p is less than or equal to q, then the last pair will be u 1 paired with u p paired with v p. Now, the remaining, p, v p, that is, v p plus 1 to up to v q are not going to be paired, but what we are trying to say is that whether we can obtain a number k say, which is less than p and say, that k of these canonical variable pairs are enough to capture the covariance structure or the correlation structure between x and y, that is what we are going to achieve.

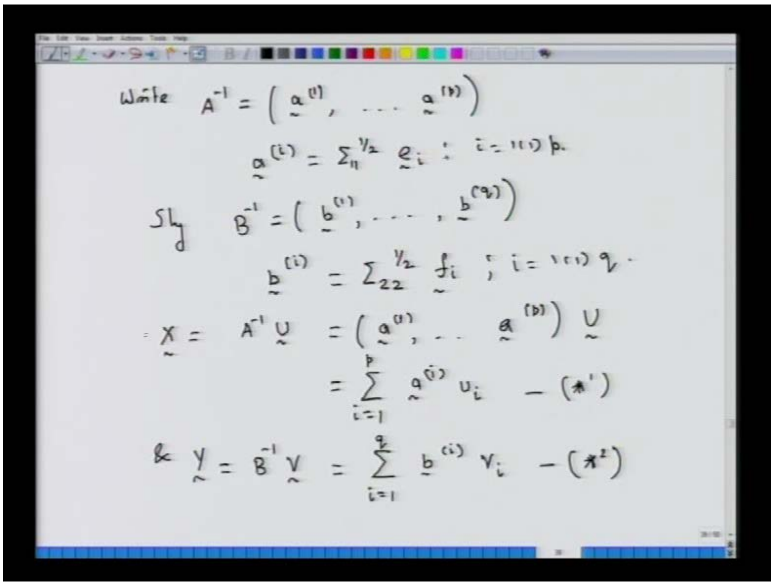
Now, how we are going to do that is through this particular technique. Now, when we had defined this canonical variables, we had seen, that this u 1, u 2, u p, we are not actually working with the standardized variables and hence, I am just writing this as A x to work with standardized variables are once again not difficult, it just follows from these calculations. So, that was given by this A times x, wherein we had obtained this A as e 1 prime, e 2 prime, e p prime, this times sigma 1 1 to the power minus half. And similarly, we had the other component, this is, remember p by 1 and this is q by 1. This we had denoted by matrix B times y, the original random vector, which was q dimensional and here, this B matrix, as we had seen, was given by this f 1 prime, f 2 prime, f q prime times sigma 2 2 to the power minus half.

So, once we have this u as A x, v as B y, where A and B are given by this. Now, note, that what is the inverse of this matrix? Now, the characteristics of this particular matrix, which has rows as e 1 prime, e 2 prime, e p prime is that it is an orthogonal matrix because e i vectors are orthonormalized Eigen vectors corresponding to the p largest Eigen values. And hence, if we look at the inverse of this matrix, of course, its inverse exists and that inverse is going to be given by sigma 1 1 to the power plus half and the inverse of this matrix, which is just be the transpose of this particular matrix because it is an orthogonal matrix. So, this is going to be given by e 1, e 2 up to e p where these are the columns of the inverse of this matrix.

And similarly, this B matrix also is nonsingular with the inverse of B matrix given by the form, which is sigma 2 2 to the power plus half, that multiplied by inverse, which is the transpose of this matrix and which is equal to f 1, f 2 up to f q. So, this is A inverse and this is B inverse.

Why we are looking at this? Because we want to write this x vector, the set of original random variables, the vector of original random variables. Now, this x, if A inverse exists is just equal to A inverse times u and similarly, this y vector is nothing, but B inverse times this v vector. So, this is the x original vector p dimensional and this is y, the original q dimensional in terms of the canonical variables. Now, these two expressions here of x and y are going to lead us to this important observation.

(Refer Slide Time: 50:59)



Now, if we write A inverse and B inverse in the following form, then we get something very nice, that suppose this A inverse, A inverse is this matrix, we write that in terms of the columns of the inverse matrix as A 1, A 2 this is p by p. So, the pth column, let us write that as A p.

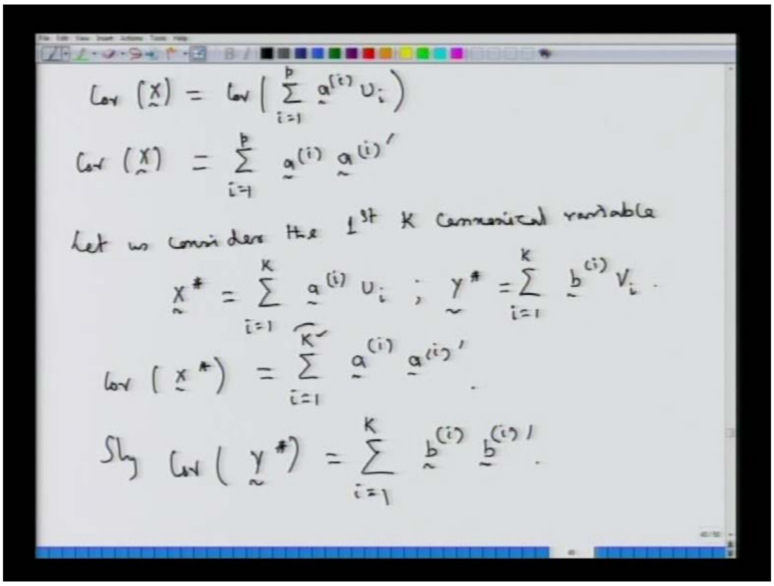
Now, what is here? A i, A i is the ith column and what would that be given by? That is given by this particular term sigma 1 1 to the power half, that multiplied by the e i vector here. So, a i is sigma 1 1 to the power half times the e i vector, this is the e i vector.

Now, similarly, this B inverse, we write as B 1 vector, the first column vector of the inverse matrix. And this last one is going to be that B q, which is the qth column of the B inverse matrix, wherein similar to the A i vector, the B i vector. Now, this i is from i equal to 1 to up to p and this B i is going to be given by sigma 2 2 to the power times this f i vector for every i equal to 1 to up to q.

So, we have this as A i and this as B i. Now, realize, that we are writing x as A inverse u here, so that once we are writing A inverse as this A 1, A 2 and A p vector, that multiplied by this u vector is just going to give us summation i equal to 1 to up to p A i vector, that multiplied by the ith component here of u.

So, we have this as A i times u i and y is B inverse times v vector, so that similar to this x term here what we can write is i equal to 1 to up to q. Now, b i vector, that multiplied by v I, now these two are importance forms, let me write as star 1 and star 2.

(Refer Slide Time: 53:37)



Now, if we look at this representation of x, what is the covariance matrix of x? This covariance matrix of x is given by this covariance matrix. Now, x we are writing as i equal to 1 to p of a i vector, that multiplied by u i.

Now, if we are writing it in this particular form, then covariance we are going to have this as summation i equal to 1 to up to p a i into this covariance between u i and the other term, which is of course, going to be equal to 1, because it is covariance between u i and u i in this summation here, and u i and u j, that is going to be equal to 0. So, what we are going to get as covariance of x here through this calculation is that a i a i transpose. Now, if this covariance of x is given by summation i equal to 1 to p a i a i transpose.

Let us under this condition consider, let us consider the first k canonical variables only. So, we are going to neglect from k plus 1 to up to q terms and consider only the first k canonical variables. So, if we consider the first k canonical variables, then we are neglecting rho k star squares from starting from k plus 1 to the n term and then, this x star vector is defined through that definition as i equal to 1 to k up to the terms we are retaining because if we look at this x here we are saying, that we are chopping off at the point k here. So, we are not looking at this a k plus 1 to up to a p vector and only retaining a 1, a 2 up to a k. And under such a situation this x star is an approximation to x and we are going to have that denoted by this summation i equal to 1 to k a i times u i.

And similar to this, since we are only considering first k canonical variables, what we are going to have as y star is the following, that this also is i equal to 1 to k b i times this v i. So, this basically is what we are having as x star and y star as when we are looking at the first canonical variables. Now, realize, that if we look at the covariance matrix between these two components here, well, that can be computed, let me first look at a simple term here in this lecture.

So, the last thing in this lecture, what we are going to see is that this, what is covariance between x star? What is the covariance of this x star vector? That by the similar logic is going to be equal to i equal to 1 to up to k a i a i transpose. So, this covariance matrix of x, which is summation i equal to 1 to p is now being approximated by the covariance matrix of A x star, which is the same summation up to the term this k. So, we basically are trying to say, that the terms from k plus 1 to up to p in this sum here are neglected when we are looking at this. Similarly, this covariance matrix of the y star vector is going to be given by summation i equal to 1 to up to k b i into b i transpose.

So, we will end the lecture, this lecture here. We will continue from this particular point in the next lecture.