Applied Multivariate Analysis

Prof. Amit Mitra

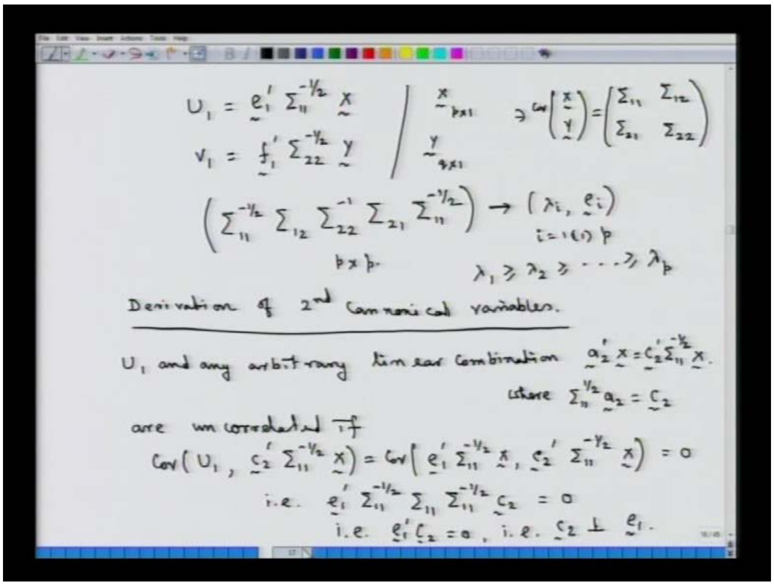
Prof. Sharmishtha Mitra

Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture No. # 41

Cannonical Correlation Analysis

(Refer Slide Time: 00:34)



In the last lecture, we had started our discussion on canonical correlation variables. We had given the basic definition, and also had looked at how to derive first pair of canonical variables. As we had seen that the first pair of canonical variables are given by the following, that this is given by U 1 equal to e 1 prime sigma 1 1 to the power minus half times X. So, X is the original setup random variables, X vector is p dimensional, and the second component of the first canonical variable was given by f 1 prime sigma 2 2 to the power minus half times this Y vector

Now, where in we had of course, define that X is that p by 1 random vector, and Y is the q by 1 random vector, this such that we will have the covariance matrix of X and Y augmented, that was given by sigma 1 1, sigma 1 2, sigma 2 1 and sigma 2 2. So, this was the covariance covariance structure of this X, Y random vector. And starting from this, we were trying to derive the first pair canonical variables, which we had denoted by

U 1 and V 1; we had derived that of this particular form, wherein of course we have that we consider this matrix, which was sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 inverse sigma 2 1 sigma 1 1 to the power minus half.

So, considering this matrix, which is p by p matrix, and then its Eigen value, Eigen vector pairs where denoted as lambda i e i, i equal to 1 to up to p. So, this e i of course, are ortho normalized Eigen vectors, corresponding to the Eigen values, which are lambda I (s). And we use those ortho normalized Eigen values, Eigen vectors e i, in framing the first canonical variable U 1, which is U 1 prime. So, e 1 is the Eigen ortho normalized Eigen vector of this matrix, corresponding to the largest Eigen value; which was lambda 1. So, we had this particular setup; that lambda 1 is greater than or equal to lambda 2 is greater than or equal to lambda p.

We at also seen that this f 1 prime or rather f 1 is the Eigen vector, corresponding to the largest Eigen value of the matrix which was sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 inverse sigma 1 2 sigma 2 2 to the power minus half. So, and the Eigen values of both these matrix a a transpose and a transpose a of course, are same the non-zero Eigen values are same. Only Eigen values with 0 multiplicities actually can differ, the multiplicity of 0 Eigen values in the 2 matrices can only differed. So, what we first going to take up today, is how we can derive the second third and in general k th canonical variable.

So, let us first look at that. So, we are trying to look at the derivation of second canonical variables. Now, in the definition of canonical variables 1 we recall that; we had said that, when we are moving on from the first do the second and in general to the k th cannot pair of canonical variables. The second pair of canonical variables will be such that, the U 2 which will be the first component of the second pair, that will be uncorrelated with the first corresponding first component of the first pair of canonical variables.

The second component of the second pair is V 2, if we have denoted by V 2. So, that V 2 is going to be uncorrelated with second component of the first pair; that is V 1. So, we would require these two follow, actually that the third pair will have will be uncorrelated with previous 2 pairs; that is the first in the second pair. And in general for the k th pair, we will have the components in the k th pair to be uncorrelated with the previous k minus 1, principal the canonical variables.

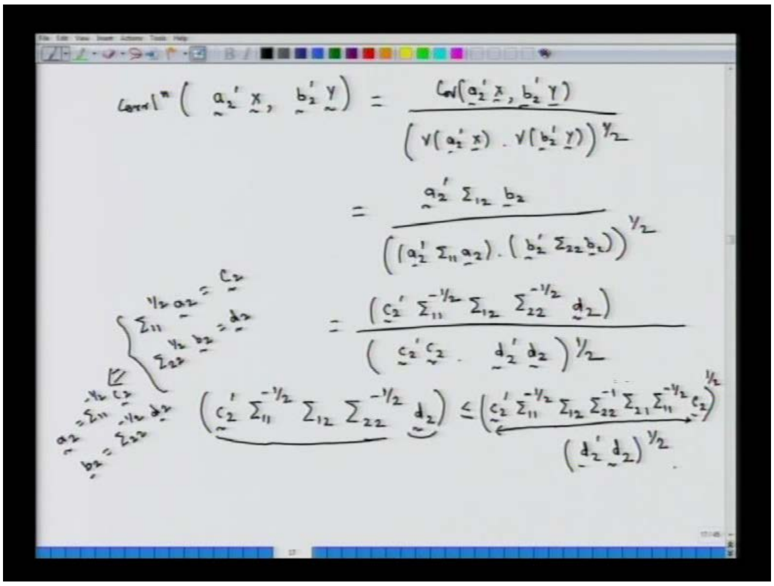
So, we will require that, and we will does see that U 1, and any arbitrary linear combination of X is say given by a 2 prime X. Now you may recall that while deriving the first pair of canonical variables. What we had done was to replace this a 2 vector, by something which was a c vector. So, in terms of that c vector, we can write this as a c 2 prime sigma 1 1 2 the power minus half times this X. Where, we had sigma 1 1 to the power plus half times a 2 that is equal to c 2. So, the c 2 vector was newly defining vector, similar to the c 1 vector; which we have defined, while deriving the first pair of canonical variables, in order to make the denominator to be equal to c prime c or d prime d as we will see.

So, with this definition, this arbitrary linear combination a 2 prime X, which is c 2 prime sigma 1 1 to the power minus half X are uncorrelated or going to be uncorrelated. If the covariance between U 1, and this c 2 prime sigma 1 1 to the power minus half X; which of courses, is equal to the covariance between now, U 1 is given by e 1 prime. So, this is this e 1 prime sigma 1 1 to the power minus half X that, and c 2 prime this is c 2 prime sigma 1 1 to the power minus half X is equal to 0.

So, if we have the covariance between U 1, and this arbitrary linear combination to be equal to 0. We would require the covariance between these 2 terms to be equal to 0. That is; if we look at, what is the covariance between this random variable, and this random variable. It is going to be e 1 prime sigma 1 1 to the power minus half times the covariance matrix of X which is sigma 1 1 into sigma 1 1 to the power minus half times this c 2 equal to 0. That is these terms here, sigma 1 1 to the power minus half sigma 1 1 to the power of sigma 1 1 sigma 1 1 to the power minus half will make this as an identity matrix. So, we would require this e 1 prime c 2 this to be equal to 0; or in other words, we would require this c 2 to be an orthogonal vector, orthogonal to the e 1 vector which is the ortho normalized Eigen vector, corresponding to the largest Eigen value, lambda 1 of this sigma 1 1 the power minus half this matrix.

So, while deriving the second canonical variable, we have to make this c 2. We have to choose this c 2 vectors, in such a way that this c 2 needs to be orthogonal 2 e 1. The first ortho normalized Eigen vector corresponding to the largest Eigen value

(Refer Slide Time: 08:37)



Now, let us look at what is the correlation coefficient between this. Let me take this as first to start with a 2 prime X that is the original linear combination forms that, and b 2 prime Y. So, we are trying to look at, what is that a 2 and b 2? Such that, we will have this correlation to be maximized subject of course to the condition, that these a 2 and b 2 should be such that this a 2 prime X will be uncorrelated with U 1, and b 2 prime y will be uncorrelated with V 1.

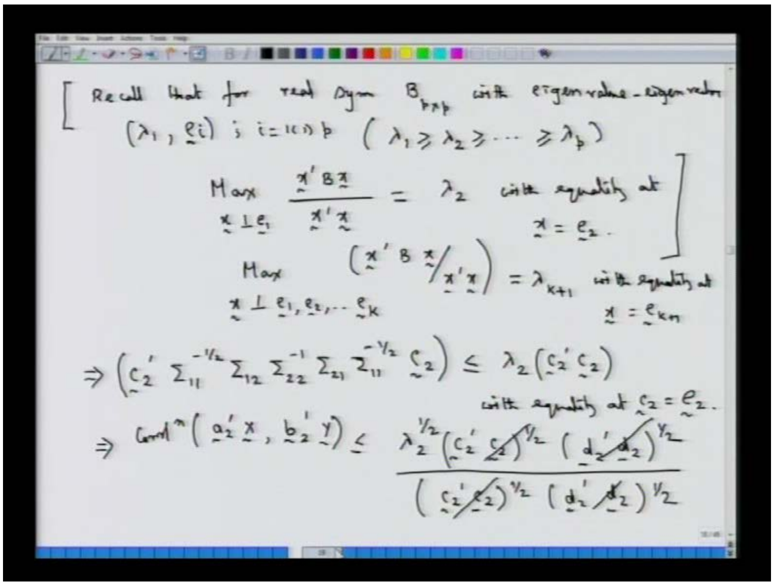
So, this correlation is nothing, but the covariance between a 2 prime X, and b 2 prime Y this divided by under root of variance of a 2 prime X, and variance of b 2 prime Y. So, this expression, we can write as simply a 2 prime sigma 1 2 times this b 2 prime, this divided by a 2 prime sigma 1 1 a 2. This is the first term, and this second term is similar to the first term, b 2 prime sigma 2 2 times and b 2 and whole raise to the power half.

So, similar to the derivation of the first pair of canonical variables, what we are going to do is redefined. So, we will take this sigma 1 1 to the power half a 2 vector, we will define that as a new vector c 2 as we have seen in here, this is the transformation. So, under this transformation and also with sigma 2 to the power half times b 2 that 2 be equal to say it b 2 vector. What we will be having in the denominator, is that this e 2 prime sigma 1 1 e 2 is just equal to c 2 prime c 2, that multiplied by this d 2 prime d 2 whole raise to the power half. And here, if we have this particular definition this would further imply that; our a 2 vector is equal to sigma 1 1 to the power minus half times c 2, and this b 2 vector is equal to sigma 2 2 to the power minus half of the redefined vector; which is d 2. So, we give a we can unplug in this values here in the numerator of this expression, which will read us to this c 2 prime sigma 1 1 to the power minus half that times sigma 1 2 V 2 prime. I am sorry, this is going to be just b 2, this is not b 2 prime; this is e 2 prime sigma 1 2 b 2

So, this b 2 is going to be given by this sigma 2 2 to the power minus half times this d 2 vector. So, the correlation between these two components is given by this, now similar to once again the derivation of the first pair of canonical variables, what we can do is to apply the Cauchy Schwarz inequality on the numerator of this expression.

And we can say that this c 2 prime sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus half times this d 2 vector, that is less than or equal to we will take this as 1 vector, say U prime as this vector, and V as this vector. So, it will be prime U 2 the power half. So, that this is going to be c 1 c 2 prime sigma 1 1 to the power minus half sigma 1 2 sigma 2 to the power minus half, and that also augmented with sigma 2 2 to the power minus half, will give us sigma 2 to the power minus 1, and the transpose of sigma 1 2 which is sigma 2 1. Sigma 1 1 to the power minus half times c 2 this raise to the power half, and that multiplied by this d 2 prime d 2, that also raise to the power half. Now, we will concentrate on this particular expression here, and try to see what can be given as an upper bound of this particular expression.

(Refer Slide Time: 13:28)



So, in order to get that, we will recall once again result from matrix theory, recall that for real symmetric matrix b say p by p with Eigen value, Eigen vectors pairs with Eigen value Eigen vector pairs as lambda i e i, for i equal to 1 to up to p this is. Such that, of course, lambda 1 is the largest Eigen value, lambda 1 is greater than or equal to lambda 2 is greater than or equal to lambda p with such a structure, if we try to look at what is the maximum over x of the expression x prime b x this divided by X prime X. Now this X here, if we try to look at all x, such that x is orthogonal to e 1.

The first ortho normalized Eigen vector corresponding to the largest Eigen value, which is lambda 1; then this maximum maximum over all x. Such that, they are all orthogonal to this e 1 is going to be a achieved at the second largest Eigen value, which is lambda 2 with equality at X equal to c 2.

Now, we cannot use the first version of this particular result, while deriving the first pair of canonical variables. Recall that, if we do not have this particular condition that we are looking at all x, such that, it is orthogonal to e 1. If we just look at maximum of this x prime V x by x prime x for all x, then this is going to be achieved at the largest Eigen value; which is lambda 1.

And the equality would be attained at x equal to e 1. Now, if we restrict the set of other region of x do the situation that all x, which are orthogonal to e 1, then what we are going to get is, this Maximum is going to be attained at the second largest Eigen values. Now in general, what also one can say is that maximum x, such that x is orthogonal to e 1, e 2 and e k. So, e 1, e 2, e k are the k ortho normalized Eigen vectors, corresponding to the k largest Eigen values of this b matrix. If we now look at all x such that they are orthogonal to the first k Eigen vectors of the quantity. Which is same as the previous quantity x prime a x divided by x prime x.

Now, this Maximum under this condition is going to be attained at the point, which is lambda k plus 1. So, it is nice result with equality here, with x equal to e k plus 1. So, this is the standard result any way in matrix series. So, we will use this particular result, in order to derive the second, third and forth, and in general k th pair of canonical variables. Now at what point we are going to use this particular result of Eigen values, we are going to use that result in this particular expression.

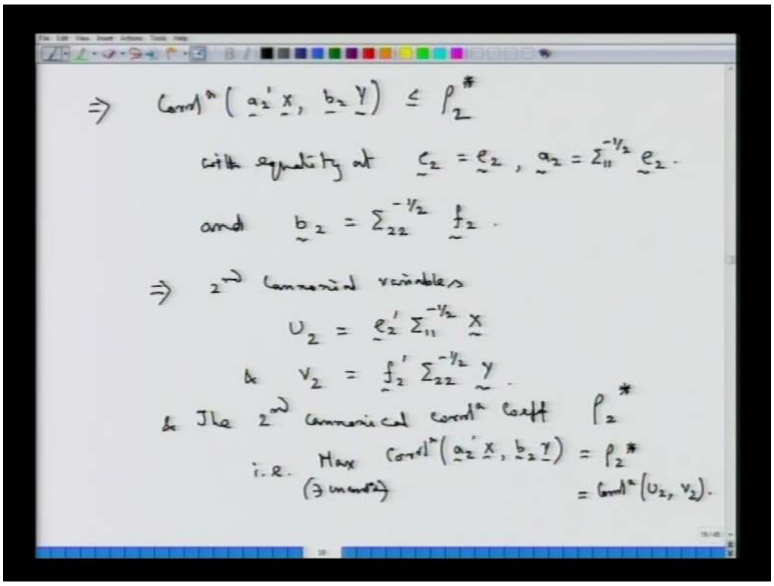
Now, let us write this expression. So, this will imply that over c 2 prime sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 inverse sigma 2 1 sigma 1 1 to the power minus half this c 2, this quantity here will be less than or equal to what if we are looking at c 2, we are looking at all c is such that, that is orthogonal to the first. That is it is a uncorrelated with the first canonical variable U 1.

And we had seen in here, in the first discussion that in order to have this c 2; c 2 should be such that, it should be uncorrelated with U 1. And the condition is that, we have to look at all such c 2s which are orthogonal to e 1. So, we will we can apply this particular result here, because we are looking at all c 2s, such that c 2 is orthogonal to e 1; e 1 is the eigen vector corresponding to the matrix defined earlier. So, this is going to be less than or equal to this lambda 2 times c 2 prime, c 2 right

In here, the equality will be attained, if we take c 2 to be equal to e 2; With equality here, with equality at c 2 equal to the e 2 vector which is the second, a which is ortho normalized eigen vector corresponding to the second largest Eigen value. Now, if we have this particular expression, then we can go back to this expression out here. And say that, this is going to be less than or equal to that particular term. So, this would imply, we can actually come to this particular point here. And say, what is the upper bound of this expression is straight away; this would imply that, the correlation coefficient between e 2 prime x and b 2 prime y. This is going to be less than or equal to lambda 2 to the power half c 2 prime c 2; this to the power half d 2 prime d 2 to the power half; this half coming from the Cauchy Schwarz inequality of the numerator and in the denominator, we as such have this as c 2 prime c 2 to the power half and d 2 prime d 2; this to the power half.

So, as we see that the 2 terms cancel out, and we see that the correlation coefficient between this is bounded by this lambda 2 to the power half.

(Refer Slide Time: 20:15)



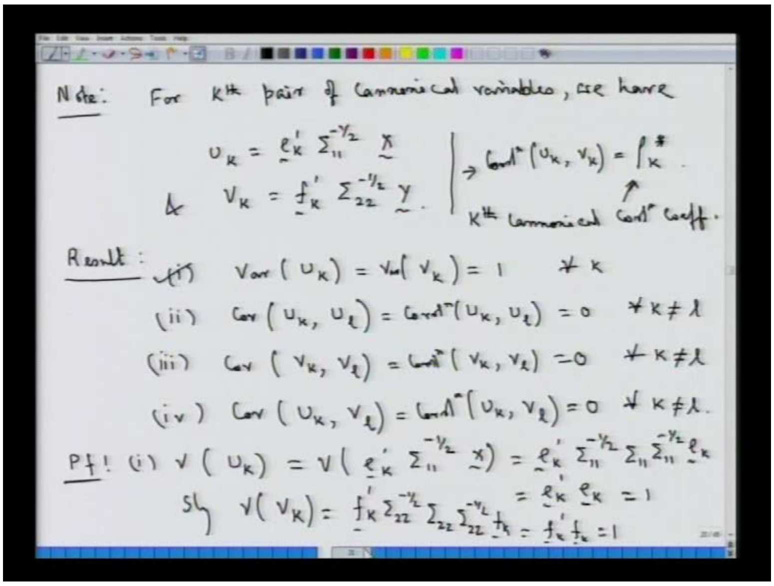
Now, we had denoted this lambda i as rho i star square. So, we will have the correlation; the Maximum correlation between this a 2 prime X and b 2 prime Y, such that the uncorrelated with the first canonical variable wholes. This is going to be less than or equal to this lambda 2 star, where lambda 2 star square is the Eigen value of that a; a transpose matrix.

And, with equality at what points a what are the points, where in we have applied inequality? So, we in order to have this equality to whole, we would require this c 2 to be equal to e 2. And in the previous expression, wherein we have got this as the inequality, we would require this d 2 vector to be proportional to sigma 2 to the power minus half sigma 2 1; sigma 1 1 to the power minus half times c 2. So, in the all the previous steps with equality at c 2 vector to be equal to e 2 vector

Now, c 2 vector is in terms of the a 2 vector a; the relationship between c 2 and a 2 is this. Thus we would require this a 2 to be sigma 1 1 to the power minus half times e 2; that is we would require this a 2 to be sigma 1 1 to the power minus of half times e 2. So, this is as for as the choice of e 2 is concerned, and we would also required the choice of b 2. Because, that is going to a play a role in this Cauchy Schwarz inequality; here the equality is going to whole if we have as a said d 2 proportional to this particular vector which would lead us to the following. That this b 2 vector, would just be equal to sigma 2 to the power minus half times f 2 vector. So, this will imply the second canonical variable pairs are going to be given by this e 2, which is going to be equal to e 2 prime sigma 1 1 to the power minus half times X vector, and V 2 vector is going to be equal to f 2 prime sigma 2 2 the power minus half times this Y vector. If we chose e 1 U 1 and U 2 to be of this particular form then not only will this U 1 U 2 and V 2, will be uncorrelated with U corresponding U 1 and V 1, We will have the maximum correlation of course, being attained

And the second canonical correlation coefficient second canonical correlation coefficient is going to be given by rho 2 star, that is the maximum correlation between this e 2 prime X, and b 2 prime Y this maximum of courses. Such that, this uncorrelatedness holds uncorrelatedness holds this is going to be attained at this rho 2 star. One can actually find out, what is a correlation between these 2 variables; this is going to be just the correlation coefficient between the e 2 and V 2; that we have derived.

(Refer Slide Time: 24:15)



So, we are been able to thus derive the second pair of canonical variables, we will a thus generalized this approach for k th pair of canonical variables K th pair of canonical variables, we have this U k to be given by e k prime sigma 1 1 to the power minus half times X, and our we k is going to be 1; that is going to be based on a f k prime time's sigma 2 to the power minus half times this vector y, with the correlation the k th canonical correlation will be the correlation between U k and V k. This is going to be equal to rho k star.

So, this is what is the canonical k th canonical correlation co-efficiency; this will be the k th canonical correlation coefficient, and the construction will be a similar to the construction of the second pair of canonical variables; that is this U k will be uncorrelated with U 1 U 2 U k minus 1, and this V k is going to be uncorrelated with V 1 V 2 up to V k minus 1.

So, entire that uncorrelated a structure will still hold, and while deriving this k th pair of canonical variables, we will precisely be using this particular result. That we would at the case k th stage would require the c k vector; similar to the a c 2 vector. Here you would require thus c k vector corresponding to the k th canonical variable pair to be orthogonal to e 1, e 2, e k. Because we would require that a particularly combination are the k th canonical variable to be uncorrelated with all the previous canonical variables. And hence that particular type of orthogonal structure would be required while deriving the k th pair of canonical variables.

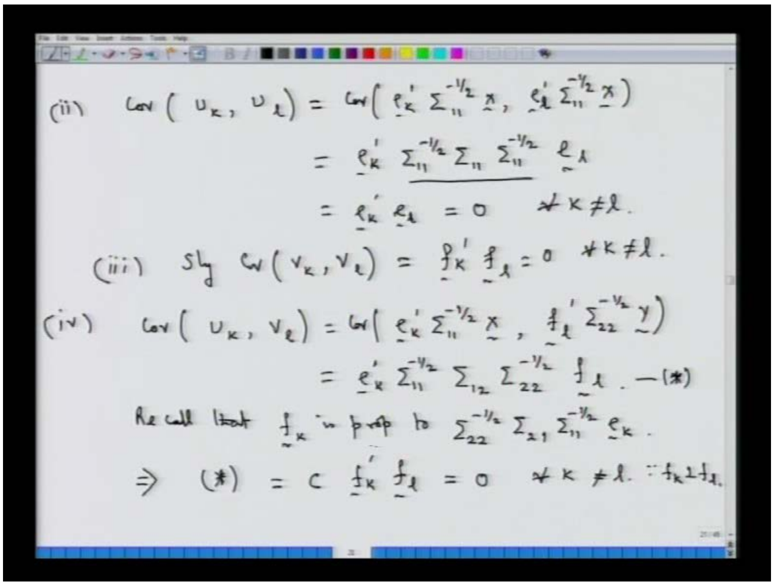
So, this is what concludes actually the derivation of all the canonical variable pairs. Now, how many such canonical variables are there? If we look at U 1, U 2, U k, that particular series of canonical variables, then we are going to have p of them. And if we look at this V k, V 1, V 2, V k, so that is going to go up to V q, because we have f (q) of the Eigen vectors present in that A transpose A matrix which is sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus 1; sigma 1 1 to the power minus 1; sigma 1 2 sigma 2 to the power minus half. So, these f 1, f 2, f case are associated Eigen Ortho normalized, Eigen vectors of that matrix; whereas, as these e is are the ortho normalized Eigen vectors of the other matrix, which is p by p. And f ks are the once which is corresponding to the q by q, Eigen q by q matrix, right.

Now, let us look at the following result which is going to summarize the variance, covariance structure among these canonical variables. The results are I will just list four of those results which are enough actually to capture the entire structure of the variance covariance fracture of this canonical variables. So, we will have as we had seen, it is prevail actually; that variance of U k is equal to variance of… I will just drop this may be, this right variance of k equal to variance of any other V k; this is going to be equal to 1; this they are 1 will have the covariance between U k and U l. This since we have this is to for all k since we have variances to be 1, the covariance between U k and U l is also equal to the correlation between U k and U l.

And they are equal to 0 for every k which is going to be not equal to l number 3 is that the covariance between U k and v l, that is equal to the correlation coefficient between V k and V l; that also be into be equal to 0 for every k which is not equal to l, and also the 4 th one which is covariance between U k and v l; that is going to be equal to the correlation coefficient between U k and v l. This is also equal to 0 for every k which is not equal to l, right.

So, if we are to prove this results, these are simple element re Results actually. Now, in order to prove that the first term that variance of U k, now U k, as we have derive it is given by variance of this e k prime sigma 1 1 to the power minus half times x. So, that is going to be equal to e k prime sigma 1 1 to the power minus half covariance matrix of X which is sigma 1 1 times; thus this sigma 1 1 to the power minus half times e k. So, this will give us an identity matrix. So, this is e k prime, e k which is equal to 1. Similarly, exactly in the same way, we can prove that variance of this V k for every k; that is also going to be given by… Now, what is going to happen? When we are looking at variance of V k, it is going to variance of this particular term; we will have this as f; we are going to have this as f k prime sigma 2 to the power of minus half times the covariance matrix of y which is sigma 2 to times sigma 2 2 to the power minus half times f k. And that is once again equal to f k prime time's f k that is equal to 1. So we will have variance of U k and V k for every k to be equal to 1.

(Refer Slide Time: 30:48)



Let us look at y the covariance between U k and V k, well in the way that we have constructed it, they are having the correlation equal to correlation or the covariance is to be 0.

Now, we can specifically see how that term is equal to 0. This is covariance between e k prime sigma 1 1 to the power minus half times X. And this U l is e l prime sigma 1 1 to the power minus half times X. So, that this term is equal to e k prime, and once again this same matrix is going to come here which is going to lead us to an identity matrix. So, this is sigma 1 1; this term here which is multiplied by this e l vector. So, this will be an identity matrix, and what we are left with is simply e k prime e l. Now, since we have e k orthogonal to e k e l will have this to be equal to 0 for every k, which is not equal to l; thus the covariance between U k and V l U l to be will be equal to 0.

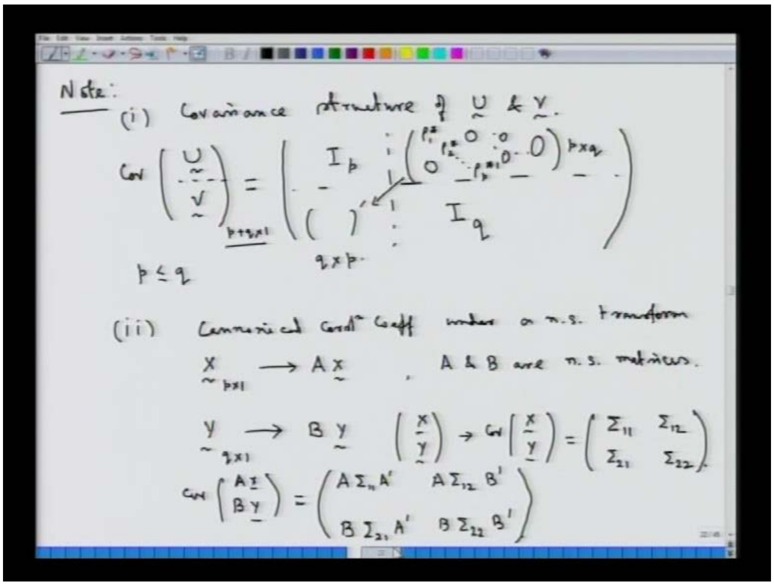
Similarly, the covariance which is equal to the correlation also; this is going to be given by f k prime f l. And that by orthogonality half f k and f l will be equal to 0 for every k not equal to l. So, we had proved the second part, the first part and the third part. In the fourth part, what we are this is the third part; in the fourth part, what we are trying to show is that covariance between U k and v l. What is this equal to this is equal to covariance between e k prime sigma 1 1 to the power minus half times; this X and V l is going to be given by f l transpose sigma 2 to the power minus half times Y. So, the covariance between these 2 would be given by this p k prime sigma 1 1 to the power minus half time's covariance matrix of between X and Y

So, that is going to be equal to sigma 1 2, and this is followed by the transpose of the coefficient vector. So, that is sigma 2 to the power of minus half times f l, right. Now, we had a result which are stated that, f l is proportional to… Recall let me put this equation as s star; recall that f k or f in general f l; so, this f k is proportional to this sigma 2 2 to the power minus half sigma 2 1 times sigma 1 1 to the power minus half times e k. right This as what we had seen; while deriving the first a pair of canonical variable. So, this f k is proportional to this term.

So, what we are going to have this 1; this basically is going to be a constant times f k vector, because if f k is proportional to sigma 2 2 to minus of sigma 2 1 1 1 to the power minus half times e k, we will have the transpose of this to be equal to c times f k prime. So, this will imply that this star expression is going to be equal to a constant C times f k prime f l. Now, this will imply that this product is equal to 0 for every k which is not equal to l this is. So, because this is once again, this f k is orthogonal to f l.

Now, since we have this, we will have the covariance between U k and V l which is also equal to the correlation between U k and v l, because these are unit variances U k and v l. And thus, we will have thus covariance going to be equal to 0 for every k which is not equal to l. For k equal to l, what is going to happen is that this will be equal to 1. And then this will be equal to the constant of proportionality which links this f k and e k.

(Refer Slide Time: 35:13)



So, we have proved this result also. So, to conclude we will look at this structure here the first noticed that we will look at the covariance structure covariance structure of U and g vector. So, we will first look at this covariance between this v vector, and the v vector. So, this is going to be given by the blocks which I write in here. So, the first block here is the covariance matrix of U vector what is that by the previous Result? This is just and I p matrix now the covariance matrix of v vector is nothing, but this is an I q matrix here and the half diagonal block here. This is what? This is the covariance between U and v right the covariance between U and v.

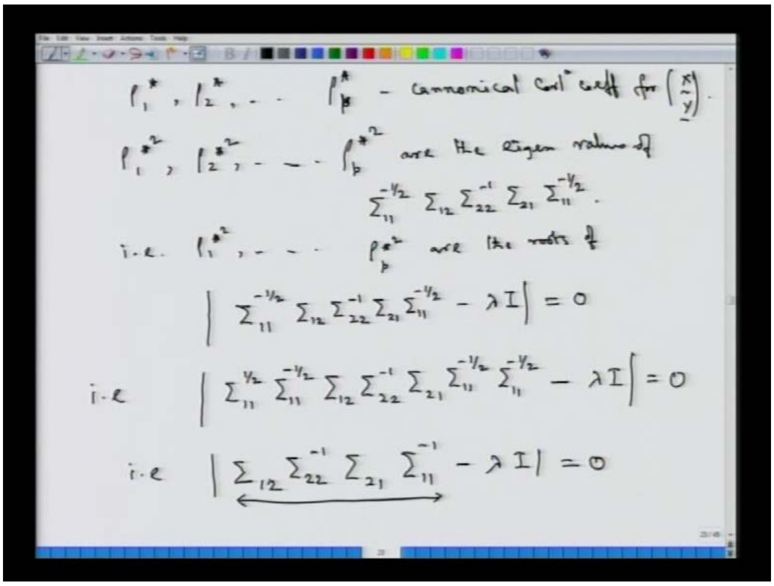
So, covariance between U 1 and V 1.What is that? That is the correlation between U 1 V 1. What is that in term first canonical correlation. So, that is rho 1 star rho 2 star, and this is rho p star after the point where we can remember what we had taken was p less than or equal to q where p is the order of X or the order of U naturally and q is the order of y which corresponds to the order also of this v. So, we will have this p of them which are the p canonical correlation co-efficiency. And what about the half diagonal entries these are all 0 (s), and there are of course, more 0 here after this p by p block here, because we can have the correlation between U 1 and v p plus 1 up to U 1 and v q.

So, we will have all this to be equal to 0 augmented, and here we will just the transpose of this particular matrix. So, this is actually this is going to be the p by q matrix here this is going to be the q by p matrix here and this vector is p by q dimensional random vector. So, its covariance matrix variance covariance matrix structure is going to be given by this p plus q by p plus q dimensional matrix as in here or in the second node. What we are going to see is that, this is canonical correlation coefficients canonical correlation coefficients under a nonsingular transformation.

So, what we are trying convey is the following that we have X as the original setup of p dimensional random vector. And y which is q dimensional random vector v make a transformation from X 2 a A x and from y 2 say B y wherein v are talking about a and b are non-singular matrices now the point of interest Note is to look at what is the change in the canonical correlation coefficient, if any in the transformed setup random vectors. So, instead of considering X and y as original setup random variable vectors, if we look at A x and B y, then what is going to be the corresponding change, if any in the canonical correlation coefficient. So, these are the 2 not singular matrices

Now, if we have this X y, then this has got the covariance structure X. And y as we had define earlier this is sigma 1 1, sigma 1 2, sigma 2 1 and sigma 2 2. Now, let us see what is the covariance structure of this; A x and B y it is easy to see that the covariance structure is going to be A sigma 1 1 A transpose; this is B sigma 2 2 times B transpose and this is A sigma 1 2 times B transpose. And this is going to be just the transpose of this of this is B sigma 2 1 times A transpose right. So, this variance covariance matrix which the this original variance covariance matrix between X and y, and this is the covariance matrix on the changed random variables

(Refer Slide Time: 40:01)



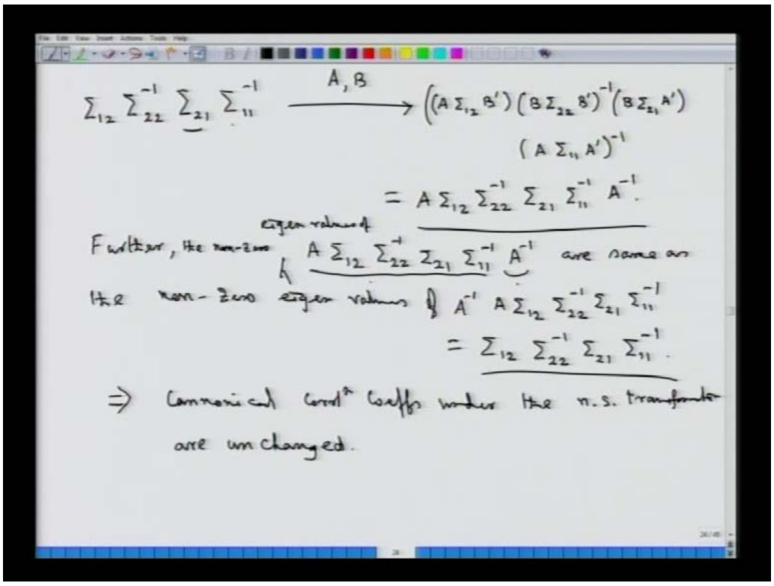
Now, we have this rho 1 star rho 2 star rho p star these are the canonical correlation coefficients canonical correlation coefficients for this X y setup; right That is what we have Derived. Now, we are going to see, what are the canonical correlation coefficients? If we are going to look at A x here and B y now this rho 1 star square rho 2 star square rho p star square are the Eigen values of the matrix that we had defined earlier which was sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 the power minus half sigma 2 1 sigma 1 1 to the power minus half that is this rho star square rho 2 star square rho p this rho p star square rho p star square are the rules of this equation which is sigma 1 1 determinant of sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 inverse sigma 2 1 sigma 1 1 to the power minus half minus lambda i this determinant of this matrix to be equal to 0 just the Eigen value equation

Now, here what we will do is to pre and post multiply with suitable matrices, such that let me right those matrices. This is sigma 1 1 to the power half; say here, pre multiplication sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 inverse sigma 2 1 sigma 1 1 to the power minus half, and this gets post multiplied with sigma 1 1 to the power minus half

So, if we do that, we are pre multiplied by determinant of sigma 1 1 to the power half and post multiplying; this equation with sigma 1 1 determinant of sigma 1 1 to the power minus half. And hence, there is no change in this term here; this still is lambda i, because the multiplication here is sigma 1 1 to the power half into sigma 1 1 to the power minus half which is an a still an identity matrix now if we have done that then this rho 1 star squawroot to rho p star square other roots also of the following equation which is this is an identity matrix sigma 1 2 sigma 2 2 to the power minus 1 sigma 2 1 and this is sigma 1 1 to the power minus 1 minus lambda i equal to 0 this equation right.

Now, what we will do is that we will look at this particular matrix here and see under the nonsingular transformation with a and v nonsingular matrices what is happening to this particular matrix.

(Refer Slide Time: 43:27)



So, let us write this matrix here this is sigma 1 2 sigma 2 2 to the power minus 1 sigma 2 1 times sigma 1 1 to the power minus 1 now this under the nonsingular transformation set A and B where is this getting transformed to now sigma 1 2 if we look at this variance covariance structure this is the variance covariance 1 matrix structure

So, what we can say is that sigma 1 1 is transformed to this sigma 1 1 is getting to a sigma 1 to be prime and so on. So, what we can right in place of sigma 1 2 is that under A B transformation this is going to A sigma 1 2 B transpose. So, this is corresponding to sigma 1 2 corresponding to sigma 2 2 what we have is B sigma 2 2 times b transpose whole inverse then sigma 2 1 is getting transformed to the transpose of this which is b sigma 2 1 times A transpose and sigma 1 1 in verse is what we have as a sigma 1 1 a transpose times whole inverse we can look at this inverses because all the matrices this B sigma 2 2 B prime A sigma 1 1 a prime all of them a nonsingular matrices. So, we can open up see that this is just equal to A times sigma 1 2 sigma 2 2 inverse times sigma 2 1 times this sigma 1 1 to the power minus 1 times A transpose, right.

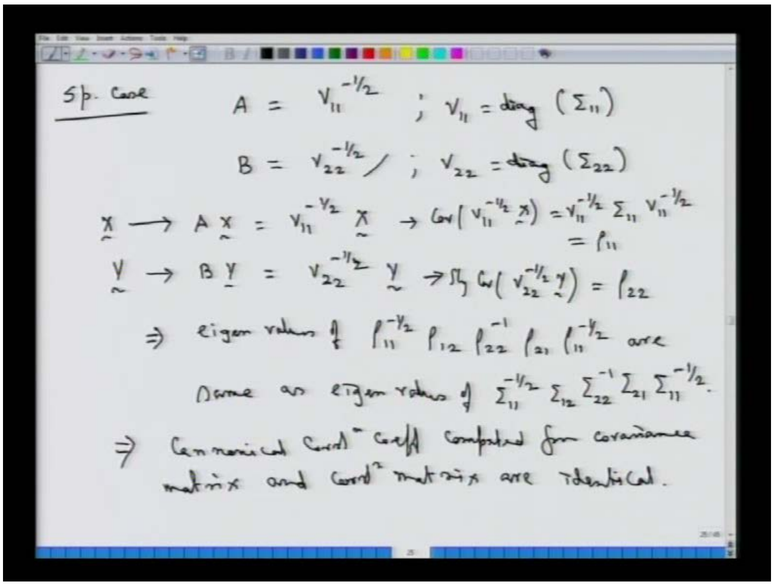
So, this matrix here which place a crucial role in determining what are the canonical correlation coefficients because this matrix is involved in solving this equation here, and the roots are the canonical correlation coefficients what we are going to have is that matrix being transform this matrix is being transformed here to this matrix.

Now, if we look at the roots of this particular matrix we can also say that further roots of A sigma 1 2 sigma 2 2 inverse sigma 2 1 sigma 1 1 inverse times A transpose the I will should write the Non-zero Eigen values of further the non-zero Eigen values of this are same as the Non-zero as the Non-zero Eigen values of… If we consider this to be c matrix then what we are going to have is that I am sorry, this is going to be a inverse here, because we are looking at the last term which is a sigma 1 1 A transpose whole inverse is A transpose here. So, this matrix is the transpose matrix here.

So, the non-zero Eigen values of this matrix times this matrix is going to be same as the Non-zero Eigen values of this matrix multiplied by this matrix. So, Non-zero Eigen values of this are going to be same as the Non-zero Eigen values of sigma 1 2 sigma 2 2 inverse sigma 2 1 sigma 1 1 inverse now what is this is just equal to sigma 1 2 sigma 2 2 inverse sigma 2 1 this now if we look at the previous line here these are the Eigen values are rho 1 star square rho 2 star square rho p star and hence if the Non-zero Eigen values in the transformed situation are same as that of this particular matrix what does that imply this implies that the canonical correlation coefficients are unchanged under this non-zero singular transformation this implies that the canonical correlation coefficients under the nonsingular transformation under the nonsingular transformation transformation A B we are unchanged

So, whether we look at the canonical correlation coefficients from the original setup variables X and y or we look at in the transform setup with A x and B y the canonical correlation coefficients in the 2 steps are going to be precisely the same right.

(Refer Slide Time: 48:51)



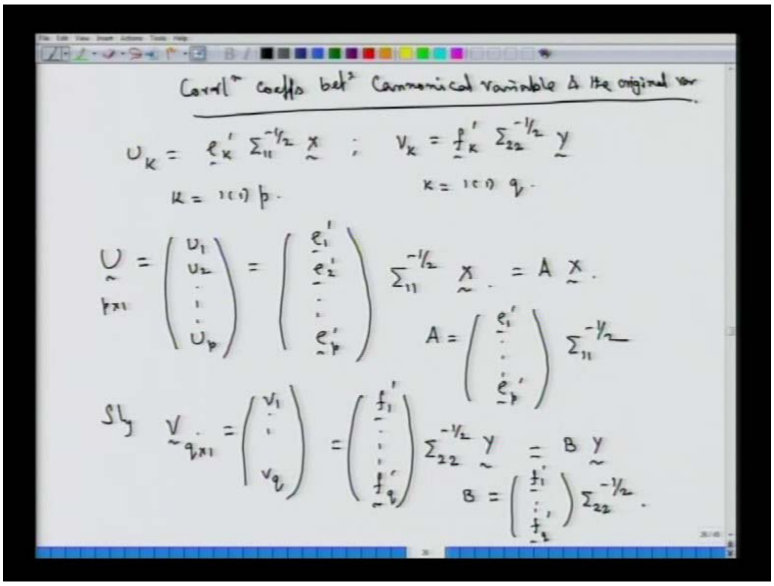
Now, as a special case of this result we will take A and B to be the following, if we consider a as the following matrix which is v 1 1 to the power minus half wherein v 1 1 is equal to the diagonal matrix of sigma 1 1, and if we chose this v matrix as v 2 2 to the power minus half wherein we have this v 2 2 to be equal to the diagonal matrix corresponding to sigma 2 2. Then if we look at this the note is the covariance matrix of this A x matrix this is a as vector other v 1 1 to the power minus half times X, similarly this is X is getting transformed to this A x and y is getting transformed to B y which is equal to v 2 2 to the power minus half times y now what is the covariance matrix of this it would be B 1 1 to the power minus half sigma 1 1 times v 1 1 to the power minus half what will that p that will we just the correlation matrix

So, this is the covariance matrix of v 1 1 to the power minus half X this is going to be v 1 1 to the power minus half times sigma 1 1 into v 1 1 to the power minus half which is nothing, but the correlation matrix if we look at the X original X vector that is a correlation matrix. And similarly the covariance matrix between a covariance matrix of v 2 2 the power minus half times y is going to be equal to v 2 2 to the power minus half sigma 2 times v 2 2 to the power minus half. So, and we will have that as rho 2 2 matrix which sigma the correlation matrix of the y.

So, this will imply that the canonical correlation coefficients let me also write the l 1 most n Eigen values of rho 1 1 to the power minus half times rho 1 2 rho 2 2 the power minus half rho 2 1 times rho 1 1 to the power minus half are same as Eigen values of this sigma 1 1 to the power minus half, because this is what is that A x matrix and B y matrixes Eigen values and this is sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus 1 sigma 2 2 1 sigma 1 1 to the power minus half.

So, this will imply that the canonical correlation coefficients canonical correlation coefficients computed from covariance matrix and correlation matrix are identical, because under the nonsingular transformation there is no change as such in the canonical correlation coefficients. If we take this particular nonsingular transformation as v 1 1 to the power minus half as a and v 2 2 to the power minus half as B, then the canonical correlation coefficients are also not going to change. And hence whether we look at this sigma matrix, and compute the canonical correlation coefficients form the sigma matrix or will look at the rho matrix is a correlation matrix of X and y the canonical correlation coefficients are going to be the same, right.

(Refer Slide Time: 52:48)

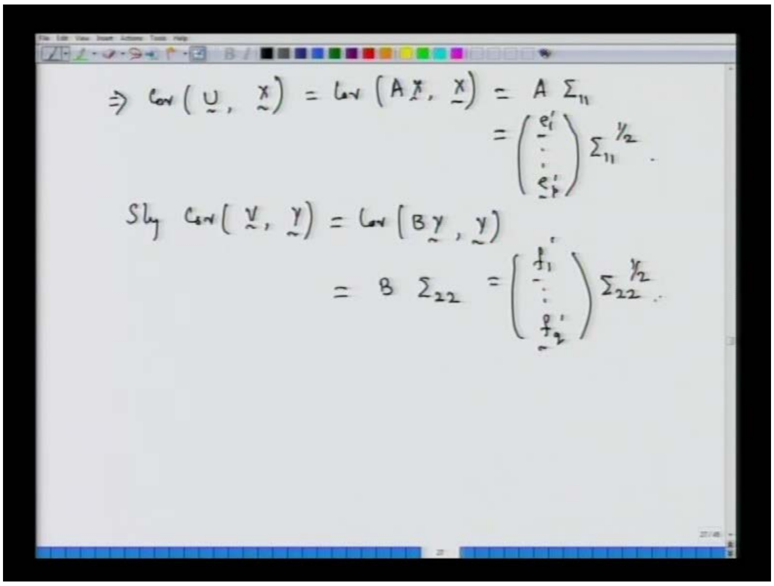


Now, we look at the next important result which is going to look at the following thing that what is a correlation between the canonical variables and the original setup variables. So, we on now looking at the correlation a coefficient between correlation coefficient between canonical variables and the original variables original variables. So, what we are looking at this these are the canonical variables that we have Derived these are e k prime sigma 1 1 to the power minus half times X we have k equal to 1 to up to p we still work with that particular order and we have V k this is given by V k f k times sigma 2 2 to the power minus half times y, and this k is from 1 2 up to q. So, if we look at the U vector the vector containing all the first components of the canonical variable pairs.

So, this is U 1 U 2 and U p. So, this is going to be given by this simple matrix which is e 1 prime e 2 prime and this is e p prime that times sigma 1 1 to the power minus half times X, right.

So, if we have this B just write that as a matrix this a matrix is different from the previous a matrix of course, well right this as A x where this a is equal to this matrix of Eigen vectors A 1 prime A 2 prime e p prime that multiplied by sigma 1 1 to the power minus half. So, if we now one to look at now this is as for as U is concerned similarly 1 can write this v vector this is p by 1 vector this is going to be q by 1 vector this is v 1 v 2 up to v q, and this will be given by f 1 prime f 2 prime f q prime that times sigma 2 2 to the power minus half times this y vector let us denote that by a matrix b times this y vector wherein b similar to this a is given by f 1 prime f 2 prime f q prime time's sigma 2 2 to the power minus half.

(Refer Slide Time: 55:44)



So, if we have U vector equal to this A x and V equal to B y it is very simple actually to see that how we are going to have compute this covariance between this U vector and the X vector now U vector in terms of the X vector is equal to A x times this X vector. So, this is simply a times the covariance matrix of X which is sigma 1 1 and what was a equal to a was equal to our this e 1 transpose e 2 transpose e p transpose that times sigma 1 1 to the power minus half. So, that combine with this gives us sigma 1 1 to the power half. So, this just is the correlation covariance matrix between U and x

Now, we can use similarly the covariance matrix of v vector and the y vector what is that going to turn out that is going to be given by covariance between B y vector and the y vector which is going to be given by b times sigma 2 2 vector. and that is going to be given by f 1 transpose f 2 transpose f q transpose times sigma 2 2 to the power half. So, we will stop this Lecture at this particular point then continue from here.