Applied Multivariate Analysis

Prof. Amit Mitra

Prof. Shramishtha Mitra

Department of Mathematics and Statistics

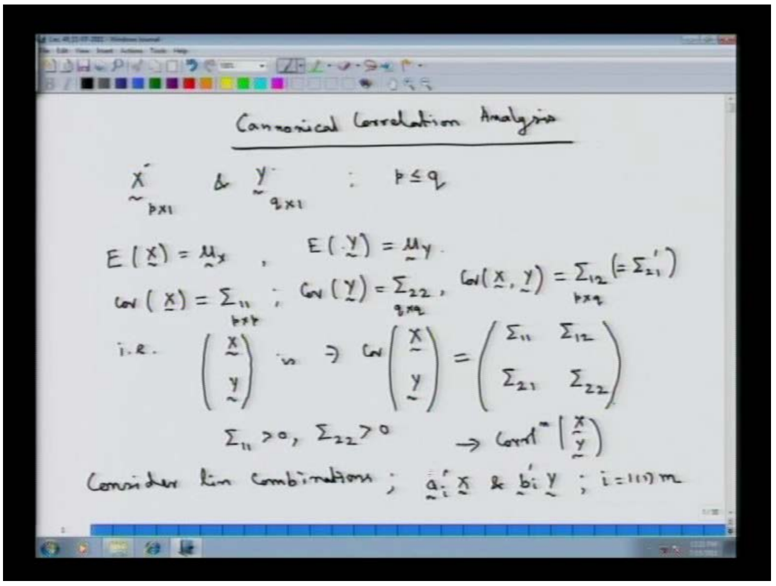
Indian Institute of Technology, Kanpur

Lecture No. # 40

Cannonical Correlation Analysis

In this lecture, we are going to start the concept of cannonical correlation analysis. Now in cannonical correlation analysis, what we try to do is the following thing; we try to explain the covariance structure or rather the correlation structure between two sets of random vectors in terms of fewer linear combinations; what I try to convey is the following.

(Refer Slide Time: 00:44)



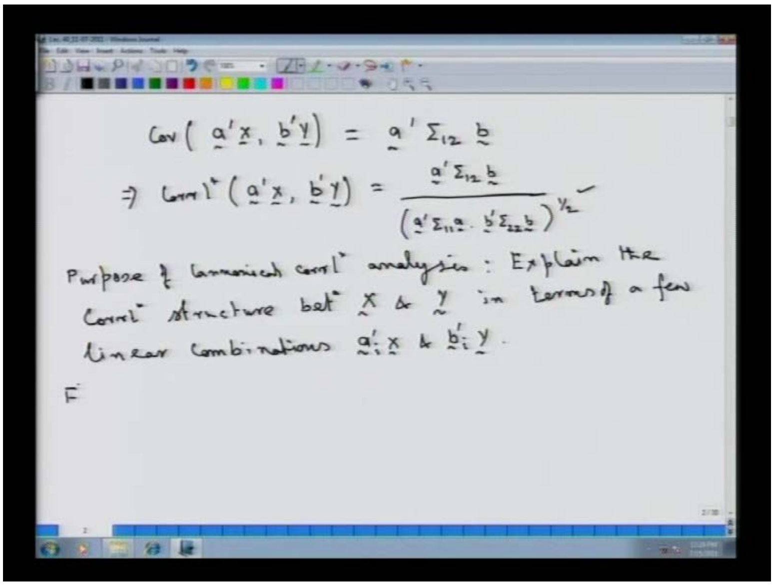
Let us look at this concept cannonical correlation analysis. Now, the setup for the cannonical correlation analysis is the following; we have a random vector p X p dimensional, and we have another random vector Y q dimensional; without loss of generality, we assume that p is less than or equal to q; if not, then we can rearrange and make the first set of random vectors, to be having components p, which is less than or equal to the number of components of the second random vector Y.

Now, these random variables are such that expectation of X vector that is equal to mu x vector, and expectation of Y vector is equal to a vector which is mu y. Now, the covariance matrix, variance covariance matrix of X is denoted by sigma 1 1; the covariance matrix of Y is denoted by sigma 2 2, and the covariance matrix between X vector and the y vector is given by sigma 1 2 say is equal to sigma 2 1 transpose, that is the covariance between Y and X would be sigma 2 1. So, the transpose of that will be equal to sigma 1 2, that is, if we look at this augmented random vector X, augmented with Y, this is a p q dimensional random vector; this is such that the covariance matrix between X and Y is going to be given by the matrix, which has sigma 1 1 in the first block, which is of the order of p by p. So, this of course, is p by p matrix, this is a q by q matrix, and this sigma 1 2 is going to be p by q matrix, because it is a covariance between X and Y. So, this is sigma 1 2 1 1 here, sigma 2 2 here, sigma 1 2 and sigma 2 1.

Now, we assume that this sigma 1 1 is greater than 0, and so is sigma 2 2. So, we will have the two variance, covariance matrices of the respective random variables X, random vectors X and Y to be having the variance, covariance matrix, which is positive definite. Now, as I said that the basic concept in cannonical correlation analysis is we look at such random vectors p dimensional X and a q dimensional Y and then try to explain the correlation structure of these two random vectors; now what will be the correlation? From here, when we have this as the covariance between X and Y given by this, pre and post multiplying this matrix by the square root of the variances, we can get to from here, we can easily get to the correlation matrix between X and Y.

Now, cannonical correlation purpose is to express or rather explain the correlation structure between this X and Y in terms of a fewer linear combinations. Now when we talk about linear combinations, we look at the following; consider linear combinations say a prime or rather a i prime X and b i prime y; for a fixed number of i equal to 1 to up to say any order m right. So, these are linear combinations of possible linear combination of X s, these are linear combinations of Y.

(Refer Slide Time: 04:47)

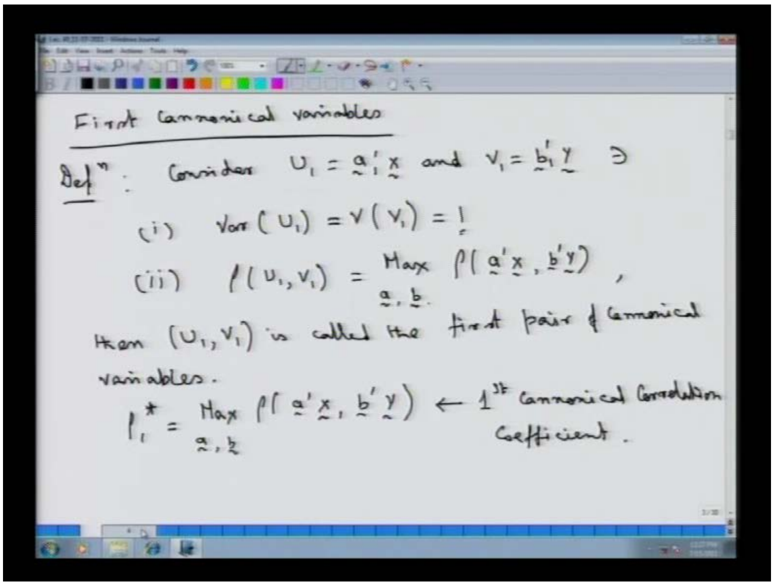


Now, if we look at the covariance between any a prime X and b prime Y, then that is going to be given by the covariance between a i prime X and b prime Y that is going to be a prime, the covariance between X and Y. Now, what is the covariance between X and Y? In the formulation that we have here covariance between X and Y is sigma 1 2; so this is going to be equal to sigma 1 2 times b primes transpose, which is b.

So, this would imply that the correlation between a prime X - linear combination and b prime y - this linear combination is going to be given by the covariance between the two linear combinations, these are scalar variables now, after we are taking linear combinations; this divided by thus standard deviation or the square root of the variances of the respective terms. So, the variances of a prime X will be equal to a prime sigma 1 1 times a, this into the variance of prime y, which is b prime sigma 2 2 times b in our defined notations. So, this is the correlation coefficient between a prime X and b prime Y.

So, the purpose of cannonical correlation analysis is to explain the correlation structure correlation structure between this X vector and Y vector in terms of a few linear combinations; linear combinations of the form that it is a i prime X and b i prime Y right. Now, how are we going to have such linear combinations, and now what sort of linear combinations should we consider, when we are looking at expressing or trying to explain this correlation structure between the two sets of random vectors X and Y; let us look at that particular formulation, how we are going to actually define the pairs of cannonical variables.

(Refer Slide Time: 07:33)

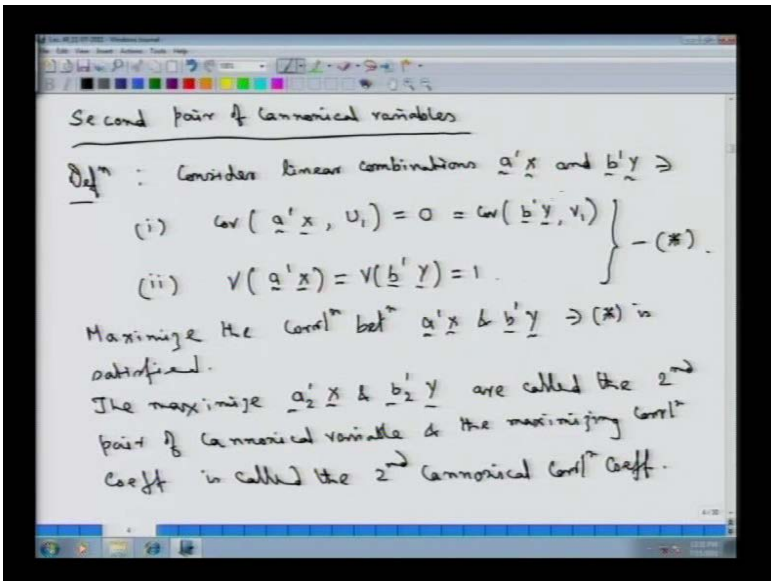


So, let me first give the definition of first cannonical variable pair, first cannonical variables, first pair of cannonical variables actually. So, let me give the definition of this consider U 1 equal to a 1 prime X, and V 1 equal to b 1 prime Y, such that number 1 variance of U 1 is equal to variance of this V I am sorry variance of U 1 equal to variance of V 1 is equal to 1. So, we will have the linear combinations, thus formed; number 2 is that the correlation between this U 1 linear combination and V 1 linear combination is the maximum correlation between a prime X and b prime Y such that this maximization is over a and b.

So, we are looking at U 1 and V 1 to be linear combinations such that U 1 and V 1 has got variance unity, and the correlation between the two linear combinations U 1 and V 1 is a maximum correlation that one can find, if one looks at all possible linear combinations a prime X and b prime y. So, this maximization is over a and b. So, if we have U 1 and V 1 such that this, these two conditions are satisfied; then (U 1,V 1) is called the first pair of cannonical variables right; and this row 1 star, which suppose denotes this maximum over a and b of these linear combinations a i prime X(s) and b prime Y(s), this is what is called is the first cannonical correlation coefficient right.

So, this is how the first canonical, first pair of cannonical variables are formed with the corresponding first cannonical correlation coefficient, which looks at all possible linear combinations; and between X and Y and then tries to look at one that maximizes that particular linear combination; the maximizes the correlation coefficient between all possible linear combinations.

(Refer Slide Time: 10:35)



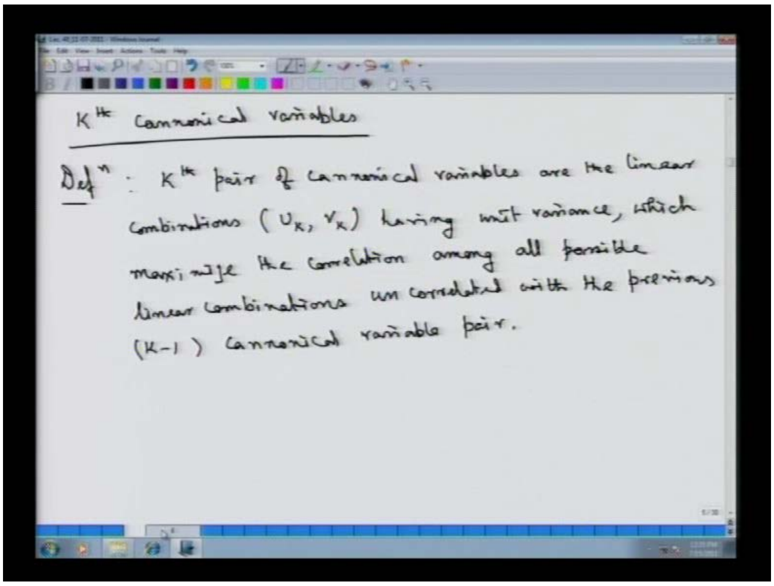
Now, given this as the definition of the first cannonical correlation pair variables; second pair of cannonical variables are going to be defined; cannonical variables; let me first give the definition of the second pair of cannonical variables; consider linear combinations a prime X and b prime Y such that number 1- a prime X is uncorrelated with the first cannonical variable that is covariance between a prime X and U 1, where U 1 is a first cannonical variable, which we have defined in here; so, this is the U 1; so, U 1 is the first cannonical variable.

So, the correlation between a prime X and U 1 is equal to 0 is equal to the correlation, I am just writing covariance, because if covariance is equal to 0, then correlation also is equal to 0; which is equal to the covariance between b prime Y and b prime Y and V 1. So, we will have these linear combinations a prime X uncorrelated with U 1, V prime y uncorrelated with V 1, then under this condition additionally, we will have this a prime X is equal to variance of b prime Y is equal to 1.

So, we are considering all such linear combinations such that these two conditions are satisfied; and under such conditions, let me put a number star here. Then maximize the correlation between a prime X and b prime Y, such that this star is satisfied right. The maximizing a 2 prime X and b 2 prime y say, let me have this notation are called the second pair of cannonical variables. And the maximizing correlation correlation coefficient is called the second cannonical correlation coefficient; cannonical correlation coefficient.

So, what is the second pair of cannonical variables? We look at all such linear combinations now, restricted to the situation that these linear combinations will be uncorrelated with the first cannonical variable, the respective components b prime Y will be uncorrelated with the first cannonical variable pair component, which is associated with Y with unit variances. And then we will try to maximize the correlation between a prime X or possible a prime X and b prime Y such that the condition star is satisfied; and the solution the maximizing, this is the maximizing, the maximizing a prime X and b 2 prime Y are called the second pair of cannonical variables, and coefficient that we are going to obtain by maximizing that expression the correlation between all such linear combinations, such that this star is satisfied is going to be called the second cannonical correlation coefficient.

(Refer Slide Time: 15:08)

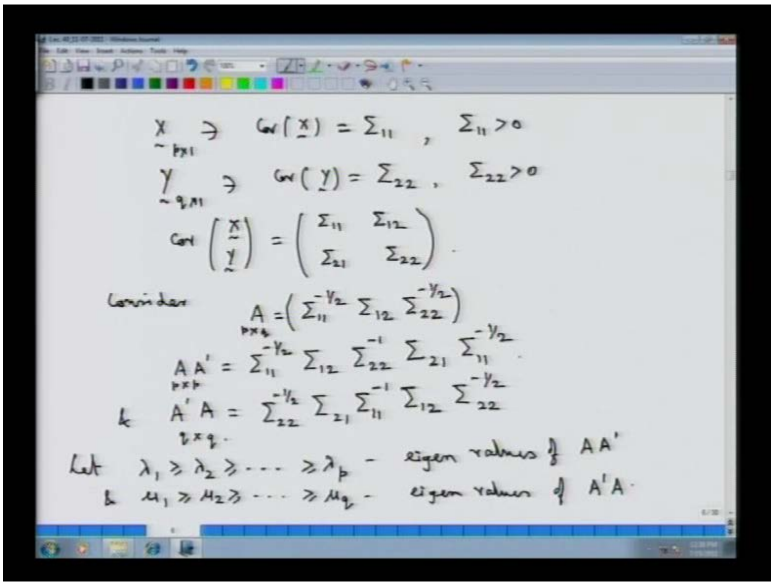


Now, we can write thus the k eth cannonical variables definition, it is going to be that pair, which going to be uncorrelated with all the previous k minus 1 pairs of cannonical variables; and subject to the condition that we are looking now at linear combinations, a k prime X(s) and b k prime X(s), such that we will have all those previously obtained cannonical variables uncorrelated with this particular setup with respect to that particular restriction, we try to maximize the correlation once again in order to get the K eth pair of cannonical variables. Let me write the definition here.

So, the K eth pair of cannonical variables are the linear combinations are the linear combinations say (U k,V k) having maximum correlation unit variance property, having unit variance, which maximize the correlation among all possible linear combinations, among all possible linear combinations, uncorrelated with the previous k minus 1 uncorrelated uncorrelated with the previous k minus 1 cannonical variable pairs. So, this is how the pairs of cannonical variables are actually obtained.

Now, we will have to look at, what finally tells out to be the cannonical variable pairs, and how we can actually get the cannonical variable pairs; sequentially starting from the first pair of cannonical variable pair, and then maximizing the correlation, and then moving forward to second, third, fourth and K th pair of cannonical variables right. Now, in order to actually derive, we will derive the cannonical variables in order to derive…

(Refer Slide Time: 18:01)

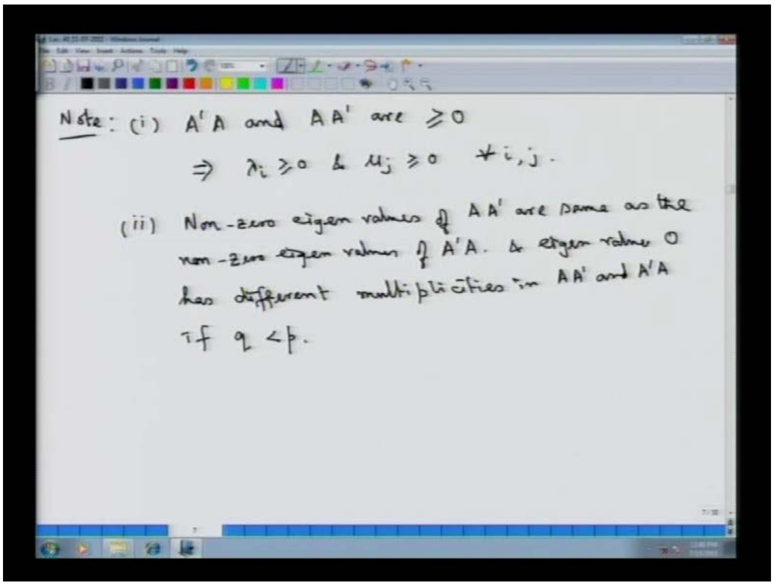


We will look at this structure as before that this has got the covariance structure as equal to sigma 1 1, why? This is p by 1, this is q by 1, such that the covariance matrix of Y is given by sigma 2 2; wherein as before we will assume that sigma 1 1 is greater than 0, sigma 2 2 is greater than 0, and thus we will have the covariance structure between X Y, this vector to be given by sigma 1 1, sigma 1 2, sigma 2 1 and sigma 2 2; now this matrix also is positive definite.

Now, consider a matrix A to be equal to sigma 1 1 to the power minus half, sigma 1 1 remembers positive definite, so sigma 1 1 to the power minus half can be defined, multiplied by sigma 1 2 times sigma 2 2 to the power minus half. Now such a matrix plays a major role in cannonical correlation analysis. Now from this matrix A, if we look at the matrix, which is A A prime; now A A prime matrix would be given by this sigma 1 2, then A transpose of this particular matrix. So, we will have this as sigma 2 2 to the power minus 1, transpose of this would be sigma 1 2 2 1, and then the transpose sigma 1 1 to the power minus half is the same matrix itself. And A prime A is going to be sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus 1 sigma 1 2 times sigma 2 2 to the power minus half.

So, these are two matrices A A prime; A prime A, where A is defined by this matrix. Now, A A prime what is the order of this matrix A? This is p by p, this is p by q, and this is q by q; so this has an order which is p by q. So, this A A prime matrix is of the order p by p, and A prime A matrix is of the order q by q. Now we note the two following important observations that this A prime A, let me define these Eigen value Eigen vector pairs. Let lambda 1 greater than or equal to lambda 2 is greater than or equal to lambda 3 greater than or equal to lambda p, Eigen values of A A prime matrix say, and mu 1 greater than or equal to mu 2 is greater than or equal to mu q, which are Eigen values of A prime A right.

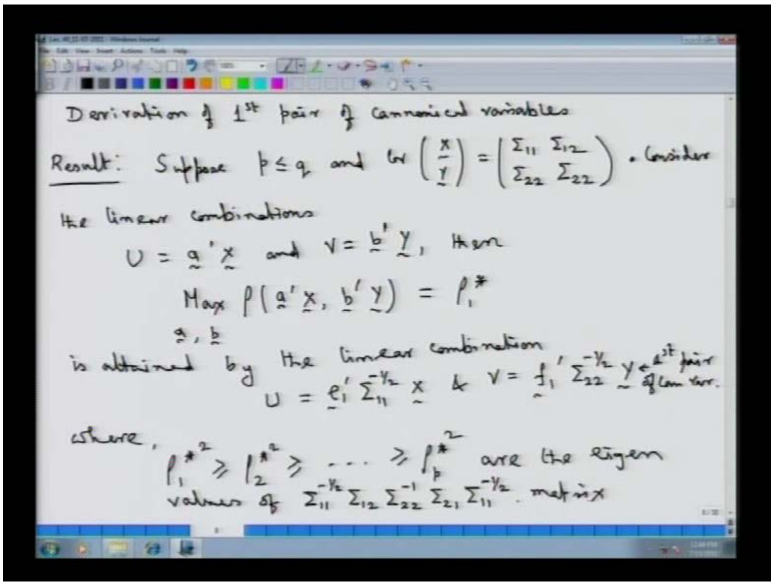
(Refer Slide Time: 21:22)



The two things are quite obvious from the formulation, which I put it as to separate notes that A prime A and A A prime are positive semi definite. So, I write it as greater than or equal to 0, this will imply that this lambda i are greater than or equal to 0 and mu i are greater than or equal to 0, for every i let me just write it as j not to mix up with the 2. So, lambda is greater than or equal to 0, and mu j is greater than or equal to 0, for every i j; what more is that, the norm 0 Eigen values of A prime A, and the nonzero Eigen values of A A prime are identical only the Eigen values 0 will have different multiplicities in a prime A and A A prime.

So, non-zero Eigen values of A A prime are same as the non-zero Eigen values of A prime A; and this Eigen value 0 has different multiplicities in A A prime and A prime A; if we have q to be strictly less than p right. So, this is the simple observation, because we will have A prime A and A A prime being defined through this particular type. And as we will see in the next derivation and furthermore that such matrices play a major role in cannonical correlation derivation of the cannonical correlation variables.

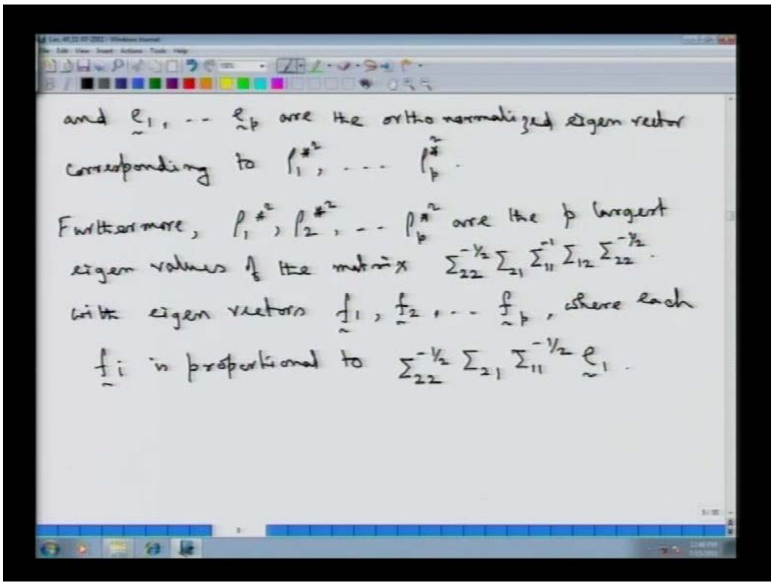
(Refer Slide Time: 23:33)



So, we are moving on to derivation of first pair of cannonical variables. We have the following result; let me first state the result, suppose we have with the previous set up p is less than or equal to q, and covariance matrix of X and Y as we are denoting that by the following matrix sigma 1 1 sigma 2 2 sigma 1 2 sigma 2 2; and consider the linear combination combinations rather consider the linear combinations say U equal to a prime X and V equal to b prime Y, then Maximum over a and b correlation between a prime X and b prime Y, let us denote that by rho 1 star this is a maximizing correlation between this is attained is attained by the linear combination linear combination U; this is the linear combination that we are going to obtain, which is going to be given by e 1 prime sigma 1 1 to the power minus half times X, and V equal to f 1 prime sigma 2 to the power minus half times, we are going to define why this e i prime and f 1 prime are…

So, this U and this V will constitute the first pair of cannonical variables. So, this is the first pair of cannonical variables, where we have rho 1 star square greater than or equal to rho 2 star square is greater than or equal to rho p star square are the Eigen values are the Eigen values of that A A prime matrix, which is sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus half. So, this maximum linear combination that the correlation coefficient between the linear combination rho a prime X and b prime Y, this is going to be attained at the value rho 1 star, where rho 1 star is nothing but they are the square roots of the Eigen values of this sigma 1 1 to the power minus of 1 2 to 1 minus 1 sigma 2 1 sigma 1 1 to the power minus 1 matrix.

(Refer Slide Time: 27:16)

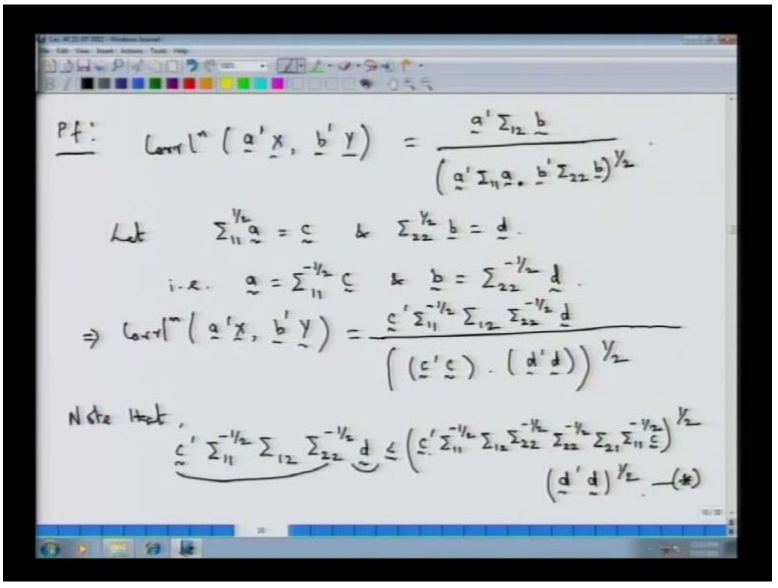


And e 1, e 2, e p are the ortho normalized ortho normalized Eigen vectors; corresponding to our rho 1 star square, rho 2 star square, rho p star square So this is, these are the Eigen values of that A A prime matrix, and these are the corresponding ortho normalized Eigen vectors, corresponding to that; and further more rho 1 star square, rho 2 star square, rho p star square are the p largest Eigen values, p largest Eigen values of the matrix, which is A prime A, so that matrix would be given by sigma 2 2 to the power min us half sigma 2 1 sigma 1 1 to the power minus 1 sigma 1 2 sigma 2 2 to the power minus half right.

So, these are also means what we had seen earlier by defining that A A prime the 2 A, and corresponding A A prime, A prime A are these two matrices, and we had noted that if we are going to clave around with A A prime and A prime A; the non-zero Eigen values of the two matrices will be exactly the same; only the Eigen value 0 will have different multiplicities in the two matrices, which are these two matrices. And we are basically writing that here that rho 1 star, rho 2 star, rho p star are the p largest Eigen values of this particular matrix. And with Eigen vectors as f 1, f 2 and f p, these are also ortho normalized Eigen vectors; corresponding to rho 1 star square, rho 2 star square, rho p star square, which are there corresponding to now, this A prime A matrix, which is given by this.

And we will also have the following, where each f i is proportional to is proportional to the following, which is sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus half times e 1. So, this is the relationship between the Eigen vectors of that A A prime matrix, which where e 1, e 2, e p; and f 1, f 2, f p are the ortho normalized Eigen vectors corresponding to the p largest Eigen values, which match the two, the Eigen values match for the two matrices. And the Eigen vectors, ortho normalized Eigen vectors basically satisfies the relationship that each of these f i prime are going to be proportional to this sigma 2 to the power minus half 2 1 sigma 1 1 to the power minus half e 1. Now this is proportionality constant can be easily computed by noting that the norm of f i(s) should be equal to 1 right. So, this is the entire result, which is going to help us or rather this is the result, which tells us, what is actually the first pair of cannonical variables right?

(Refer Slide Time: 31:22)



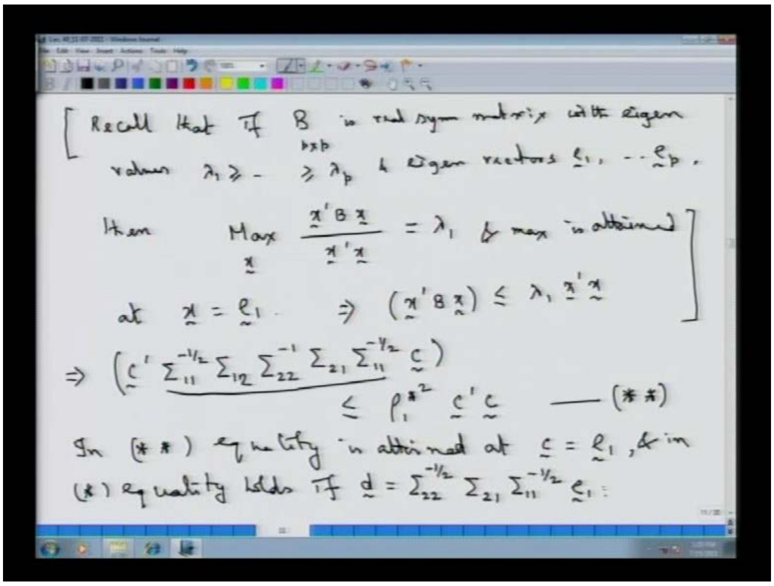
We will look at proving this particular result, because it is a fundamental result in cannonical correlation analysis. So, we will actually prove this result in, and let us in order to do that look at this correlation between linear combinations a prime X and b prime Y. So, as we have seen, this is equal to our a prime sigma 1 2 times b this divided by a prime sigma 1 1 times a that multiplied by b prime sigma 2 2 times b, this raise to the power half right. Now, in this particular expression, let us define this sigma 1 1 to the power half a that is equal to vector c; and sigma 2 2 to the power half times b vector to be equal to a vector which is equal to d, that is our a vector is equal to sigma 1 1 to the power minus half sigma 1 1 and sigma 2 2 are positive definite. So, this is defined, and this b is equal to sigma 2 2 to the power minus half times d.

Now, with this notation here, what we will be having the correlation between a prime X and b prime Y, this is these a(s) and b(s) are now going to be replaced by c(s) and c and d rather. So, what we will be having in the denominator is the following; now a is equal to this particular term, sigma 1 1 half a is equal to c. So, what we will be having is c prime is a prime sigma 1 1 to the power half; so this just is equal to c prime c, this is what this boils down to; and the second term is nothing but d prime d whole raise to the power half; and what we have in the numerator of this particular expression is that a prime is going to be given by c prime sigma 1 1 to the power minus half, and then this is sigma 1 2, which remains as it is; and b is going to be given by this sigma 2 2 to the power minus half times d; so this is what is the correlation coefficient between the two linear combination in terms of the redefined vectors c and d.

Now, note that if we consider the numerator, which is c prime sigma 1 1 to the power minus half sigma 1 2 times sigma 2 2 to the power minus half times d by using Cauchy Schwarz inequality by considering this as one vector, and this as the second part, this will be less than or equal to c prime sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus half times the transpose of that. So, that is equal to sigma 2 2 to the power minus half, once again multiplied by sigma 2 1 sigma 1 1 to the power minus half this raise to the power half that multiplied by this term d prime d whole raise to the power half.

So, this quantity, which is in the numerator is less than or equal to this term, when we are applying the Cauchy Schwarz inequality; let me give an equation number to this, because later on we will be using these equations in the derivation . So, let us now look at this particular term, well a term c is missed out here, this is c; so this is c prime sigma 1 1 to the power minus half sigma 1 2 sigma 2 to the power minus half by combining these two matrices that multiplied by sigma 2 1 sigma 1 1 to the power minus half times this vector c. Now, we will look at this term here, and see if we can provide an upper bound to this particular term here which of course, is very simple we recall this result.

(Refer Slide Time: 36:04)



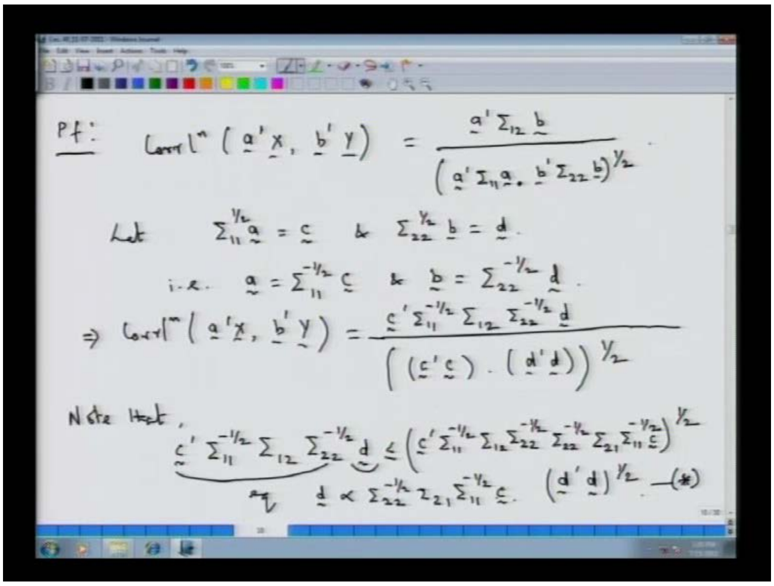
So, if we look at the right hand side here and the first expression if we are trying to get an upper bound of this expression, we make use of the following result that and say as you recall of matrix theory result, recall that if this b is real symmetric matrix with Eigen values and Eigen vectors. Let this B p by p Eigen values lambda 1, lambda 2, lambda p in this order; and Eigen vectors ortho normalized Eigen vectors ortho normalized as e 1, e 2, e p; then we have this following result that if we are trying to maximize the quantity, which is x prime d x that divided by x prime x, this is maximization over all possible x(s); this Maximum value is nothing but equal to lambda 1, the largest Eigen value of this real symmetric matrix. And this maximum is attained at is attained at x equal to e 1 vector right, where e 1 is ortho normalized Eigen vector corresponding to the largest Eigen value lambda 1.

So, this actually will tell us that we can say that x prime B x, if w look at this expression equal to be less than or equal to lambda 1 times x prime x, where the equality once again is attained, if we choose x to be equal to be equal to e 1; in that cases will be equal to 1, and this side will just be equal to lambda 1. So, this is what is the result, which we will make use of in order to provide an upper bound of this particular quantity; why do we look at that we consider we will consider this entire matrix here, which in our earlier notation is equal to A A prime matrix, and then look at its Eigen values and Eigen vectors, and using this particular result we will be able to say that this c prime, where is that this is c prime sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus 1 sigma 2 1 sigma 1 1 to the power minus half times this c vector this, using this expression would be less than or equal to the corresponding Eigen vector of this matrix, which we have earlier denoted by rho 1 star square times this c prime c; with equality attained at I will give this as equation number double star, equation number star is defined earlier for this expression

So, we will have this less than or equal to this terms; in the above, in the equation number double star, equality is attained at now where will the equality be attained? We are looking at rho 1 star square as the Eigen value, the largest Eigen value of this particular matrix; we are going to say that this is less than equal to rho 1 star square times c prime c; and equality of course, will be attained as in the previous general result, at the Eigen vector ortho normalized Eigen vector corresponding to rho 1 star square. So, this equality will be attained at c equal to e 1.

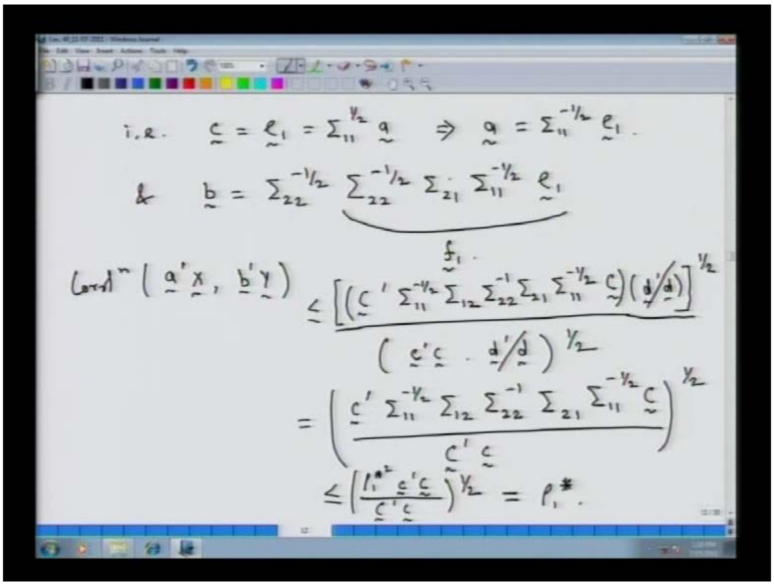
Now, if we have equality attained here, we will also look back at the equation number star as to when the equality is going to be attained in this expression star. Now equality in expression star will be attained if we have this vector proportional to this particular vector. So, we will put in that condition also and in single star equation, equality holds if we have if we have t to be equal to sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus half times e 1; why is that so, because if we are looking at this expression, here equality will hold.

(Refer Slide Time: 41:05)



If we have d equality, if d is proportional to this particular vector, which is sigma 2 2 to the power minus half sigma 1 and sigma 2 1 sigma 2 1 times sigma 1 1 to the power minus half right times this vector c. Now, in double star we require c to be equal to e 1 in order to have equality, in order to attain the maximum value; and here in order to attain its maximum value, we would require this. And since, in double star equality holds for c equal to e 1, here also c needs to be replaced by e 1. So, we will have these two to hold if we want to have equality in all the previous steps.

(Refer Slide Time: 42:01)



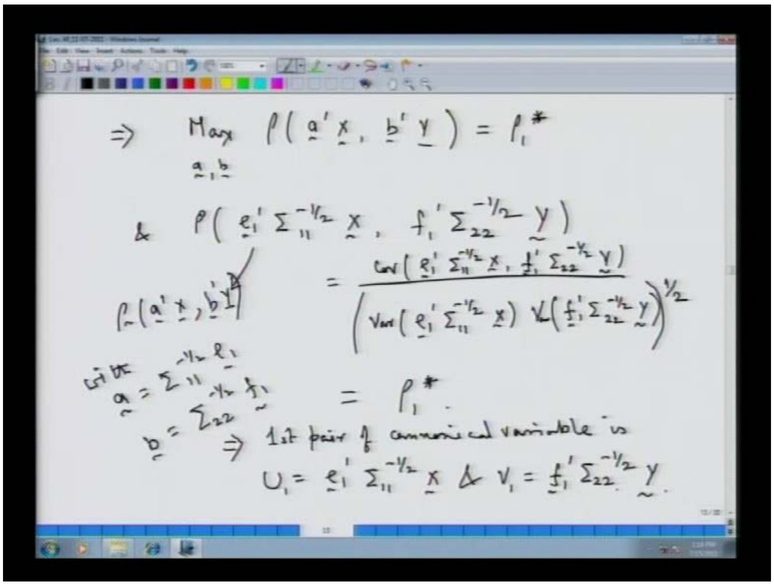
That is we would require this c vector to be equal to a 1, which is equal to c is connectively a vector, because a was the original vector. So, this will imply that the maximizing a vector is going to be given by sigma 1 1 to the power minus half times e 1 right; and the b vector, which is connected with the d vector through this that b is equal to sigma 2 2 to the power minus half times d. And in order to have equality, we would require this condition that d is sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus half e 1, so that we will have b to be equal to sigma 2 2 to the power minus half sigma 2 2 to the power minus half, which multiplied by sigma 2 1 sigma 1 1 to the power minus half times this e 1 right. Now this entire vector is what we are going to call as f 1; look at the result statement here, what we had was this f i is proportional to this particular quantity here, and hence this is what we are going to have hence f 1.

Now, if we look back at the expression of correlation of those linear combinations of the two linear combinations, which were of the form a prime X and b prime Y; we had seen that this was less than or equal to… Let me look at that expression where we had that. So, this correlation between this term was equal to this, which we had seen that the numerator is less than or equal to this; and hence, we will have this correlation between a prime X and b prime Y to be given by this c prime sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus 1 sigma 2 1 sigma 1 1 to the power minus half times c, this divided by this, our c prime c, because that d prime d term cancels out, I think it would be better if I write one more step here; so, this multiplied by this d prime d; we had the entire thing here raise to the power half.

What we are doing is that we are looking at this expression here that the correlation between a prime X and b prime Y is equal to this. So, if we put the upper bound of this numerated term here, we will have the correlation coefficient between a prime X and b prime Y to be less than or equal to this to the power half, and this to the power half divided by c prime c d prime d to the power half. So, what we have is this less than or equal to this, this divided by c prime c multiplied by d prime d whole raise to the power half. So, as we that these two terms cancel out, and this is equal to c prime sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus 1 sigma 2 1 sigma 1 1 to the power minus half times c, this divided by c prime c whole raise to the power half right.

And this expression by using this result what we had done here this one numerator is further less than or equal to rho 1 star square times c prime c. So, what we have finally is that this expression is less than or equal to this one is rho 1 star square, this is rho 1 star square multiplied by c prime c, this divided by a c prime c entire thing raise to the power half, so that this term is just equal to rho 1 star. So, this correlation between a prime x and b prime y is less than or equal to rho 1 star; so, that rho 1 star is going to be attained if we choose a equal to this, and b equal to this particular term right.

(Refer Slide Time: 46:40)

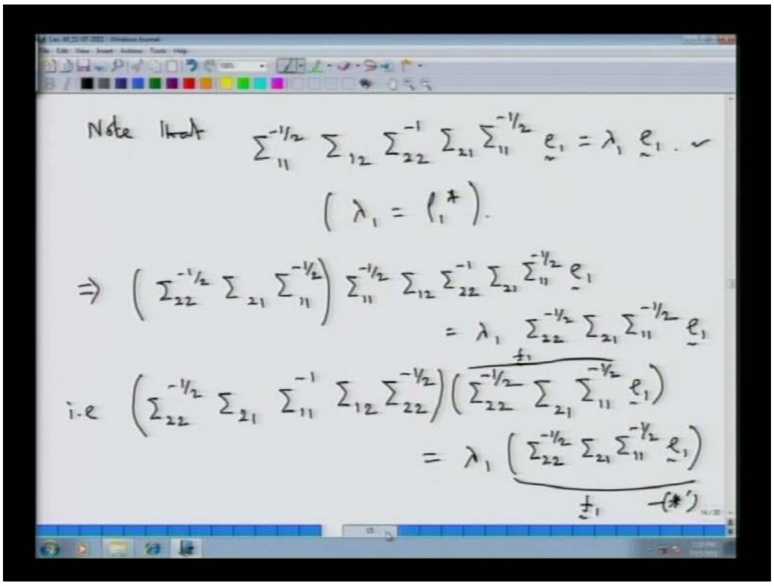


Let us finish that. So, this will imply that maximum correlation between a prime X and b prime Y, maximum over all possible a and b vector, this is equal to rho 1 star; and the correlation coefficient between the two vectors, which is e 1 prime sigma 1 1 to the power minus half X and f 1 prime sigma 2 2 to the power minus half Y; so, these are the linear combinations that we are talking about. So, by choosing b this is just b equal to sigma 2 2 to the power minus half times f 1, where f 1 is the Eigen vector ortho normalized corresponding to that A prime A matrix.

So, what we are going to have is this correlation; it is elementary actually to look at what this expression is equal to? This is e 1 prime sigma 1 1 to the power minus half times covariance between X and Y that is equal to sigma 1 2 this times sigma 2 2 to the power minus half… I am writing the entire thing here, not exactly the covariance term. So, this correlation is equal to the covariance between e 1 prime sigma 1 1 to the power minus half times X; f 1 prime sigma 2 2 to the power minus half times Y that divided by the variance of the respective terms e 1 prime sigma 1 1 to the power minus half X that into the variance of the other term, which is f 1 prime sigma sigma 2 2 to the power minus half times Y.

So, this is that particular this whole raise to the power half of course, this is that we are looking at this rho a prime X and b prime Y with our a equal to the maximizing coefficient, which is sigma 2 2 to the power minus half e 1. So, this is sigma 1 1 to the power minus half times e 1; and with the b, which is sigma 2 2 to the power minus half times f 1; and this straight away leads us to the value, which is rho 1 star. So, that is the maximum correlation coefficient attained with a equal to this, and b equal to this particular term. So, this will imply that the first pair of cannonical variable is given by U 1 say U 1 equal to sigma, that equal to e 1 prime sigma 1 1 to the power minus half times X; and the second component V 1 is equal to our f 1 prime sigma 2 2 to the power minus half times this Y. So, this is a pair of the first cannonical correlation.

(Refer Slide Time: 50:37)



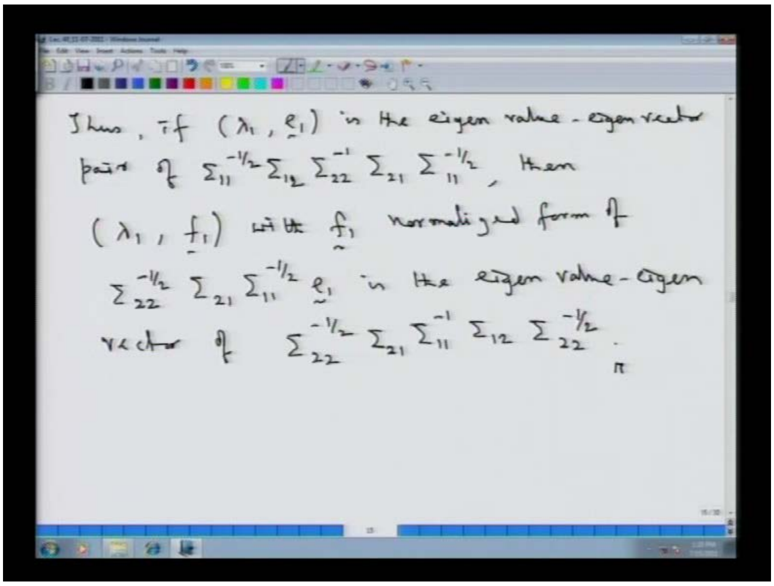
Now, from this expression also, we had another point to prove in the result. Now, note that we have this sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus 1 sigma 2 1 sigma 1 1 to the power minus half times. Now this is the matrix, which has Eigen values as lambda is and Eigen vectors ortho normalized as e i (s). So, this times e 1 vector will be equal to lambda 1, I will just write lambda 1 or rather lambda 1 is nothing but this rho 1 star only right; so, that is the Eigen value; so, I am just writing lambda 1 that times this e 1.

Now, starting from this particular equation here, this will imply that if we pre multiply the left and the right hand side with this expression, which is sigma 2 2 to the power minus half sigma 2 1 times sigma 1 1 to the power minus half. So, this matrix if we pre and post multiply on both the sides what we are going to get is the following that it is sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus 1 sigma 2 1 sigma 1 1 to the power minus half times e 1 that is equal to lambda 1 or rho 1 star times sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus half times e 1.

Now, let us look at the left hand side here, this sigma 2 2 to the power minus half sigma 2 1 and this expression is sigma 1 1 to the power minus 1, we take the next term also sigma 1 2 and split this particular term in terms of sigma 2 2 to the power minus half and another sigma 2 2 to the power half. So, let us write this as sigma 2 2 to the power minus half times whatever is left. So, this has still sigma 2 2 to the power minus half times sigma 2 1 into sigma 1 1 to the power minus half times this e 1 that equal to lambda 1 or rho 1 star times sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus half times e 1.

So, we are basically denoting this as f 1 that is, this is f 1 and if we look at this particular matrix here, and go back to the result that we had stated when deriving the first pair of cannonical variable, there you will find that this is the matrix that we had to talked about at the end of the result, which is that furthermore these are the p largest Eigen values of the matrix sigma 2 2 to the power minus half sigma 2 1 1 to the power minus 1 sigma 1 2 sigma 2 2 to the power minus half with these as Eigen values, where f is are proportional to this particular quantity; and that is what precisely we have obtained that this is the matrix, the Eigen vector is f 1, which is proportional to this term that is equal to lambda 1 lambda 1 is rho 1 star times this f 1. Thus we conclude from the previous expression what we have got that if we have this equation to be satisfied say star prime equation.

(Refer Slide Time: 54:37)



Thus, if we have (lambda 1,e 1) is the Eigen value, Eigen vector pair; Eigen value Eigen vector pair vector pair of this sigma 1 1 to the power minus half sigma 1 2 sigma 2 2 to the power minus 1 sigma 2 1 sigma 1 1 to the power minus half; then this (lambda 1,f 1) with f 1 normalized form of form of the vector, which is sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus half times e 1 is the Eigen value, Eigen vector pair Eigen value Eigen vector pair of this sigma 2 2 to the power minus half sigma 2 1 sigma 1 1 to the power minus 1 sigma 1 2 sigma 2 2 to the power minus half, that concludes the proof of this particular result.