Neural Network and Applications
Prof. S. Sengupta
Department of Electronic and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 32
Introduction to Principal Components and Analysis

Today's lecture is going to be on the Introduction to Principal Components Analysis. It is a new chapter for us and in short form principal components analysis is referred to as the PCA in the literature. So, this is what I want to introduce today. Now, before we introduce this, let me tell you about the basic motivation of this technique. First of all we have to understand that and thereafter we will be presenting the theory of the principal components analysis.

And once, the theory is known to us we will also see that, what is it is relevance to the neural network. In fact, as you will see later on, that the principal component analysis problems are easily solvable using the artificial neural networks. So, that is where it has assumed. So, much of importance in the study of neural network is a very important thing.

Now, one thing which has been discussed over the last chapter is that, we deal with the input data. And input data is in the form of an m dimensional vector. So, we consider x vector to be an m dimensional vector, which in general we know is nonlinearly separable. So, to make it linearly separable what we do is to map this m zero dimensional vector into a higher dimensional, because we argued last time that in higher dimension the separability increases.

And then, we can once the separability is ensured by mapping it into a higher dimensional space or by mapping nonlinearly, not that we always require a higher dimensional mapping. But, what is essential is we need a non linear mapping preferably on a higher dimension and then the classification accuracy definitely improves. And the classification is to be carried out based on that, but there are some cases, where the dimensionality of the input itself is considered to be excessively large.

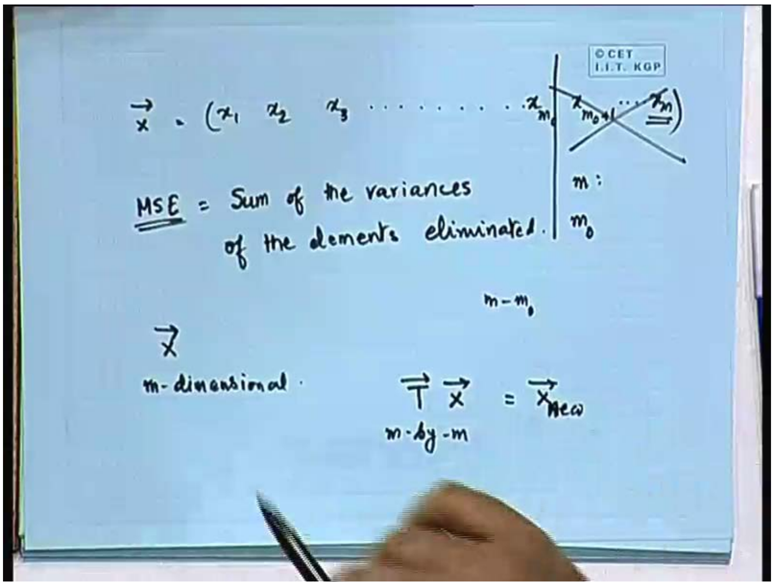
And in such cases the problem that one has to face is that, if we have to deal already with a large dimensional input vector. And we want to map it nonlinearly into and even higher dimensional. Then of course, the size of the neural network that increases enormously.

So, where the number of dimensionality of the input itself is very large, then our objective should be to reduce the dimensionality.

And how, there we have to find out that whether in the input data itself, there is any redundancy that is present. If supposing it is m dimensional input, then what we have to find out is that within that m dimensional data that, we have got is there any redundancy. Whereby, we can let us say we keep some of the inputs, let us say we keep m 0 number of inputs we keep out of m.

And we discovered, where m 0 is a number which is less than m and we discovered the remaining m 0 minus m data points all together is that possible to do. It is only possible to do if, whatever we are eliminating is really something very insignificant, then we can afford to do given an m dimensional data set.

(Refer Slide Time: 05:32)



Let us say we have got data this x vector composition is let say x 1, x 2, x 3 all the way up to let us say here x m 0, then x m 0 plus 1, etcetera up to let us say x m. Supposing this is the composition of the vector. And we say that this m is very large number and we do not want to deal to with such a large number. We want to reduce that to m 0. In that case what we are simply deciding upon is that, we will keep this much and we will truncate the rest, the rest we will not use at all.

Not use means, we will keep it as 0 data or to say in the representation of the data, these m 0 points should be good enough for us, we do not require the remaining m minus m 0 number of data points that we have got. Now, we can do that only when we are eliminating some insignificant data. Now, how to make sure that, whether something is insignificant well, that analysis has to be done.

Because, otherwise somebody giving me an m dimensional data and I just decide on my own, that I will keep the first m number of points. And all the remaining first m 0 number of points and all the remaining data I will truncate to 0, if I do that definitely what I am doing is I am introducing error. And what will be the amount of error that I incorporate, in that process by simply chopping of the data mercilessly.

In that case, the mean square error that I will be incurring in the data should be equal to what, should be equal to the variance of the data which I have eliminated. So, the MSE will be equal to the some of the variances of the elements, which are eliminated, so that should be our mean square error. Now, naturally we can effort to tolerate some definite mean square error, if the some of the variance of the elements eliminated, they are a small quantity.

Only, if it make some sense whatever we eliminated results in a very insignificant amount of mean square error. Now, how to do that, because we are in the x vector space, x vector is an m dimensional space. And how do we decide that, whether it contains anything that is insignificant or not or how do we decide that this way we will be picking up first m 0 number of significant coefficient.

So, for that matter unfortunately the x domain will not be of useful to us, because we cannot say that the first m 0 points are important. And the rest m minus m 0 points are not important, because the next data that arrives, the next pattern that we choose, could be having more significant information here. Some other time may be, some other points are important. So, naturally in the x domain we cannot really decide that, something is of low variance, so we can eliminate that.

So, to do that, what we do is that, we have to map this x vector into another vector space which, we achieve by transforming this x vector. So, let us say that x is an m dimensional vector, so if we can design a matrix, let us say that we design a t matrix that is m by m. And we just pre multiply the x vector with this T matrix, then what results, this is x is m dimensional vector, so what results is another m dimensional vector.

So, this is the new vector T x is a new vector, let us say this is x n that is a new or x new let us say. So, we have got a new vector x new and it may be possible that in the transformed space, if we transform the x into x new could be that in the transformed space out of the m number of points. That, we have got in x new m 0 points are significant and the rest m minus m 0 points are insignificant, it may be easy for us to decide in the transform space.

So; that means, to say that we have to very cleverly design in this transformation t. In fact, this transformation design is something that we are doing in several fields. Especially, several fields of Electrical Engineering, Electronic Engineering fields related to the electrical sciences, we require the design transforms enormously.

One of the things which all of you must have done is the use of the transforms, use of orthogonal transforms in order to which serves as the basis functions. And we can reconstruct signal on the basis of using those orthogonal functions, we known about Fourier transform, we known about the cosine transforms. And it is discreet versions in the form of discreet Fourier transform, discrete cosine transformers, etcetera.

All these things are giving us this t matrix in some form, it is based on some definite kernels which exists. But, one difficulty is that this is based on some fixed kernels, you talk of the d c t or d f t these things, they are based on some fixed kernels. Which are say for example, in the case of Fourier transform, the kernel is then exponential complex exponential kernel, consisting of cosine terms and the sign terms.

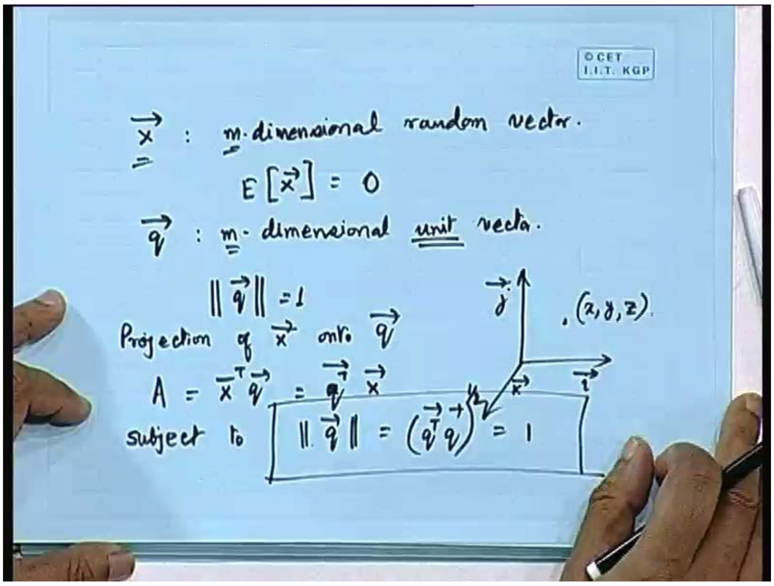
Or the discreet cosine transform that consist of the cosine terms used as the kernels as the basis function. The difficulty is that, if instead of designing it based on the fixed kernel we can design our own kernel and try to find out the optimal kernel or the optimal transformation. So, that in the transformed vector that we are getting x new, it is possible to find out an optimal, if we can find out an optimal transformation t such that, in the x new, if we eliminate the last few points, then we will be eliminating the terms with very low variance.

And eliminating terms with very low variance will lead to lower mean square error, so that we should be able to use those kind of transforms in an optimal sets. So, that is exactly what we are attempting to do, so now our main objective will be to design a good transformation function t.

Such that, in the transformed space that is in x new that we have got, we have got low variance terms, which can be easily chopped off. So, with that motivation in mind we are designing, what is called as the principal component analysis. And I will also touch upon the point, that why at all it is called as the principal component what is, so principal in it, those things will come little later on.

But, let us first of all develop the theory which will enable us to design a good transformation t, that is what with that motivation in mind let us proceed. So, what our objective is, to maximize the rate of decrease of variance, because if we can maximize the rate of decrease of variance in that case, the last terms that we have got, they will be having very low variance. And we can easily get away with it, that is what are we are looking for.

(Refer Slide Time: 14:54)



So, to give the mathematical analysis for that let us consider m dimensional input vector, in fact that will be treated as a random variable, a random variable used as a vector. So, this is an m dimensional random vector and further more what we assume is that, the expectation of this x vector is equal to 0. Now, this can develop some doubt in your mind, that can we assume all the time, that the input that we are getting should have an expectation equal to 0.

That means, to say that the positive coefficients and the negative coefficients are equally likely. In practical data's it is very often not in fact, most of the cases it is not the case, you may be having always some positive values, defined as the elements of the x vector in such case what you do. So, what you do is that simply you subtract the mean from the individual elements, if it is not a 0 mean random vector, you make it 0 mean by subtracting the mean from it.

So, if x is a non zero mean vector subtract x bar from it and get a 0 mean random vector. So, you are assuming that you have done all that. And you have got a 0 mean random vector x available with you and then, we are defining that there is another m dimensional input vector which we are calling as q. So, we are assuming this to be an m dimensional unit vector I will explain the significance of that.

So, this is an m dimensional unit vector, dimension same is that of x, this is m dimensional, this unit vector is also m dimensional. Now, what we want to do is that, we want to project this x vector, the given x vector into, this m dimensional unit vector space. Something like this, that supposing we have got a three dimensional input vector, let us say we have got i j and k available as input vector. And supposing we have got a point in space, which is having coordinate let us say x, y, z.

So, what we do we in order to represent this in the vector form, what we do is that we project this point into i direction, z direction and k direction, it is a projection that we get and then we get i x, j y, k z like that we get. So, essentially what we have to do is in this case is, that this is a three dimensional thing projected into a three dimensional unit vector space. And likewise here, we have got an m dimensional random vector, which we are projecting into an m dimensional unit vector.

So, when we projected into now because it is a unit vector, so one property that we must have is that the Euclidian norm of this q vector, should be equal to 1, this is unit vector. So, now what we can represent is that, the projection when we consider the projection of the x vector on to q, if we are going to write that let us say A is the projection what will be the projection.

Student: ((Refer Time: 18:58))

X.

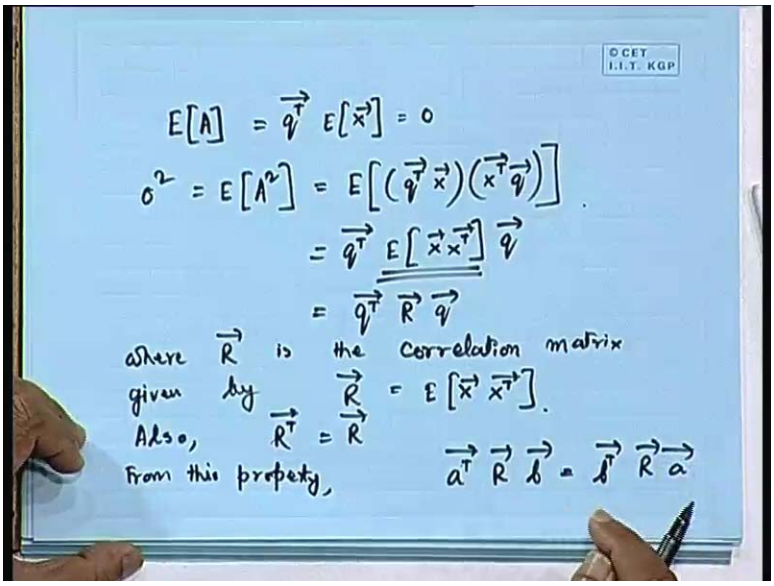
Student: Transpose q.

X transpose q very good, in fact this is going to be a scalar quantity, it is a projection. So, it is a scalar quantity it is a dot product that you said very rightly. So, it is x transpose q this is the projection of x vector on to the q vector, the m dimensional input vector, either you write it as x transpose q or you write it as q transpose x it means one and the same.

So, it is as if to say that projecting q into x or projecting x into q anyway either way it can be looked at. But, this is the representation and this we have subject to the norm of q vector equal to 1. So, the norm of q vector in this case can be written as what, it can be written as q transpose q, but q transpose q gives us norm of square.

So, what we have to do is write it as q transpose q to the power half and this is equal to 1. So, this is the constraint on that vector, because this is unit vector, so that is why norm of q equal to 1 and this is the projection. So, what we can write down now is that, in the projected space what is the exception and what is it is variance, this two quantities we want to determinate.

(Refer Slide Time: 20:45)



So, given this let us see what is going to be the expectation of A, A means the projected values to say. So, what is the expectation of A.

Student: ((Refer Time: 20:54))

Y 0.

Student: ((Refer Time: 20:57))

Yes, you see expectations of A means expectation of this entire quantity, x transpose q, now naturally q not a random variable. So, q comes out of the expectation operator, so it is q vector coming out and expectation of x and what is the expectation of x that is equal to 0. So, it is quite obvious that this is equal to q transpose expectation of x, which is equal to 0.

And what is the variance, the variance of this quantity is to be represented as the expectation of A square. And expectation of A square can be written as A square, how can we write A square, A square we can write as q transpose x this thing multiplied by x transpose q is that right, q transpose x, x transpose q.

So, this can be now represented as q transpose and q, they will be coming out of the thing. So, this is q transpose and within the expectation operator what are we going to have, it is going to be x x transpose and this is going to be q, this is q transpose and this is q. Now, what is this quantity expectation of...

Student: ((Refer Time: 22:32))

Yes, it is the correlation in fact, it is x multiplied by x is outer products with itself. So, which is a correlation matrix of the x vector, so this we are going to represent by the matrix R, this is going to be a matrix, this is an m by m matrix that results x x transpose, x is of dimensional m, so it is m by m vector correlation vector. So, we are going to have it as q transpose the correlation matrix R times q vector where, R is the correlation matrix given by R matrix equal expectation of x x transpose follow.

Also we note here, that this correlation matrix will be; obviously, symmetric, because you are taking the same vector x and x transpose, x transpose is the different way of writing the x vector only. So, you are having absolutely symmetrical terms, so that is why R transpose and R, they will be the same.

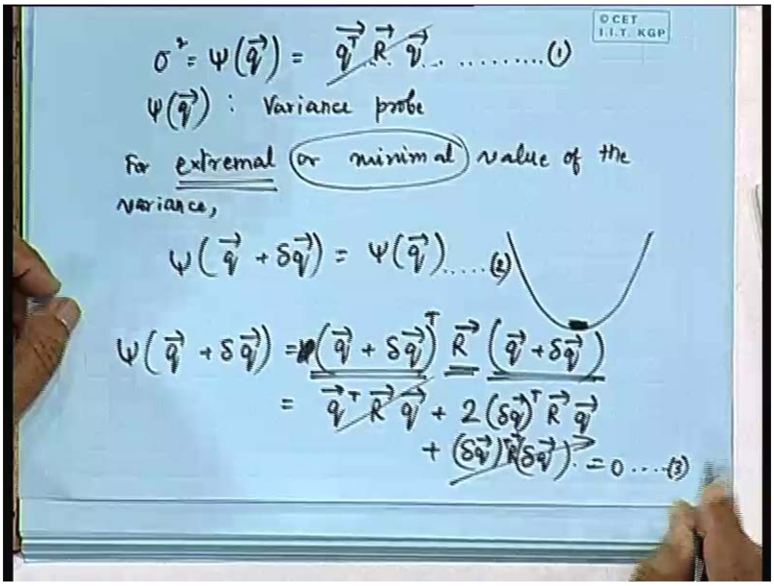
And in fact, from the property of this sort of matrices, which are similar matrices, properties of similar matrices also tell us that supposing we have got two vectors a and b which are of dimension m, we take two vectors of dimensional m. And those vectors are a and b. So, from this property means from this symmetric property, from this property it is possible for us to right a transpose R b vector do you know this, this is a very important property a transpose R b is equal to b transpose R a for symmetric matrix it will be valid.

Now, what we are more interested in is this expression, this the variance expression and what we want to do, we want to minimize the variance. Now, this correlation matrix is not exactly under our control, because this is controlled by the x vector itself. So, this is controlled by the random vector, the random process itself will control R, so we do not have any control of that.

So, what is in our hand in trying to design the minimum variance in trying to have a minimum variance design, the proper choice of q. That means, to say that we kept very proper an appropriate choice of q, it is possible for us to going for some form of minimization of the sigma square. So, this is what we should look for; that means, to say that using the q vector as a search, if we use the q vector as a search to find the minimum.

In that case, we can use the q vector as if to say like a probe to find out, that now you vary q vector and find out that where do you find the minimum of the variance. So, this variance sigma square that we have got is defiantly some function of the q vector.

(Refer Slide Time: 26:44)



So, what we have to find out is that, we have to find out a function psi of q which is equal to this q transpose R q. So, with basically the q that we are going to search through is a function like this, so psi q, so we have to minimize this function psi q. So, this is exactly equal to sigma square, so we write sigma square is equal to this equal to q transpose R q and this we are going to write as our equation number 1.

So, psi q because we are finding out the variance based on this, this we are calling as the variance probe. Because, by varying this q we are going to find out, that where lies the minimum, so at the point where this psi q has got some extremal values, extremal values means in this case the extremal is definitely going to be in a minimal sense.

The minimal will indicate the form of extremal, that we are looking for because on the maximum side yes, variance can reach anything sky is the limit there. But, for the minimum yes there should be a one extremal point, that we are looking for. And at the point of extremal what happens is that, if the q has already found out some extremal, in that case if we try to alter a q slightly, if we perturb the q vector from it is minimal position q.

If we perturbed the vector q by an amount, let us say delta q, then at the place of extremal or at the place of minimum of the variance probe, we are going to have psi of q equal to psi q plus delta q is that understood. So, at the for extremal values for extremal or in this case minimal value of the variance, what we should have is a psi of q plus delta

q. What is the significance of the delta q I will just explain it once more, this is equal to psi q and what is the significance of this.

That means, to say that q is a vector which we have already found out to be a minimum. So, if q is a vector which reaches the minimum of the variance, in that case at the minimum if we try to perturb the vector q slightly, it would not change the variance much, because it has reached the bottom. So, this is the variance probe it has reached here, so that means, to say that if you try to disturb it little from q, you want to disturb it to q plus delta q, the variance more or less remains the same.

Student: ((Refer Time: 30:01))

But, one thing that local maxima do not arise, local maxima cannot arise in this case, because from a set of data you can never try to achieve. Because, maxima is that you can go anywhere in the maxima, you have to consider only the minimal thing and for minimal you will get a point case arising like this.

Student: ((Refer Time: 30:40))

Yes, so we will be coming to that concepts will be little more clear later on. Here, there is a definite minimal existence that we are looking for. Now, if we assume this condition for the minimal, then by the application from the very definition of the variance probe, because the variance probe what we have decide. Yes, this is equal to q transpose R q that is right, so q transpose R q how do we write down. So, if we write down q plus psi plus delta q if we want to write, then this becomes equal to psi q plus delta q transpose R transpose r times what q plus delta q followed, this transpose R then q plus delta q.

Student: ((Refer Time: 31:59))

By applying the same definition.

Student: ((Refer Time: 32:04))

It is not psi that is wrong yes thank you very much for the correction yes. So, this by applying the definition of equation 1 only, psi q plus delta q is a is equal to q plus delta q transpose R q plus delta q. So, now what we have to do is to expand this whole thing and by expanding we will be getting a term like this, we will be getting in fact, three terms one is q transpose R q this is one term, then we will be getting two times delta q transpose R q.

And then we will be getting the third term as delta q transpose delta q.

Student: ((Refer Time: 33:07))

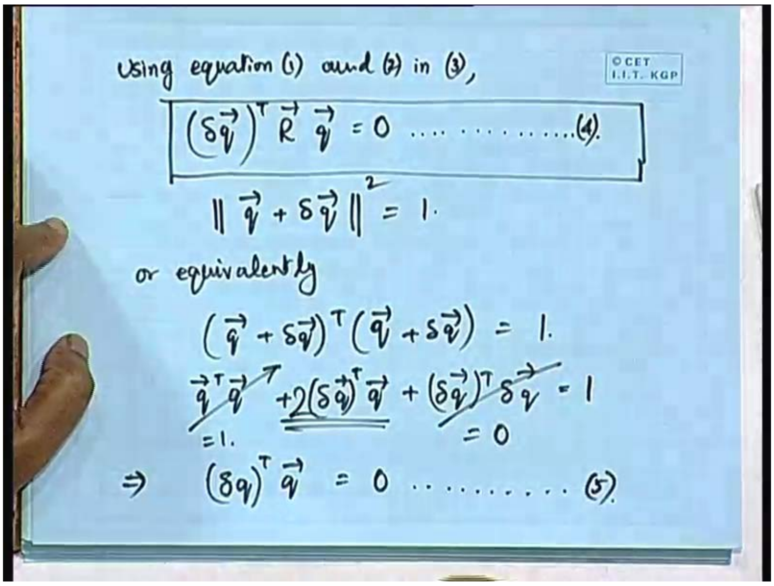
R delta q correct delta q times R delta q, thank you for correction again. So, there are three terms in this expansion, now out of these three terms, this one that is delta q transpose R delta q this term is not of significance, why because delta q is a small perturbation that we have assumed. So, delta q values itself being very small, the contribution of this delta q transpose R delta q will be much less, as compared to the contributions of the first two terms.

So, the first two terms will be dominating and if we assume this extremal or the minimal condition over here then, we can one thing that I would like to say is that even if you let us say that for the sake of argument, you treat it as extremal, if the student finds it difficult to feel convinced, that why cannot we get a maximal. Let us say extremal, let us proceed with analysis, then we will study it is minimal aspect in a better way.

Now, external if we take then this condition is getting valid, so this term is equal to q transpose R q. So, which means to say that if this thing we if we equate psi q plus delta q with psi q, in that case this q transpose R q here and here getting canceled. So, what we are getting is that this time anyway is neglected, so this is equal to 0 and this gets canceled with this. So, what remains is that delta q transpose R q is equal to 0, because psi q and psi q plus delta q are the same.

So, this if we say equation number 2, so then what we obtain is that delta q transpose R q is equal to 0 anybody who found it difficulty in following this step.

(Refer Slide Time: 35:35)



I may be going a bit too hurriedly, but tell me if anything you are missing, ((Refer Time: 35:46)) this is by comparing, if we say that this is equation number 3. Then, we can simply using equation 1 and 2 in 3, this is what we get at the point of extremal. Now, this is the let us say this one we are calling as equation number 4. Now, one thing is there that what we said in delta q, we have perturbed q by perturbing q we have got p plus delta q.

Now, one constraint that will always be followed is that, because we have assumed q to be unit vector even after perturbation it has to remain as a unit vector only. So, that the Euclidean norm of q plus delta q should be equal to 1, so what we enforce upon is that because q is a unit vector. So, the norm of q plus delta q the Euclidean norm of q plus delta q should be equal to 1.

Or equivalently we can express it as this square, if we take equal to 1. In that case q plus delta q, that is the term this transpose times q plus delta q should be equal to 1. And now, we can simply again we expand it and we find that what happens is this is q transpose q plus we will be having delta q transpose q plus again q transpose delta q which are all the same. So, this will be 2 times delta q transpose q plus there will be a term delta q transpose delta q, which is again negligible term this should be equal to 1.

So, this is equal to 0 almost equal to 0 and q transpose q is equal to 1. So, this is equal to 1, which implies that this term is equal to 0, that is delta q transpose q is equal to 0. So, this is equation number 5, so let us now remember that we have got an important condition from equation number 4 and we have got another important condition from equation number 5.

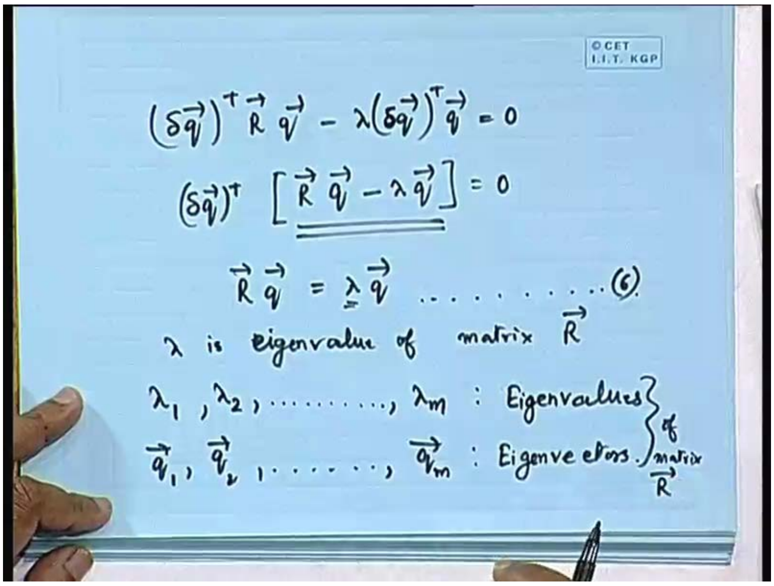
And what is the physical significance conclusion from equation number 5. It means to say that q and delta q they happen to be orthogonal. So, that means, to be say that only a change of direction of q is permitted, orthogonal means that you can only change the direction of the q. So, this part we have to accept and this one is already known to us this is equal to 0.

Now, since both of them are equal to 0 there must be some way of combining these two, but we find that there is some difference between this equation number 4 and equation number 5. And what is the difference here a term r is there whereas, here there is no such terms that is r. Now, q as such is a dimension less vector q's elements are not having any dimension, because they are unit vector. So, it is elements are not having any dimension.

So, the only dimension that is there is in this equation, whatever dimension we are getting is the dimension of the elements of R matrix. Whereas, in this case we do not have any such things, so if we want to combinedly express this equation 4 and 5 we have to introduce the term, which has got a dimensions same as that of R. So, we can decide to introduce some factor, some multiple or let us say some scalar quantity. Let us say that we introduce a scalar quantity lambda, which is of dimension R.

If we can introduce a scalar quantity lambda, which is having the same dimension as that of R, in that case this equation number 4 and equation number 5 can be combinedly written as this thing, this left hand side path plus or minus anything we can write. Because, lambda we chose as positive and negative depends upon that, let us say that we chose as minus lambda times this is equal to 0 we can write that. So, let us write that and see the results.

(Refer Slide Time: 41:34)



So, if we write that way then delta q transpose r q minus lambda, lambda having the same dimension as that of the elements of r matrix lambda delta q transpose q is equal to 0. Or equivalently now we can write down that this is one in the same as writing, we take delta q transpose outside and write it as R q minus lambda q is equal to 0. Now, certainly delta q transpose is not 0, it is a nonzero quantity, because we indeed wanted some perturbation.

So, what we have got from the extremal condition is that, necessarily this term has to be equal to 0. So, if we equate this term written with in the square bracket to 0, in that case what results is R q is equal to lambda q. Now, in this case you see dimensionally we do not have any problem q is of dimension m, then R is of dimension m by m. So, this results in a m dimensional vector, here lambda is a scalar, this q is an m dimensional vector. So, what results is simply equating this thing R q simply equating very identical quantities.

So, now we have got an equation like this and we call this as equation number 6. Where, there is a definite relationship that we are getting between this R here on an left hand side lambda on the right hand side. And this equation one should be able to solve, but it is not that this will be satisfied the solution can be obtained for any values of lambda and any values of q, given some R given a matrix R we have to solve for a solution of this lambda and q.

There, will be some combinations of lambda and q that should result in this. So, what is the significance of this lambda, it is called Eigen value, Eigen value of the...

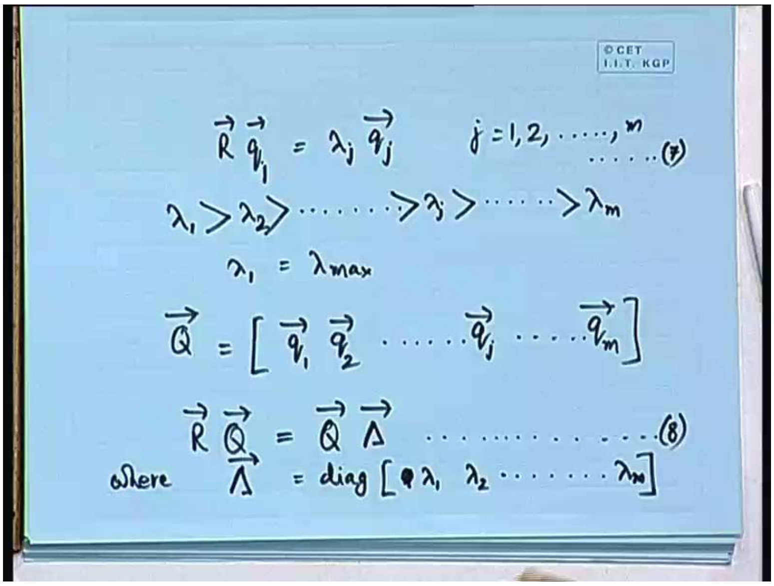
Student: ((Refer Time: 44:21))

Matrix r that is right. So, lambda is called as the Eigen value of the matrix R and R is of dimension m by m. So, how many such Eigen values will be there, there will be m such Eigen values. So, we can find out m such solution, so there will be lambda 1 and lambda 2 up to lambda m, there should be m such solution of Eigen values. And the associated values of cubes that we get, with those r given as q 1, q 2 up to q m these are the vectors, which are associated with it.

So, this lambda 1 to lambda m are the Eigen values and correspondingly all these q vector to which space you are projecting the vector x, you are calling those things as Eigen vectors. So, these are the Eigen vectors, so these are the Eigen values and Eigen vectors of the correlation matrix R. So, first a fall we have to obtain the correlation matrix; that means, to say we definitely know the statistics of this x vector.

So, now given that if we can design these quantities, in that case what results is that we will be projecting the x vector into a space, that results in a low variance quantities. Now, actually if speaking there are as we agreed, that there are m such solutions to equation number 6.

(Refer Slide Time: 46:40)



So, we can write down in the following way we can write down as R q j is equal to lambda j q j, for j equal to 1, 2 up to m. And let the corresponding Eigen values the arranged in the decreasing order. So, we place them in this order that lambda 1 greater than lambda 2 up to let us say here we have lambda j, the general term and them last we have the lambda m.

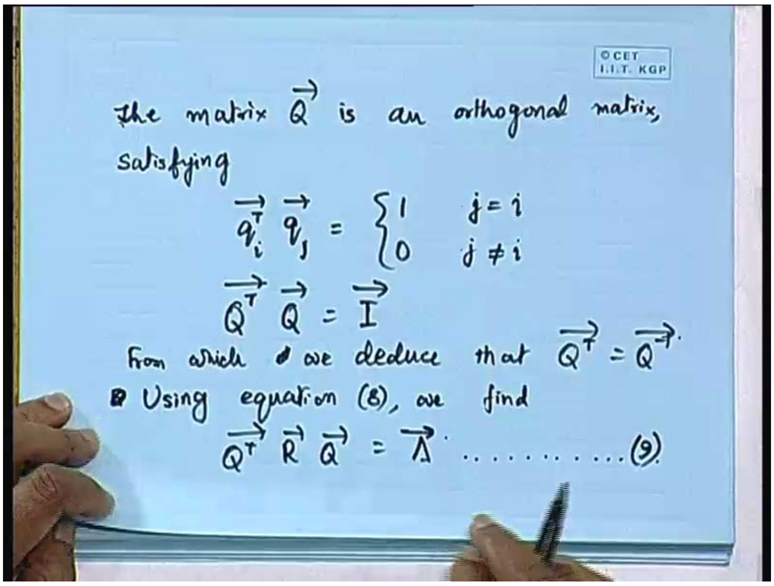
So, we have arranged them in a in decreasing order, so lambda 1 to lambda n. So, that lambda 1 is equal to lambda max. So, and then we use the Eigen vectors to constitute a matrix, so we now define a matrix capital Q, which is composed of the individual Eigen vectors. So, we have q 1 vector, q 2 vector, here q j vector lastly q m vector, so how this definition helps in that case, this one let us say that this is equation number 7 where, we have got m different such equations, now this m different such equations can be written in a more compact form by using this capital Q matrix representation. So, what results is R Q, if we write all this m equations in a combined way, in that case what results is R Q is equal to lambda. But, do we get lambda we get lambda 1 for the first one, we get lambda 2 for the second one, lambda 3 for the third one and combinedly what do we get, we get a lambda matrix and in which elements will those lambdas like...

Student: ((Refer Time: 48:57))

In the diagonal, so what we have here is this is equal to q lambda matrix we call this as equation number 8 where, this capital lambda matrix that we are writing is equal to the diagonal elements of this lambda 1, lambda 2 up to lambda m it is right. So, it is a diagonal matrix, so if we write down it in the form of capital lambda matrix consisting of these diagonal elements. Then, this is the equivalent matrix representation of all this m different equation which we heard of got otherwise.

So, now one thing that one should note down from this is that, what is the property of this q matrix. You see q matrix is composed of all these different q vectors and if we take the dot product if we try to project, one of these vectors into other, then what is it that we are going to get 0. So, this is actually an orthogonal matrix, so q becomes an orthogonal matrix and why it satisfies the condition of orthonormality.

(Refer Slide Time: 50:41)



So, we can say that the matrix q is an orthogonal matrix which satisfies. So, this is orthogonal matrix satisfying you take any column q i let us say you take the transpose of this. And just take the dot product of this with q j you will find that this is equal to 1, when j is equal to i and this is equal to 0 otherwise. So, far j naught equal to i, this is equal to 0.

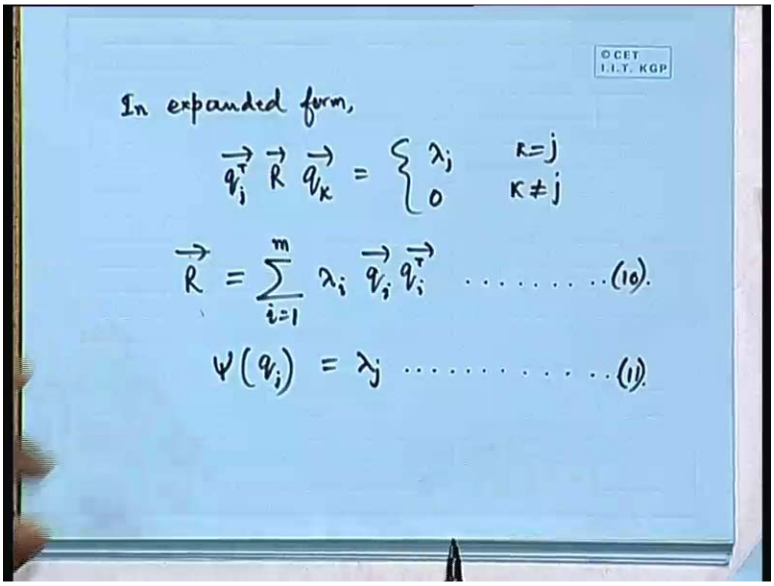
So, q is an orthogonal matrix or to say from the orthogonality property we can write down that Q transpose Q is equal to identity matrix. Because, this follows from this very basic property like, because it is orthogonal it results that q transpose q is equal to i. But, we know that q inverse q is also equal to i is not it Q Q inverse equal to i. So, what it mean is that...

Student: ((Refer Time: 52:05))

Q transposes is q inverse. So, from which we deduce that Q transpose is equal to Q inverse. So, that the equation that we have got the equation number 8 ((Refer Time: 52:34)) that we had obtained just see, look very carefully. If we now pre multiply the left hand side and the right hand side by q transpose. Then, what results on the left hand side Q transpose R Q and what results in the right hand side, Q transpose Q, Q transpose Q lambda up and Q transpose Q is equal to identity matrix.

So, what remains is the lambda. So, we can write, so from equation 8 using equation 8, 8 we find what Q transpose R Q is equal to the lambda matrix. So, this is the most compact matrix representation in equation number 9, which can be in the expanded form we can write it down as.

(Refer Slide Time: 53:45)



So, in expanded form the same thing that is Q transpose R Q is equal to lambda can be rewritten as q j transpose R matrix q k that is equal to lambda j for k equal to j and that is equal to 0 for k not equal to j, this is the expanded form of representation of equation number 9. And another thing that you should see is that this correlation matrix r itself could be expanded in terms of the Eigen values.

You just see, that you can pre multiply this by Q and you can post multiply that with Q transpose. If you do that what results is the R matrix on the left hand side and on the right hand side, what results is the Q transpose lambda Q lambda is a just a diagonal matrix consisting of the scalar elements. And then, we can represent the R matrix in the expanded form will be written as i is equal to summation i is equal to 1 to m lambda i q i q i transpose this we can write as equation number 10.

So, this is and one thing which you can see is that from equation number 1, that we had got from the equation number 1 ((Refer Time: 55:50)) that we had got that psi of q is equal to q transpose R q and q transpose R q is equal to what, it is equal to lambda j. So, also we find here that psi q j which are the variance probes, that is equal to lambda j. So, the variance probes that we have got are actually the Eigen values of the correlation matrix.

So, this is another important relation. So, this much for today's class we will continue with the mathematical analysis and the development of the theory in the next class.

Thank you.