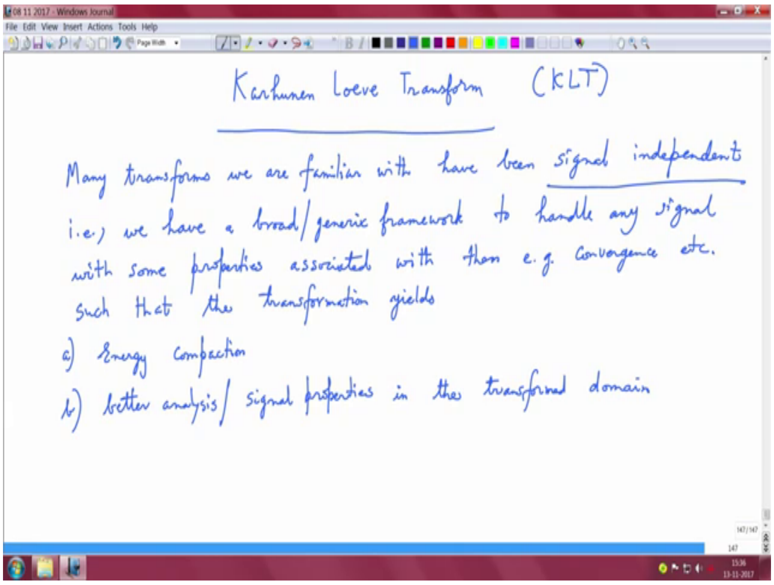
Mathematical Methods and Techniques in Signal Processing - I
Prof. Shayan Srinivasa Garani
Department of Electronic Systems Engineering
Indian Institute of Science, Bangalore

Lecture – 75
KL transform

So, let us get started with today's lecture this would be on KL transforms. So, you have seen at least two transformations as part of this course; one is the Fourier transforms and the other is the Wavelet transforms, you we have gone into the detail of course, your z transforms Laplace transforms and many other transforms that you have studied as part of your electrical sciences courses. So, in today's lecture we will see what is common between these transforms and then we will delve into the KL transform Karhunen Loeve transform.

(Refer Slide Time: 01:02)



So, if you see the broad structure, right many transforms that we are familiar with have been signal independent, this is a very important thing. Many transplants we are familiar with having signal independent that is we have a broad or generic framework to handle any signal with some properties associated with them, example convergence etcetera such that the transformation yields a, energy compaction b, perhaps better analysis or I would say signal properties in the transformed domain.

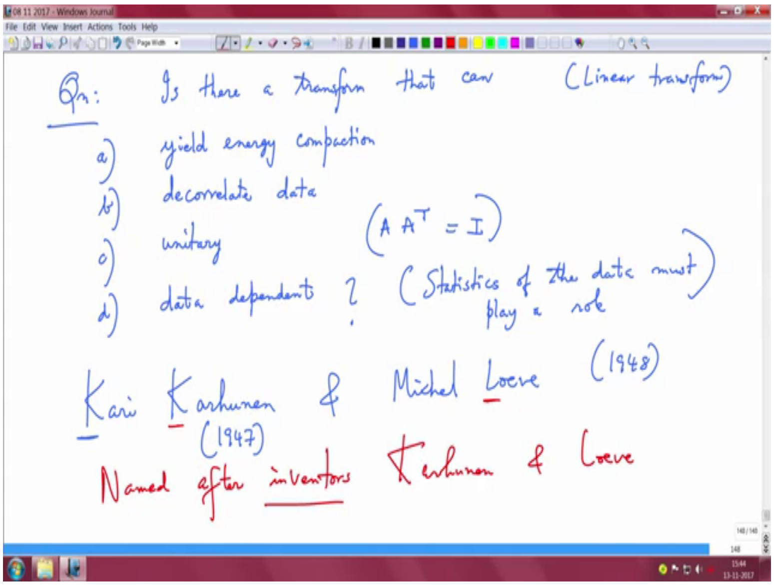
So, a good example of energy compaction these wavelet us right you look you think about wavelet us the compact energy and because of some compaction of energies which is happening that is you look at the original band and then you split it into two bands, right high pass and low pass and again perhaps within low pass you further decompose into low pass and high pass and so on. You do this level of decomposition as much as you want so that you can have energy in the in the topmost sub band, right and that basically I mean you have all the energy concentrated in the region in the sub band region that you want up to the level of resolution that you intend to right and that is basically energy compaction and that leads to automatically to compression.

And, another example that we saw was analyzing the sinking properties in the transform domain right. So, for example, if you look at wavelet us if I give you a 63 hertz versus perhaps a 65 hertz you would give a certain emphasis to 63 hertz versus 65 hertz though it is just 2 hertz separation they are build and compartmentalised in different sub bands alright there is some sort of unequal resolution that we can make in place alright and this gives us better analysis, it can also give a spike suppression etcetera etcetera.

So, therefore, signal properties can be better analyzed in the transform domain by using the frequency domain tools like the Fourier transform etcetera, right and we have seen such properties. Now, one of the questions that we have to ask is there a framework where the transformation depends upon the data right. So, this is sort of a philosophical question to ask. So, one is a regular transformation I give you any signal I can compute using Fourier that is a single independent as long as it satisfies certain notions of convergence, right example you know it is absolutely integrable or square integrable we have different notions we that we place for the signal right as long as it behaves is well behaved within those notions of conversions then we can compute the transformation and then we can analyze further, right.

Now, what is the main issue with data independent transform sources related data dependent transforms, right. If you think about data independent transform irrespective of a signal as long as that satisfies certain properties we can do energy compaction we can compress we can transform it into the equivalent domain figure out what it is required etcetera, but one of the questions that may come is if we were to maximally compress in some way in a signal processing framework it compression can also be thought about from an information theoretic framework right there are entropy based compression engines that is out of the scope of the current course, but if you think about from signal processing perspective statistics of the signal should play a role for certain if you were to look at certain properties of the signal in terms of compression etcetera and that is the reason why people conceived data dependent transform transforms such as the principal component analysis which is also the KL transform, ok.

(Refer Slide Time: 07:37)

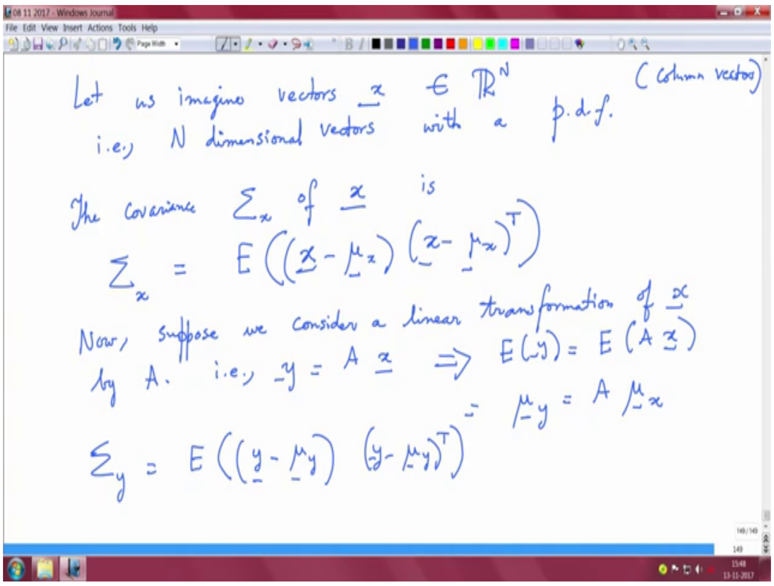


So, philosophically put. Is there a transform that can a yield energy compaction energy; that means, you compact the energy into a set of few coefficients that you want, ok. Next is decorrelate data correlations are good if you want to sense intelligence, right they are bringing an intelligent machine. If there are correlations or patterns that are correlated you can predict a good example is if you are gone through the course and you try to correlate with the professors scheme of examinations you can predict what can come if you cannot correlate then probably you donate you cannot expect what the type of questions can be right. So, this is I think is an important we need to decorrelate data we would like to have unitary transformations because it eases out certain properties in terms of simplifications if we can. So, therefore, we bring this notion of unitary unitariness which means if you are given a linear transformation A, AA transposes identity.

And, then this is very important it is data dependent; that means statistics of the data must play a role, ok. If you pause these problems, I need a transformation and I would make it more specific as a linear transformation because we like to play with linear operators, you need a linear transformation that can do all the following that is it should be able to do energy compaction, decorrelate data, maintain unitariness and then it should be data dependent. So, if you satisfied all these things then we have a transformation, right.

And, fortunately for us Kari Karhunen and Michel Loeve is a 1948 this is 1947 they developed this notion of this is for K and this is L Kari Karhunen Loeve transforms, named after the inventors. So, we will study the properties of KL transforms. I will study the properties that leads to KL transforms set up the problem in the framework of optimization that automatically leads us to the transform. So, I think it will be a two step approach in the lecture. So, first we will study some properties and then we know the answer to some extent and then we will post the right problem and then we will derive the transformation from first principles, and it has a lot of applications PCA is a is a very well known algorithm in machine learning and applications from search engines to many things you can do with PCA, ok.

(Refer Slide Time: 12:16)

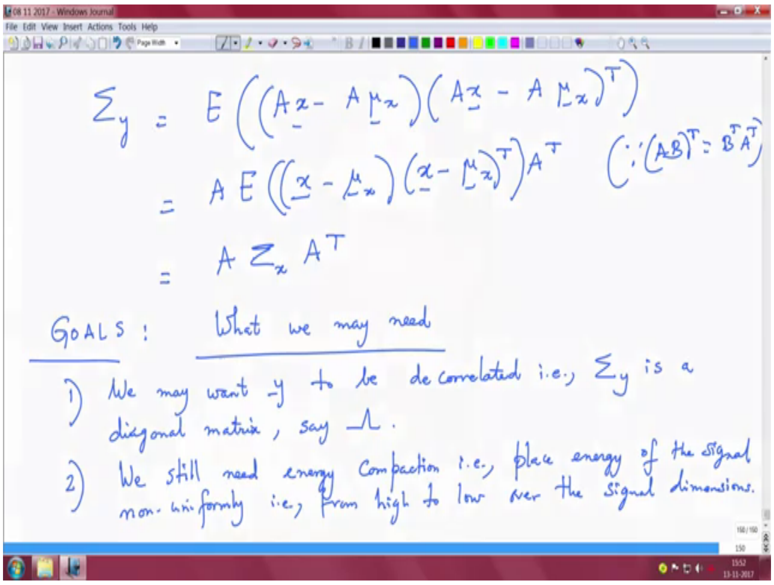


So, let us start with the following let us imagine vectors x belonging to R N right they are column vectors N dimensional vectors and they are associated with some probability density function, ok. So, to be precise they are column vectors.

Now, we can compute the covariance of x as follows. So, sigma x is the covariance matrix of x which is expectation of x minus mu x times x minus mu x transpose ok. So, this is a matrix right it is an N by 1 this is a 1 by N you have a N by N. So, this is a covariance matrix. Now, suppose we consider a linear transformation of x by A let A be some linear transformation of x which means y equals a times x right familiar transformation. So, this implies to take the expectation of y of this is expectation of A times x which is basically this is mu y we will designate it as mu y which is a times mu x right you can pull A outside because it is not stochastic expectation over the random vector right, when you take expectation should look at the multi domain multivariate distributions right, because it is a vector it of several coordinates and you should look at the multivariate density.

Now, sigma y is expectation of y it is a covariance you have to remove the mean right we have to remove the bias and then we have to compute the expectation y minus mu y times y minus mu y transposed this is what we need to compute.

(Refer Slide Time: 16:02)



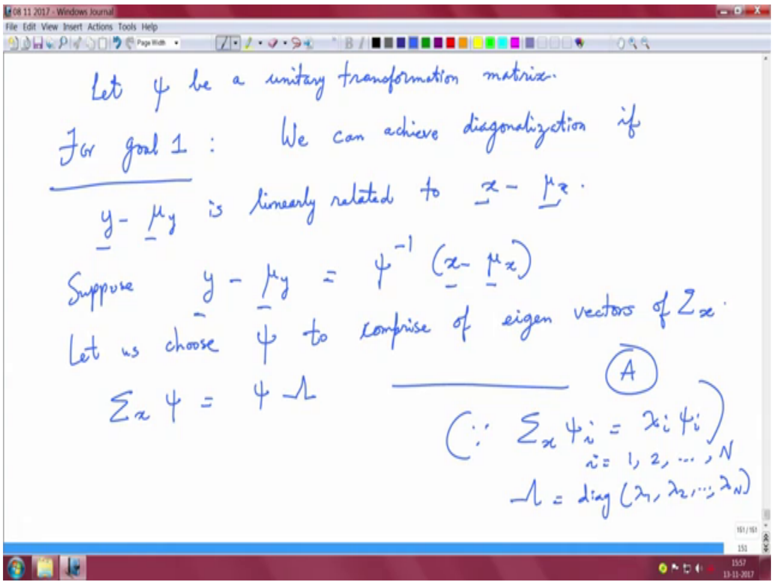
Now, if we plug this in y equals A x and we will simplify sigma y is expectation A x minus A times mu x times A x minus A times mu x transposed. Now, we can pull A outside here expectation x minus mu x times this is A into x minus mu x transpose the whole thing transposes A B whole transposes B transpose times A transpose, right and we can write this as x minus mu x transpose times A transpose ok, because just A B transpose is B transpose A transpose we apply this property.

Now, I think this brace must be here this is A times sigma x times A transpose this is how we would compute the covariance of y. So, we are slowly building what we need what our goal is, right. So, we start with the following goals. So, what we may need one they may want y to be decorrelated that is sigma y is a diagonal matrix say lambda that we would not decorrelate the data; that means, look at one coordinate it is not having any correlations with the other coordinate I mean you take expectation in the in the stochastic sense.

Now, second, second property is we still need energy compaction that is place energy of the signal non uniformly that is from high to low over these signal dimensions. So, what do we mean? Suppose, data is such that I have let us say it is a 2 dimensional data energy in x is 95 percent versus energy in y right if you take the some energy that has to be 100 percent, right if you look at all the coordinates if you look at the energy in each of the coordinates that should be the overall energy in the system, but some of the coordinates may have more energy than the rest and we need some way in which we can sort of have a gradation in the energy over the coordinates, that is sort of the I would say an idea.

Now, you get these two points, right we have these two goals now and we want a linear transformation right you can question that if you proceeded if you proceed with a non-linear assumption your optimization will be according to what you started off with, but often having a linear constraint helps us,. So, let us proceed further with this.

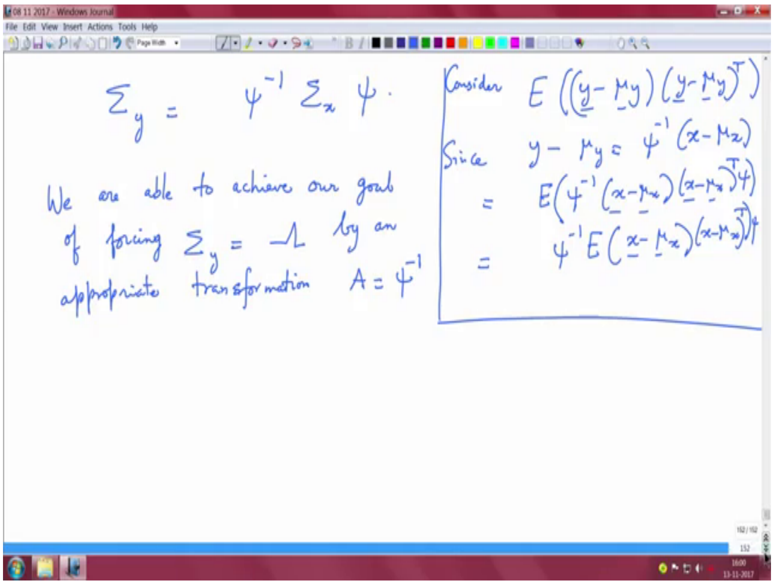
(Refer Slide Time: 21:04)



Let psi be a unitary transformation matrix for goal one we can achieve diagonalization if y minus mu y is linearly related to x minus mu x right. We can achieve diagonalization if y minus mu y is linearly related to x minus mu x. So, suppose y minus mu y is some psi inverse x minus mu x let us choose psi to comprise of eigenvectors of sigma x. So, it is in a way I have given you the answer for this linear transformation it is basically stacking all the Eigen vectors of the covariance matrix of x right, but we will arrive that this is indeed the right front right transformation to choose with the set of optimization constraints that we have,.

So, now if we assume this solution right then sigma x times psi is psi times lambda I call this equation A why because of course, we can say sigma x times some say psi i is one such vector right is basically lambda i times psi i this is your Eigen value equation for i equals 1 2 3. So, on N till N right you have N coordinates when Eigen vectors and Eigen values and this is the Eigen value equation and you stack all these lambda 1, lambda 2, lambda 3 becomes a diagonal matrix of lambda right lambda is basically diag I express lambda is diag lambda 1, lambda 2, lambda N, and then you can compute. Similarly, you stack the corresponding Eigen vectors and you can get this.

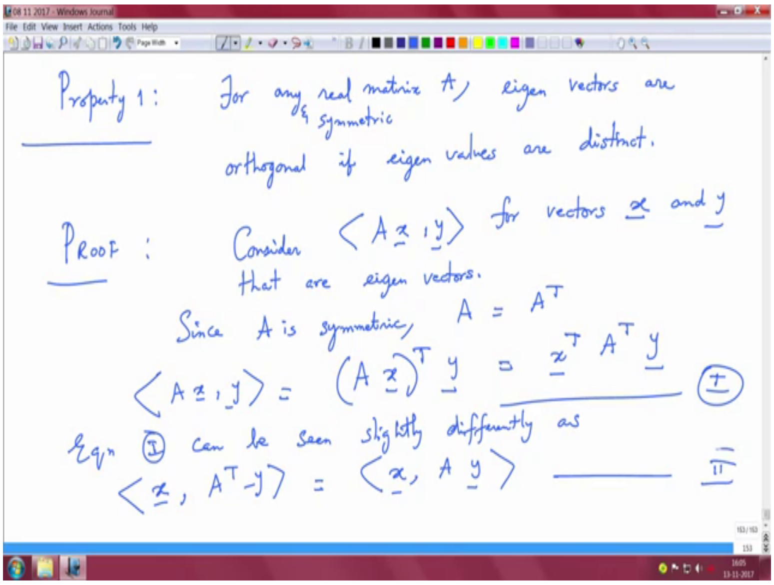
(Refer Slide Time: 24:32)



Now, sigma y is psi inverse sigma x times psi this is not too difficult to solve right I mean all you have to write is write it in terms of expectation of y minus mu y times y minus mu y transposed sort of vectors right. So, and you use this solution that y minus mu y is psi inverse of x minus mu x right, you plug this here you will land up with expectation of psi inverse x minus mu x x minus mu x transposed and then you have a psi here right expectation is over this is basically psi inverse expectation of x minus mu x x minus mu x transposed that is right. Put this just aside some steps right just follow the math you will you will land up with this,.

Now, we are able to achieve our goal of forcing sigma y to be lambda by an appropriate transformation A equals psi inverse right we can get a diagonal matrix I mean we can say that if sigma y is going to be lambda we can say the sigma y is going to be lambda if A is chosen to be psi inverse,. Now, let us we got a hint already that is you take the data x that is random vectors x, pull them up, compute their covariance matrix get the Eigen decomposition for the covariance matrix stack all the Eigen vectors and that forms your linear transformation, for mapping x to y and let us investigate the properties of y and we should be able to have d correlation etcetera, etcetera ok. So, let us prove some basic properties associated and then we will carefully delve into the rest.

(Refer Slide Time: 28:06)

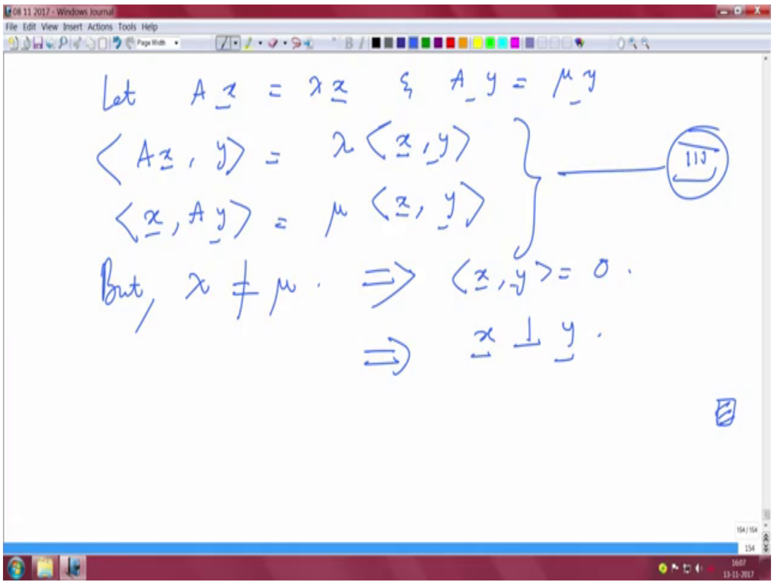


Property 1; for any real matrix A it is real and I also need to bring in the symmetric property real and symmetric matrix A Eigen vectors or orthogonal if Eigen values are distinct an interesting property right I give you a real symmetric matrix and I compute I do the Eigen decomposition right and Eigen vectors are orthogonal if Eigen values are distinct and why do you have to bring in this real symmetric matrix what is the connection think about the covariance matrix right expectation of x 1, x 2 is same as x 2, x 1.

So, if you start transposing it you will see symmetry properties there unless you are bringing any data dependent notions on the coordinates itself in some way that may not give you that kind of structure that you want right normal conditions expectation of x 1, x 2 same as expectation of x 2, x 1. So, let us prove this property and that is the link where we are reading this is more a general notion to state this property, but keep in mind that this matrix a you have imagine the covariance matrix because that is our goal that is how we are proceeding ok. So, let us start with the proof consider the inner product of x with y for vectors x and y that are eigenvectors. Now, since A is symmetric equals A transposed now, in product of A x with y can be written as A x transpose y right. So, dot product and this is going to be x transpose A transpose y let me call the equation-I.

Now, equation- I can be seen slightly differently right. So, you can see this as x you have put in these vectors here x dotted with A transpose y right. So, by our definition you transpose this and multiply with this right that is what we did here we took A x we transposed it you may multiplied it y and we got this right I can interpret it like this x with A transpose y, but this is symmetric matrix A transpose equal say I can write it as x with A y right. So, A x with y is same as x with A y if A is a symmetric matrix ok. So, let us call this equation II.

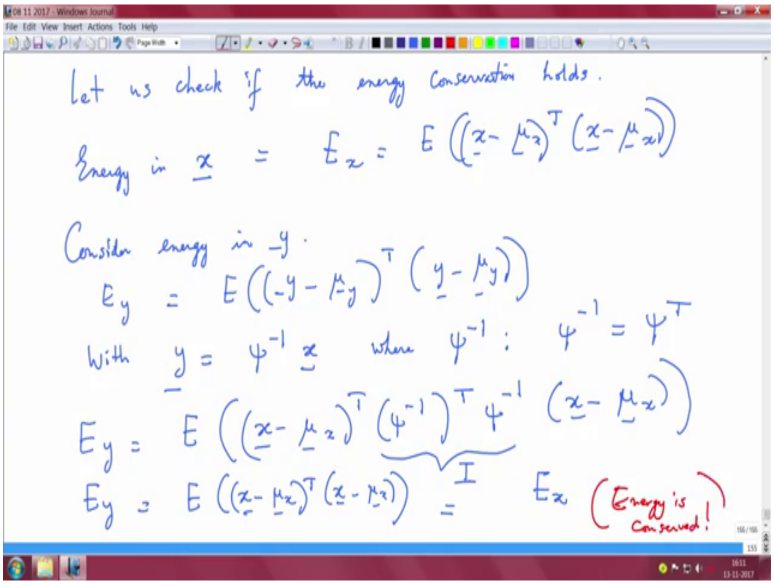
(Refer Slide Time: 32:34)



Let A x equals lambda x and A y is some mu y now A x with y is basically lambda because A x is lambda x right I use A property that I can pull the scalar out lambda inner product of x with y similarly I have x with A y this is mu times I pull the A y is basically lambda mu y. So, I pull the mu scalar outside this is again x with y right, I have set of equations-III.

Now, they both are equal right; that means, if I take the if I subtract these two equations inside 3 that has to be 0, but lambda is not equal to mu because we assume that they are distinct this implies inner product of x with y has to be 0 which means from geometric sense x is orthogonal to y first an interesting result it says that the Eigenvectors are orthogonal and you know what is the deeper implication of this result because they can form a basis right and that is what people think about it is Eigen basis from that perspective,.

(Refer Slide Time: 34:49)

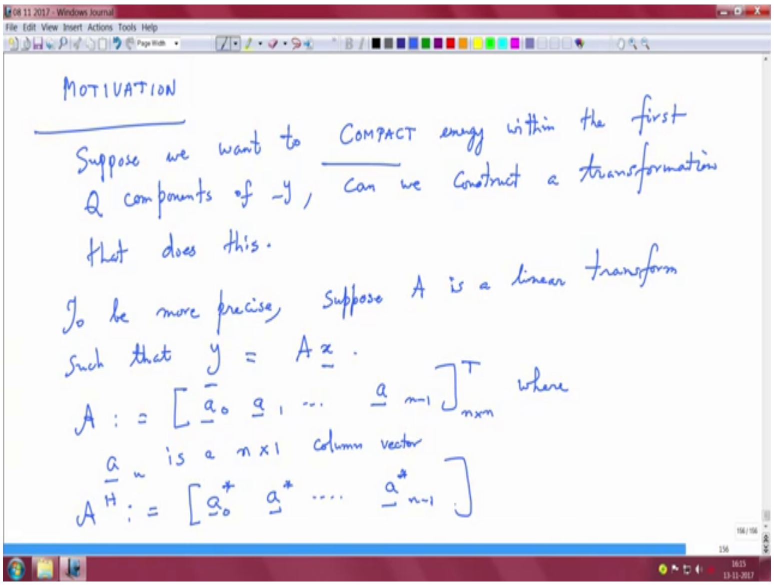


Now, let us check if the energy conservation holds this is very very important because if you cannot satisfy the energy conservation property then there is some loss in the transformation right and it be cannot recovered now the energy in x is.

Basically, the covariance which is given by expectation of x minus mu x transpose times x minus mu x. Now, let us consider energy in y right energy in y is E y which is expectation of y minus mu y transpose times y minus mu y right with y equals psi inverse x, where psi inverse is some matrix such that psi inverse equals psi transpose it says property satisfied. Now, E y is expectation of x minus mu x transpose psi inverse transpose psi inverse times x minus mu x and observe this because we want to bring in the unitariness right A A transpose is 1. So, your A is like psi inverse and here transpose is like your psi inverse transpose or your a is psi inverse and then you know a transpose is psi inverse transpose because of this is unitary this is identity and we have seen these unitariness in filter banks as well right.

So, therefore, E y is expectation of x minus mu x transpose times x minus mu x and this is basically E x and the implication of the statement is energy is conserved. So, this is a very important property that you have to bear in mind when you do these transformations, you do not want to lose the original energy in the signal in the transformation the energy has to be preserved, but you decide what components. You want to throw them off right you decide which of the how much energy you want to retain and depending upon your threshold you decide what you can do with the representation. So, this picture should be very clear in your mind.

(Refer Slide Time: 38:58)

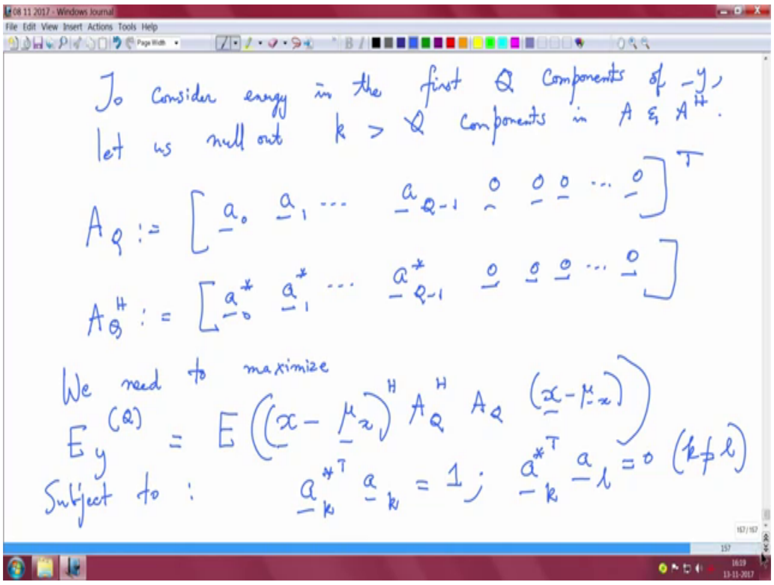


Now, with some of these properties that we saw they will motivate towards the derivation of the transformation. Suppose we want to compact this is very very important compact energy within the first Q components of y can we construct a transformation that does this; that means, I take some transformation of x and I get this vector y and I will observe the energy in y I want to compact it within the first Q components of y some transformation is happening such that say let us say 95. So, let us say the dimension is 10 the first 5 components has 95 percent energy and the rest has 5 say suppose, I want to realize such a kind of transformation can I do that with the given properties and you've already seen how statistics is playing a role here because they are bringing in the covariance matrices of the data. So, and therefore, we are bringing data dependent statistics right we have to apply other properties and motivate this problem.

Now, to be more precise suppose A is A linear transform such that y equals A x let us assume a such that I have a stack of all these vectors where a u is a n by 1 column vector say suppose. Now, let us deal with Hermitian because they can be complex valued. So, therefore, we can think about Hermitians now a Hermitian. So, instead of real symmetric matrix I could you know instead of real I can bring in perhaps complex and when I bring in complex then you have to be careful when you do the transposition your take conjugate transpose, right that is only subtle detail.

So, just to be mindful about. So, now, a Hermitian is this matrix which is a naught conjugate a 1 conjugate dot dot dot a n minus 1 conjugate ok.

(Refer Slide Time: 43:02)

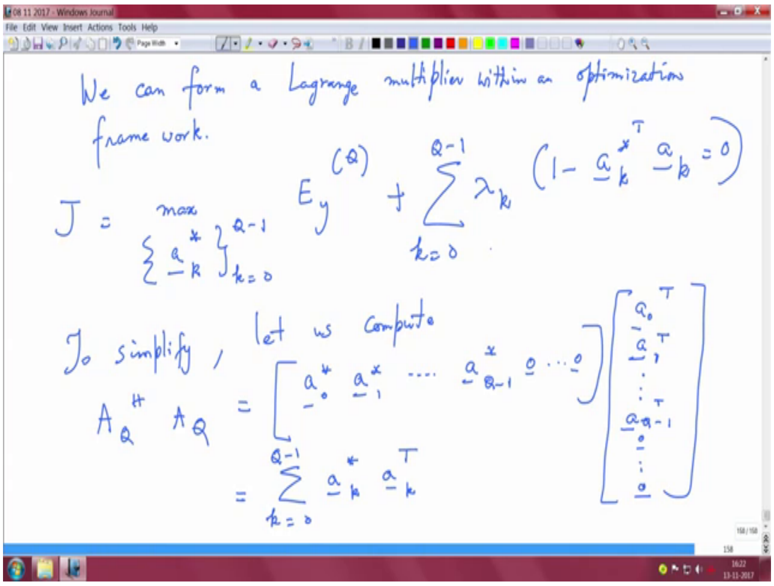


Now, to consider energy in the first Q components of y let us null out k greater than or equal to Q components say greater than Q because you want to retain the first Q and then anything post Q you pull null out components in a and a Hermitian, ok. Now, let us form A Q by reading the first Q components and this is a naught a one dot dot N dot A Q minus 1 0 0 0 transposed and A Q Hermitian is this matrix which is a naught conjugate a 1 conjugated dot dot dot A Q minus 1 conjugated null.

Now, what we need is we need to maximize the energy within the first Q components which means we want to look at the expectation of x minus mu x Hermitian with A Q Hermitian A Q times x minus mu x subject to the following conditions which is a k conjugate transpose a k is 1 it is ortho normal conditions that is the inner product between two Eigenvectors are not the same is 0 and when it is the same it is normalized to 1 right otherwise I would have had if I considered if I did null everything then I would have just a full a right in this if I would have I would have I would have had a full a here right.

Since, I am running out components k greater than Q right then I have I retain the first Q components and therefore, this is the energy in the first Q components and I want to figure out some transformations that can maximize the energy in the first Q components that is maximized E y of Q which is given by this quantity subject to these conditions subject to these conditions. So, now, it is not too difficult we can take and formulate these into the Lagrange multiplier framework.

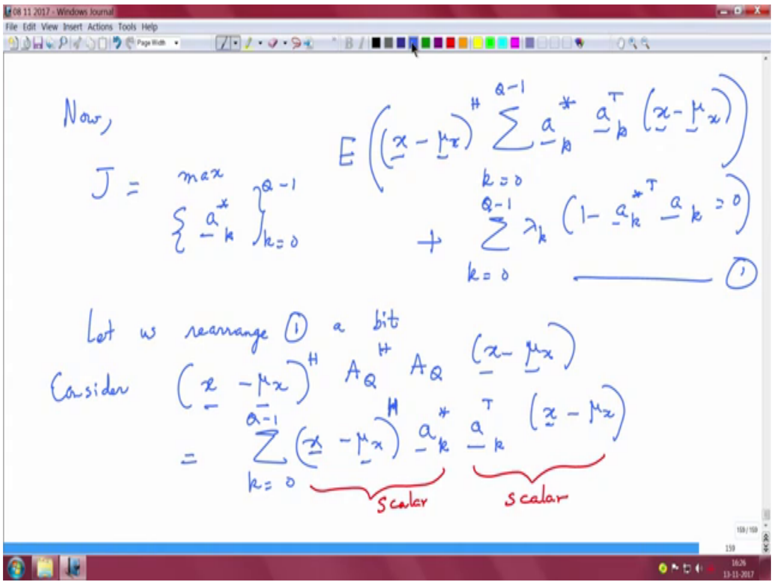
(Refer Slide Time: 47:09)



Let us see how we can do this now, J is maximized a k conjugate k equals 0 to Q minus 1 of E y of Q subject to the condition this is 0. Let us you bring in the condition that a k conjugate transpose a k equals 1 and that is the equation there is a associated parameter lambda k for the Lagrangian. So, you have all such equations, Q equations for k equals 0 1 2 3. So, on till Q minus 1 and you said that within the constraint set up.

So, now to simplify let us compute some terms and one of them is A Q Hermitian times A Q. So, this is basically a naught conjugate a 1 conjugate dot dot dot a Q minus 1 conjugate nulls times a naught transposed a 1 transposed dot dot dot a Q minus 1 transposed and then everything is null, ok. This can be compactly written in the form of a summation this is k equals 0 to Q minus 1 a k conjugate times a k transposed, ok. The moment you understand how to set this right the rest is routine algebra I mean you just have to get the right set up formulated.

(Refer Slide Time: 50:13)

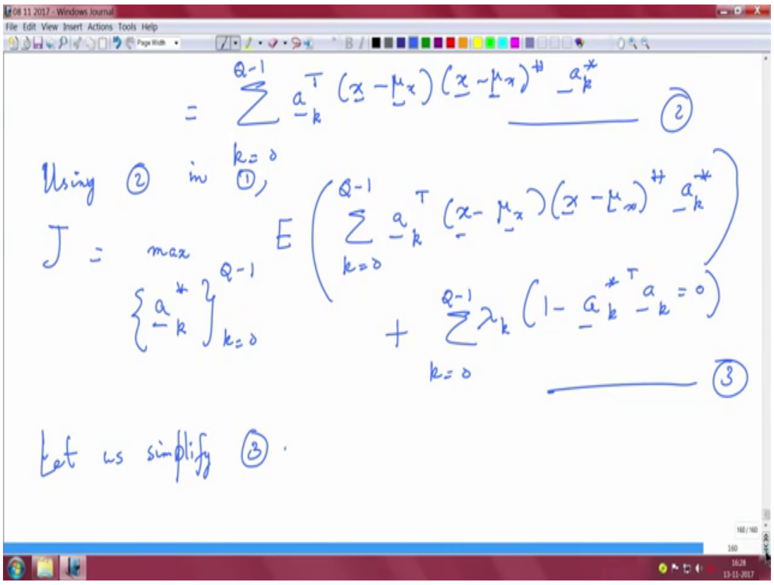


Now, J equals maximization where all a k conjugates k equals 0 to Q minus 1 expectation of x minus mu x Hermitian summation. So, a Q Hermitian times a Q can be written with the summation k equals 0 to Q minus 1 a k conjugate a k transpose times x minus mu x right and then subject this to the following constraints correct I did this looks more messy then what we started off of it, but of course, sometimes doing a lot of messy things will give you good results we can simplify things carefully it looks a very ugly equation, but you can simplify this.

Now, I call this equation – 1. Let us rearrange – 1 a bit let us rearrange this a little bit. So, to rearrange this we need to consider x minus mu x because the ugly term is one which is in this expectation right I consider x minus mu x Hermitian times A Q Hermitian A Q times x minus mu x and I said this could be written as summation k equals 0 to Q minus 1 x minus mu x times there is a Hermitian here a k conjugate times a k transposed times x minus mu x.

Now, this is very important to just recognize the fact that this is a scalar and this is a scalar and you are taking two scalars and you are multiplying them and you are basically adding them. So, therefore, if I can reverse the scalars; so what I can do is the following.

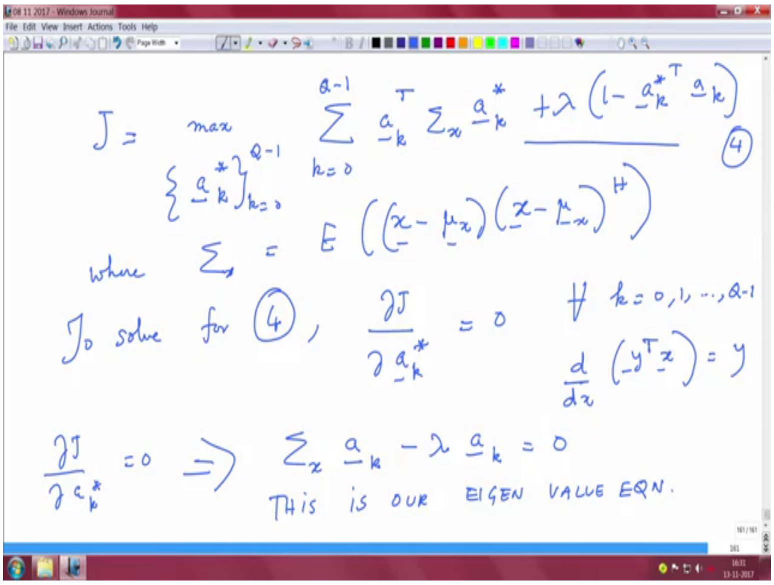
(Refer Slide Time: 53:19)



I can rewrite this as summation k equals 0 to Q minus 1 a k transpose x minus mu x times x minus mu x Hermitian times a k conjugate.

Say, if a routine algebra will not help you somewhere you should pause a little bit and think what if you can simplify this further and the simplification will help us because we can write J I would call this as I said this is earlier was equation one let us call this equation – 2. Using – 2 in – 1 we can write J as maximize overall a k conjugates k equals 0 to Q minus 1 expectation I have to bring this expectation out summation k equals 0 to Q minus 1 I am just dropping this directly drop into here a k transposed x minus mu x times x minus mu x Hermitian a k conjugate, ok. I want to maximize this subject to these constraints. It looks still messy, but we can simplify because you can pull this expectation inside expectation is a linear operator right that property we will apply.

(Refer Slide Time: 55:50)



Now, J let us simplify – 3. So, by doing so what we land up is now J can be written in the form maximize overall a k conjugates k equals 0 to Q minus one summation k equals 0 to Q minus 1 a k transpose expectation is pulling inside. So, the expectation of x minus mu x Hermitian x minus that is become sigma x covariance matrix times a k conjugate subject to these constraints. This is now easy for us right, where sigma x is expectation of x minus mu x times x minus mu x Hermitian right this is straightforward.

Now, what we do is I have called this 4 now to solve for equation – 4. So, you want to maximize something subject to this you set the partial derivative of this cost j with respect to a k conjugate to 0 for all k equals 0 one dot dot dot Q minus 1 right, you do this and you apply the properties of the vector differentiation that if you have two vectors y and x d by dx y transpose x is basically y. Now, using this property dJ by d a k conjugate equal 0 implies sigma x with a k minus lambda a k equals 0 and this is our Eigen these are Eigen value equation.

So, all we did is we post the solution to the linear transformation in the framework of optimization subject to certain constraints. This is basically a constraint optimization problem and then we land up with the Eigen value equation. So, before we started off with the optimization we never assumed that they are they are Eigen they are they are 0 or that they are Eigen vectors etcetera that satisfies this equation we assume that they wanted we wanted them to be orthogonal we wanted. So, that it forms an easy way to represent in the form of a basis if we brought in the orthonormality constraints and subjected the energy to be maximized over the first Q components and we did this we land up with the Eigen value equation. So, it is to choose Eigen vectors of the covariance of x and then work out backwards to get to our transformation.

So, I really do not know how Karhunen and Loeve thought about in their original thought process of thinking through this problem did they have an intuition that they have to look into the Eigen basis or they just posted it in an algebraic framework and solved this I do not know how they had they had, but I think remarkably there should have been some clear intuition behind going about in a very structured way to solve this problem, right.

So, this concludes the KL transformation we will now see the application of KL transformation to dimensionality reduction, ok. So, we will stop here.