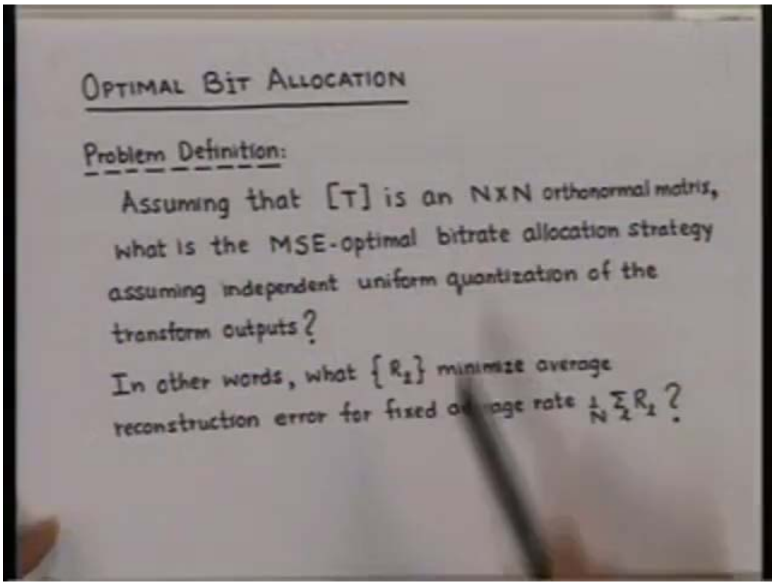
Information Theory and Coding
Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 41
Transform Coding Part – II

In today's class, we will continue our study of transform coding. Transform coding is a block quantization approach, it exploits the statistical correlation between the samples of the source output vector. The goal of the transformation is to obtain the uncorrelated transformed coefficients and compact the energy into a few of this coefficients. We have also seen that the mean squared reconstruction error is equal to mean squared quantization errors.

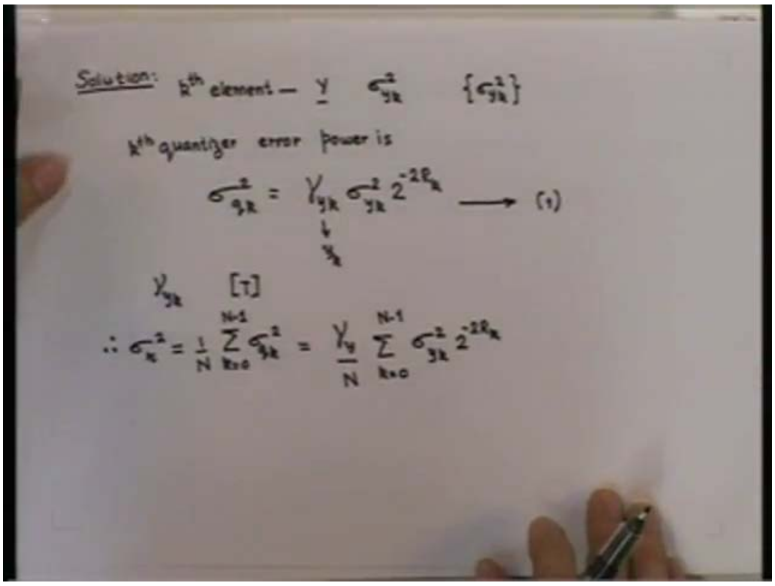
Now, because these transformed coefficients are uncorrelated it is reasonable to quantize these coefficients independently according to their individual statistical characteristics. Now, the next logical question is that if we are given total number of bits then how do we allocate the bits to this various quantizers, so that the mean squared reconstruction error is minimized with the constraint of the given fixed average bit rate.

(Refer Slide Time: 02:45)



So, this is optimal bit allocation problem and the definition would be as follows. Assuming that T is an N by N orthonormal matrix, what is the mean squared error optimal bit rate allocation strategy assuming independent uniform quantization of the transform outputs. In other words what is that set of rate for the quantizers, which will minimize average reconstruction error for fixed average rate.

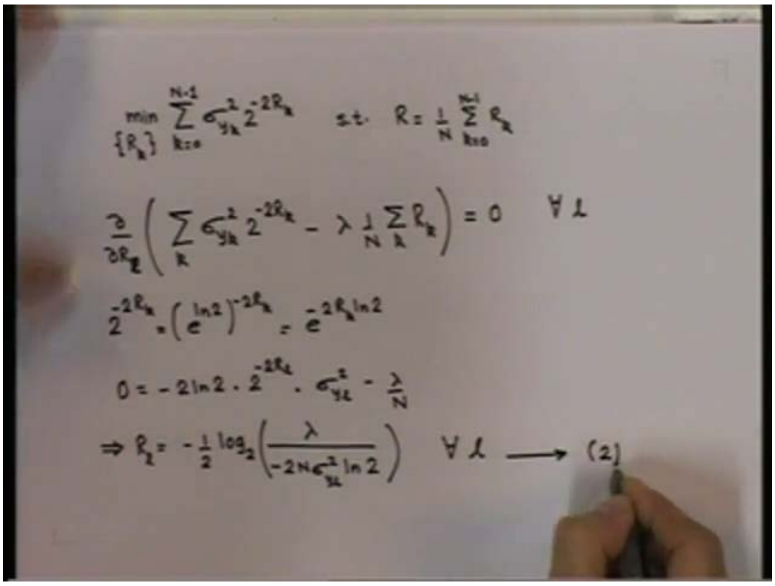
(Refer Slide Time: 03:17)



Now, the solution to this let us say that the k-th element of the element of the transform output vector y has variance sigma squared y k. So, we will add set of variances for the different transform coefficients. Now, we have shown that with uniform quantization, the k-th quantizer error power is given by the following expression. This we proved it in the class where R k is the bit rate allocated to the k-th quantizer output, and where gamma y k depends on the distribution of y k that is the k-th transform coefficient in the output vector y.

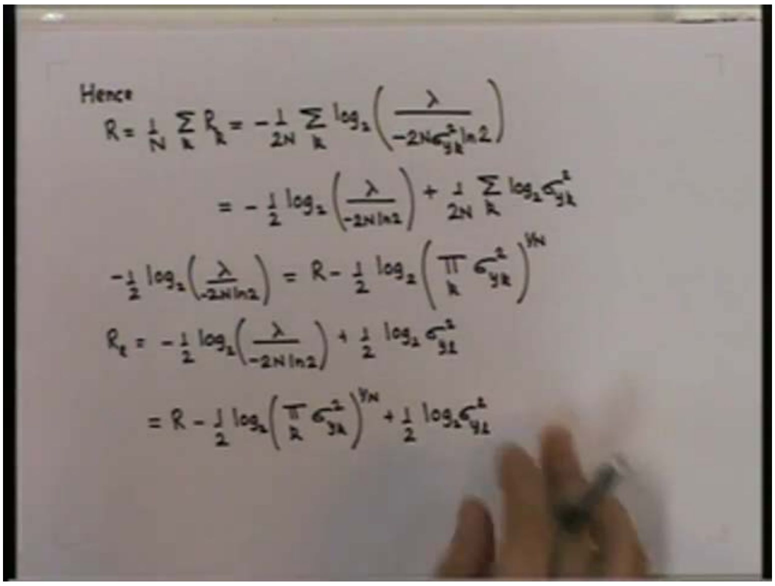
Now, we make the simplifying assumption that gamma y k is independent of k, now because we are using orthonormal transformation T, this implies that the mean squared reconstruction error equals mean square quantization error. Therefore, we can write mean square reconstruction error as 1 by N summation of mean quantization errors from different quantizers. This can be rewritten using this expression as y by N where we have assumed that gamma y k is independent of k and is equal to gamma y.

(Refer Slide Time: 06:26)



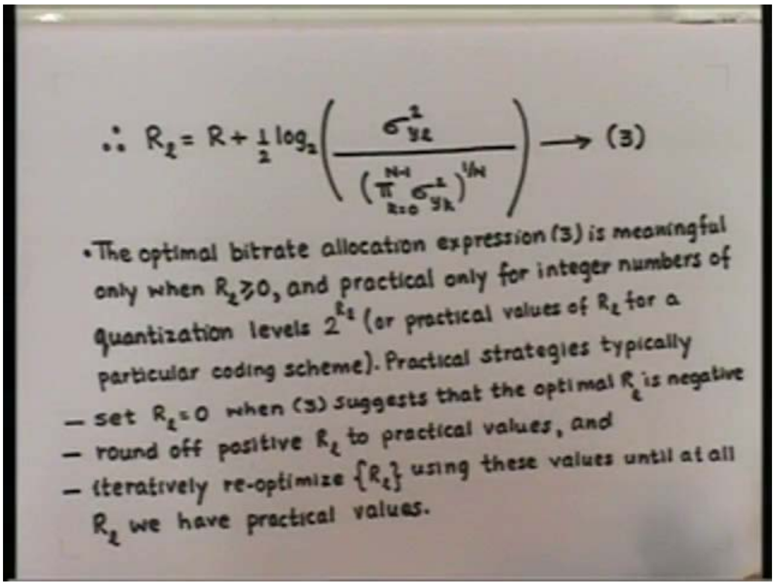
Therefore, we have the constraint optimization problems as minimize this quantity over the set of R k such that we have this constant satisfied. Now, using the Lagrange technique, we first set differentiation with respect to r l of the following quantity equal to 0 for all l. Now, we can write 2 raise to minus 2 R k as e raise to log 2 minus 2 R k equal to e raise to minus 2 R k log 2. Using this relationship, the zero derivative implies R l is equal to minus half log to the base 2 of the quantity lambda divided by minus 2 n sigma squared by l log 2 for all l, this is equation number 2.

(Refer Slide Time: 09:16)



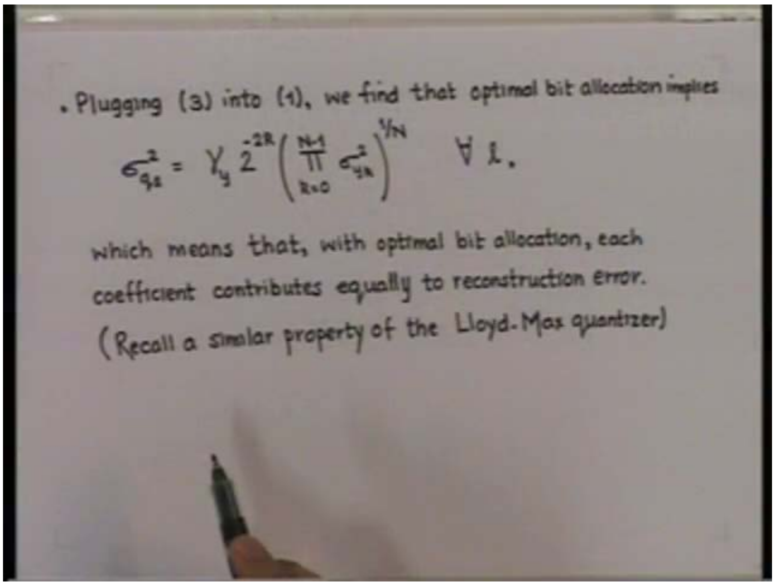
Hence, R equal to 1 by N sigma k, R k can be rewritten as minus 1 by 2 N summation over k of the quantity log 2 lambda 2 N sigma squared y k log 2 and this can be rewritten as minus half log 2 lambda minus 2 N log 2 plus 1 by 2 n summation k log to the base 2 sigma squared y k. So, from this we get this quantity that is minus half log 2 lambda minus 2 m log 2 equal to R minus half log to the base 2 product of variances. Now, rewriting two we get R l is equal to minus half log 2 lambda by 2 N log 2 plus half log 2 sigma squared y n. Now, plugging in this expression into this, we get R l equal to R minus 1 by half log 2 1 by n plus half 2 sigma square y l and this can be simplified as shown here.

(Refer Slide Time: 12:09)



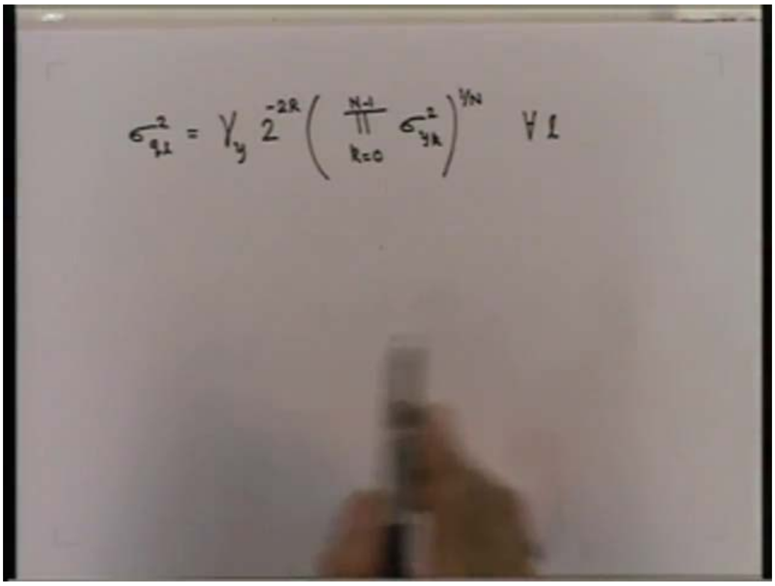
So, this is the final expression which we get for R l, so the optimal bit rate allocation expression is meaningful when R l is greater than equal to 0 and practical for integer numbers of quantization level or practical values of R l for a particular coding scheme. So, practical strategies typically set R l equal to 0 when three suggests that the optimal R l is negative, round off positive R l to practical values and iteratively re-optimize the set of R l using this values until at all R l we have practical values.

(Refer Slide Time: 13:23)



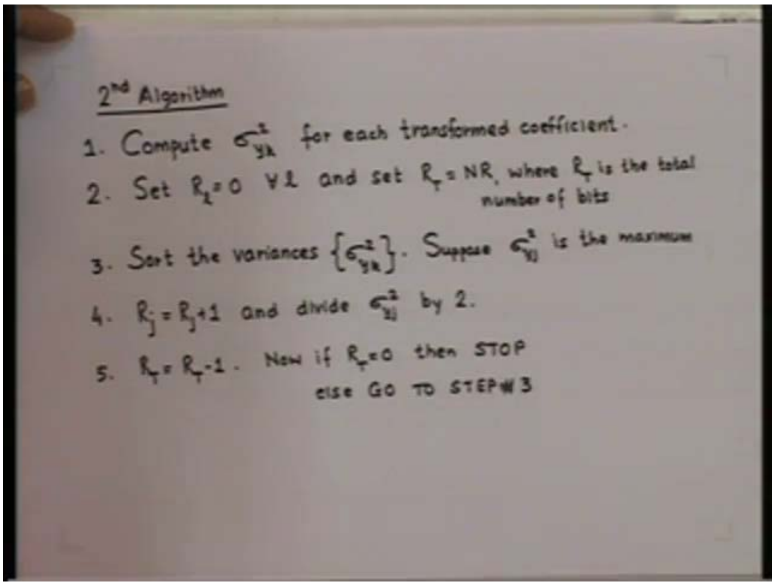
Now, plugging this equation three into equation one, we find that the optimal bit allocation is given by now plugging equation number three into one, we find that optimal bit allocation implies sigma square q l equal to gamma y 2 raise to minus 2 R. Now, what this equation means is that with optimum bit allocation, each coefficient contributes equally to reconstruction error.

(Refer Slide Time: 13:44)



Now, recall a similar property of the Lloyd-Max quantizer, now another algorithm that uses estimates of the variances in a recursive algorithm is as follows.

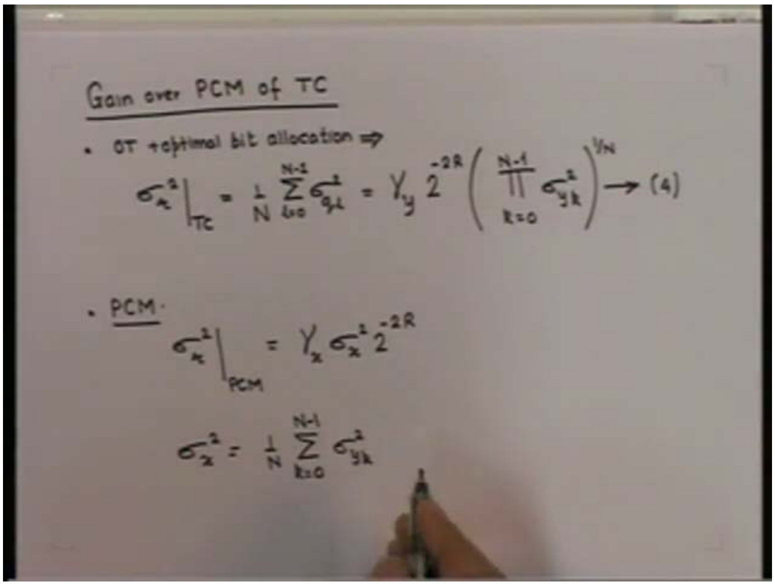
(Refer Slide Time: 14:40)



First we compute the variances for all the transform coefficient next step, we set all R l equal to 0 and set R t equal to n times R, where R t is the total number of bits available for allocation among all the n quantizers. Next step is to sort the variances, suppose the variance for the j th transform coefficient is the maximum, then the next step is to increment R j by 1 and divide the variance by 2 and next step is to decrement R t by 1. Now, if R t is equal to 0, then we stop else we go to step number 3.

Now, at the end of this algorithm, the transform coefficients with higher variances will be allocated more bits. Now, having discussed the optimal bit allocation, let us discuss the performance transform coding in comparison with pulse code modulation. So, let us find out how much gain we get over pulse code modulation gain over pulse code modulation of transform coding.

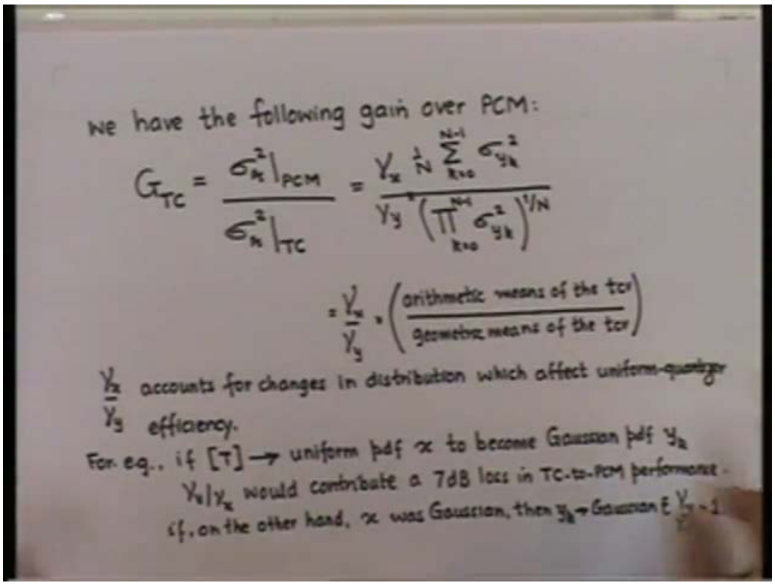
(Refer Slide Time: 18:12)



Now, with an ortho normal transform and the optimal bit allocation as given by 3, the total reconstruction error equals. So, this means total reconstruction error for transform coding is equal to 1 by N times summation sigma square q l, l is equal to 0 to N minus 1 and this is equal to gamma y and 1 raise to minus 1 R times n minus 1 k equal to 0 sigma squared y k.

Now, compared to uniformly quantized P C M, we have the quantization error equal to reconstruction error. P C M is equal to gamma x sigma x squared 2 raise to minus 2 R, where we assuming the same fixed average bit rate. Now, since we are using an orthonormal transformation, this implies that sigma x squared is equal to 1 by N times summation for k equal to 0 to n minus 1 of sigma squared y k.

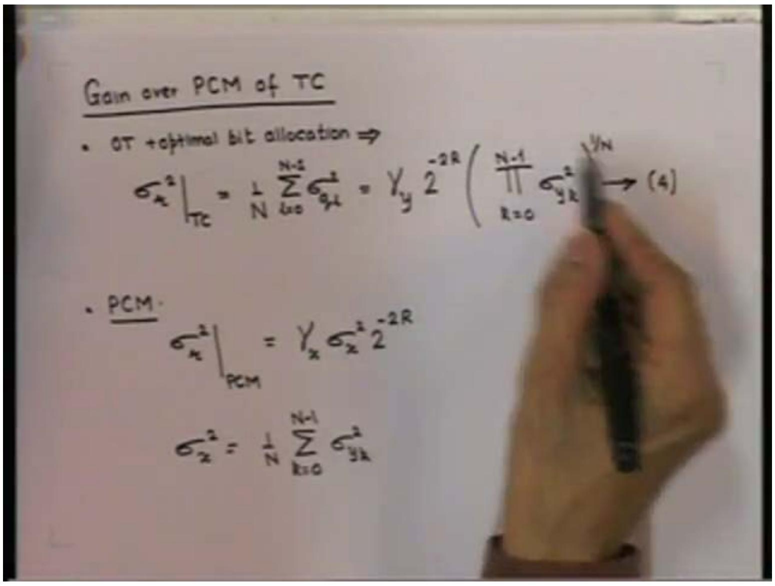
(Refer Slide Time: 20:53)



So, using this we have the following gain over P C M, so gain of transform coding over P C M is given by this expression is equal to this, and this can be rewritten as gamma x by gamma y times arithmetic means of the variances of the transform coefficients. This is the geometric means of the variances of the transform coefficient gamma x by gamma y accounts for changes in distribution, which affect uniform quantizer efficiency. For example, if the choice of the transform T converts uniform p d f x to become Gaussian p d f y k, then this factor gamma y by gamma x would contribute a 7 dB loss in transform coding to P C M performance.

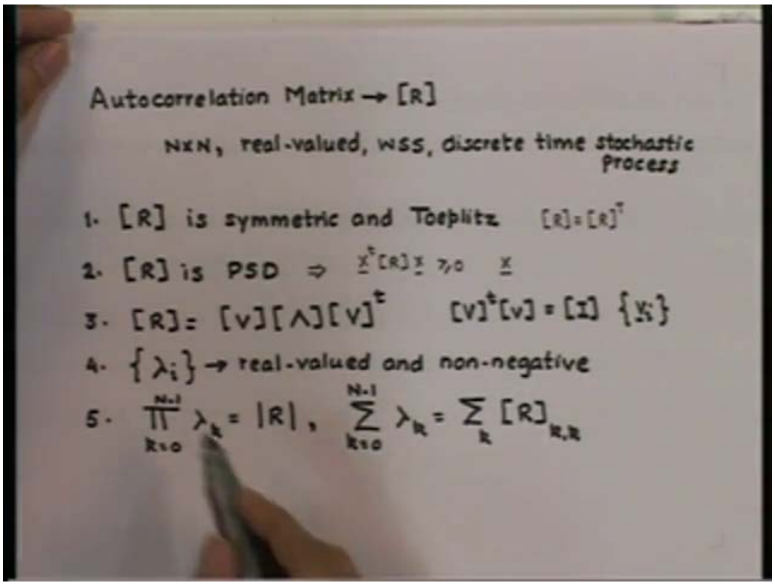
Now, we can show this it is not very difficult, but we will not do this at the moment, now if on the other hand if x was Gaussian then y k would be also Gaussian, and in this case gamma y by gamma x would be equal to 1. Now, in literature this term out here is also known as transform coding gain and it is denoted as G t c this is transform coding gain. This transform coding gain is a measure of the amount of energy compaction afforded by a particular transform, and deficiency of a transform depends on how much energy compaction is provided by the transform. So, indirectly G t c provides a measure of the efficiency of a transform.

(Refer Slide Time: 23:25)



Now, coming back to this expression, if we ignore the effect of transform choice on uniform quantizer efficiency is gamma y. Then this expression four suggests that transform coding reconstruction error can be minimized by choosing the orthogonal transform T that minimizes the product of coefficient variances. It is to be remembered that orthogonal transform preserve the arithmetic average of coefficient variances. So, the next question is what the optimal orthogonal transform which will minimize this quantity. So, before we try to answer that let us discuss some of the properties of an autocorrelation matrix.

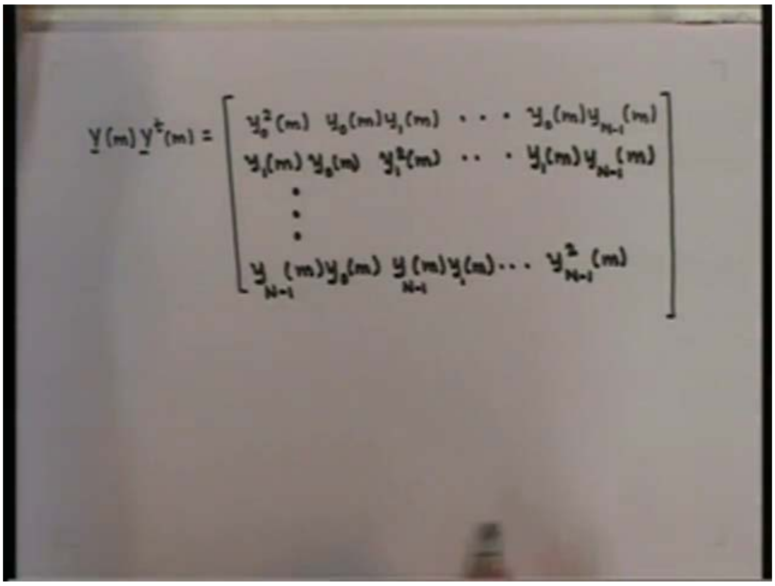
(Refer Slide Time: 24:27)



See that R is an N by N autocorrelation matrix of a real valued wide sense stationary discrete time stochastic process. Now, the following properties are associates which are useful with this autocorrelation matrix autocorrelation matrix denoted by R is symmetric and Toeplitz, what it means R is equal to R transpose while Toeplitz has equal elements on all diagonals. The next property is that R is positive semi definite what this implies is that x transpose R times x is always greater than equal to 0 for any real valued x.

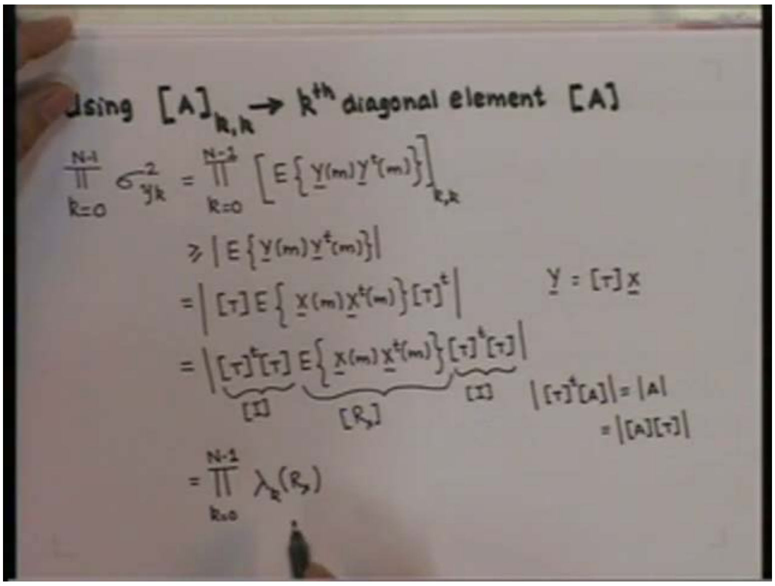
The next property is that this autocorrelation matrix has an Eigen decomposition, it means that R can be written in this matrix form where this matrix V is a orthonormal matrix which implies V transpose V is equal to I and columns of this matrix are Eigen vectors V i of R. So, V can be written as and this is a diagonal matrix whose elements are the Eigen values lambda i of R. So, matrix is equal to diagonal lambda 0 lambda 1 up to lambda N minus 1. The Eigen values of R are real valued and non negative and finally, the product of the Eigen values is equal to the determinant of the matrix, and the sum of the Eigen values is equal to the trace of the matrix.

(Refer Slide Time: 27:59)



This denotes the diagonal element of the matrix R. Now, let us define the outer product as follows y 1 m, y 0 m, y 1 squared m. So, this is a dot product of the vector y with itself.

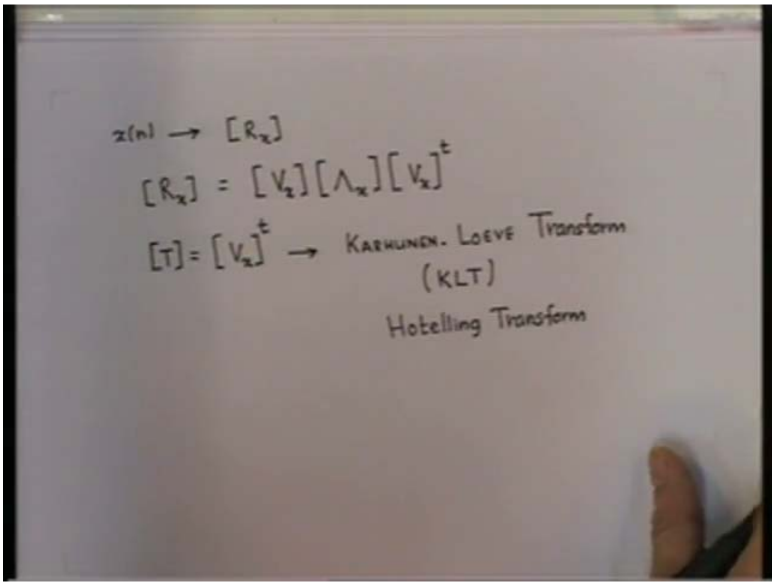
(Refer Slide Time: 29:23)



Now, using the notation to denote the k th diagonal element of a matrix A, we can write equals, so the product of the variances is equal to the product of the diagonal elements of this output autocorrelation matrix. Now, this is always greater than equal to the determinant of the matrix, now this can be rewritten as because Y is equal to T transform on k Y is equal to transform T of rating on the vector X. So, this is equal to where we are using the fact that determinant of orthogonal transform transpose operating on matrix A is equal to the determinant of the matrix A is also equal to the determinant of the matrix obtained by multiplying A and orthonormal transform T.

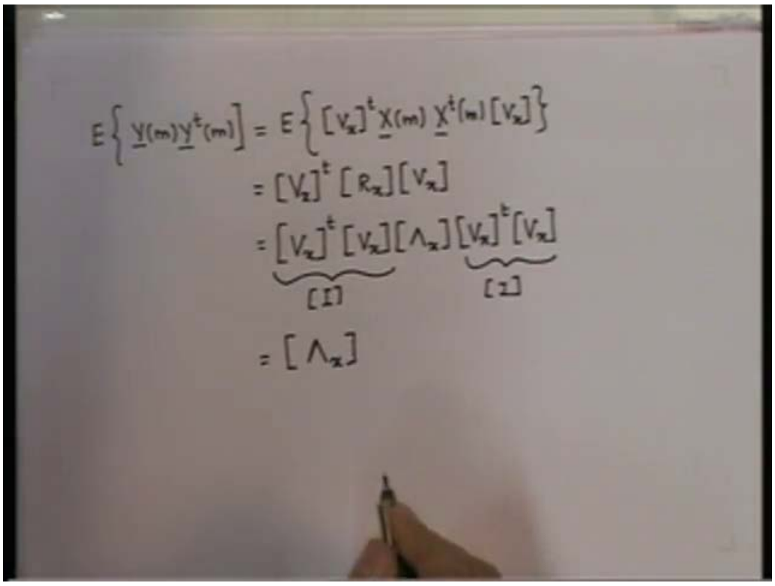
So, based on this relationship we have gone from this step to this step, now this is equal to I because T is an orthonormal matrix, this is equal to the input autocorrelation matrix and this is equal to again I. So, this implies that this is equal to N minus 1 k equal to 0 Eigen values of the autocorrelation matrix are x. So, thus minimization of this quantity would occur if equality could be stable at this point.

(Refer Slide Time: 34:10)



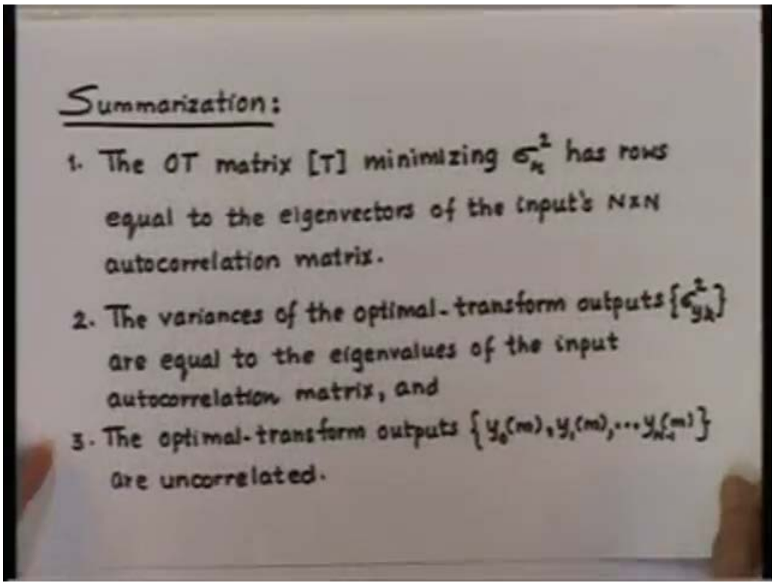
So, say that the Eigen decomposition of the autocorrelation matrix of x n which we denote by R x is R x is equal to V x, for orthonormal Eigen vector matrix V x and diagonal Eigen value matrix shown here. Now, if we choose the transformation t equal to V x transform, this will result in the desired properties that equality would be stable at this point and this transform is known as Karhunen-Loeve transform or in short K L T is also known sometime Hotelling transform or principal component analysis. Now, if we choose this then the desired properties are satisfied and this can be easily verified as follows.

(Refer Slide Time: 36:00)



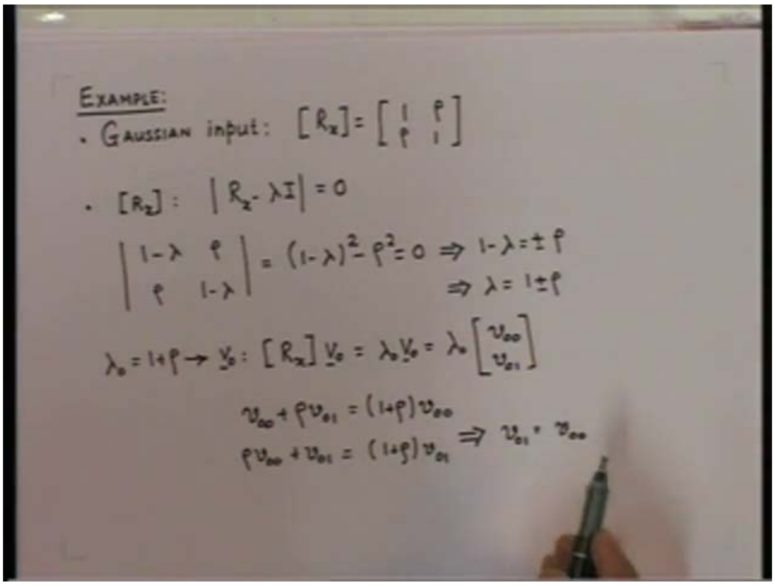
E is equal to expectation of V x transpose times X m X transpose m V x and this is equal to V x transpose R x V x. This can be rewritten as V x transpose R x can be decomposed as V x times the diagonal matrix composed of Eigen values V x transpose V x. Now, this is equal to I this is equal to I and therefore, this is equal to.

(Refer Slide Time: 37:23)



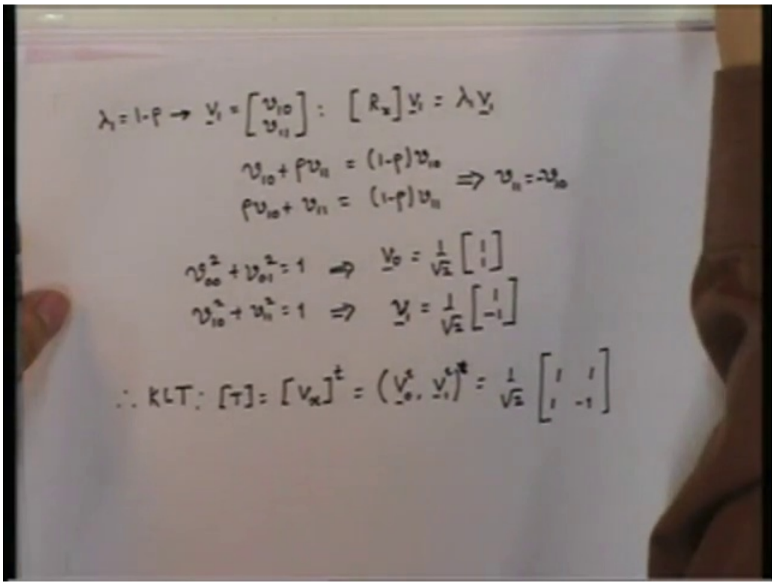
So, in summarization the orthogonal transform matrix T, minimizing the mean square reconstruction error has rose equal to the Eigen vectors of the inputs N by N autocorrelation matrix. The variances of the optimal transform outputs are equal to the Eigen values of the input autocorrelation matrix and the optimal transform outputs are uncorrelated, this is a very important result. So, the presence of mutually uncorrelated transform coefficients supports our approach of quantizing each transform output independently of the others. Now, if we have the source output vectors as Gaussian then the output will also be Gaussian and K L T will decorrelate these components and for Gaussian case, decorrelation also implies independence. So, let us take a simple example and obtain the K L T transform for a particular autocorrelation matrix.

(Refer Slide Time: 38:55)



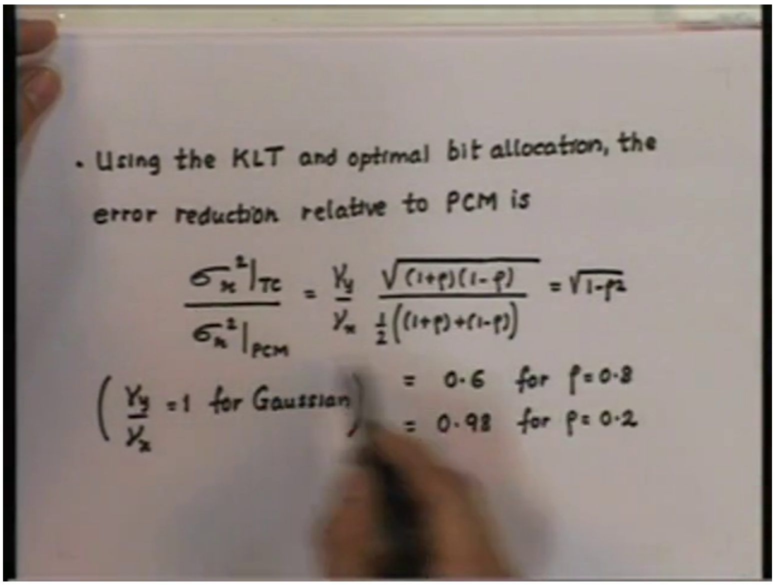
So, let us assume that we have Gaussian input with the autocorrelation matrix given as follows where rho indicates the correlation coefficient between the adjacent samples. Now, we can obtain the Eigen values of R x by solving the characteristic equation R x minus lambda I determinant of this equal to 0. So, if we do that, we get 1 minus lambda rho, 1 minus lambda equal to 1 minus squared minus rho squared equal to 0 which implies 1 minus lambda equal to plus minus rho which implies lambda equal to 1 plus minus rho. Therefore, there are two Eigen values, 1 plus rho and 1 minus rho, so for the Eigen value given by 1 plus rho, we can obtain the corresponding Eigen vectors V 0 as follows is equal to. This implies V 0 1 is equal to V 0 0. So, both the components of the Eigen vectors V 0 are equal.

(Refer Slide Time: 41:14)



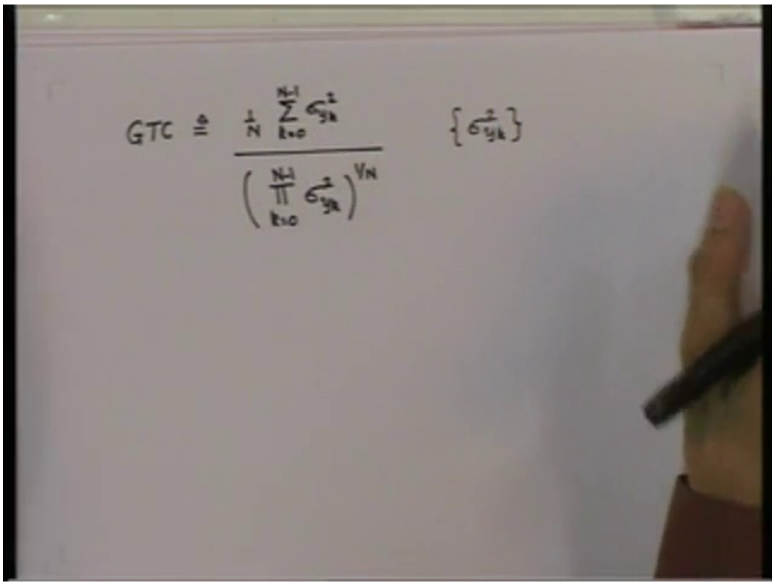
So, similarly for the Eigen value, lambda 1 equal to 1 minus rho, we can obtain the Eigen vector V 1 equal to V 1 0, V 1 1 and as the solution of the following equation. This implies V 1 0 plus rho V 1 1 is equal to 1 minus rho V 1 0 rho times 0 plus V 1 1 equal to 1 minus rho times V 1 1 and this implies V 1 1 is equal to minus V 1 0. Now, we also require that the K L T transform matrix should be orthonormal, so which implies equal to 1 and V 1 0 square plus V 1 1 squared is equal to 1. This implies V 0 is equal to 1 by root 2, 1 1 and this implies V 1 vector is equal to 1 by root 2, 1 minus 1. Therefore, the K L T transform for this case is equal to V x transpose is equal to V 0 V 1 is equal to...

(Refer Slide Time: 43:26)



So, using the K L T and optimum bit allocation, the error related to the P C M can be found out for this case is given by this expression here. We are assuming that gamma y by gamma x is equal to 1 because our input is Gaussian, so the value at this is equal to 0.6 for rho equal 0.8 and this is equal to 0.98 for rho equal to 0.2.

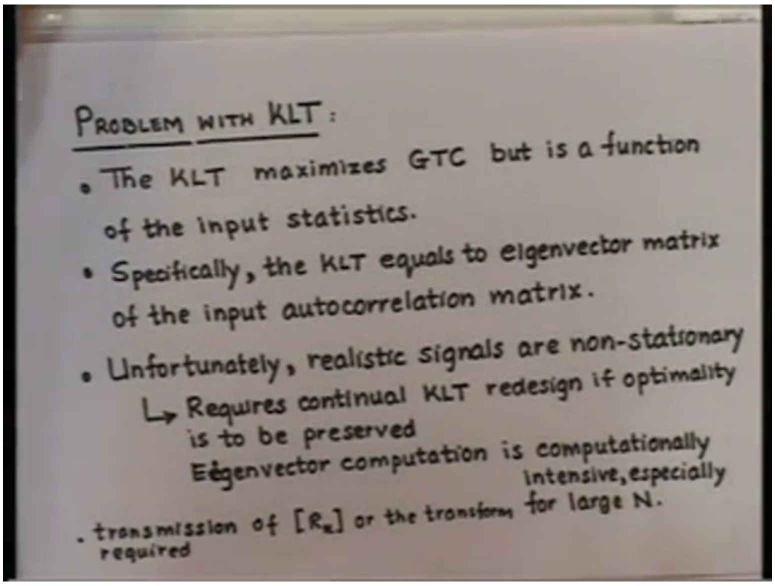
(Refer Slide Time: 44:12)



Now, the goal of the optimal orthonormal transform is to maximize the ratio of arithmetic mean of the variance of the transform coefficients to their geometric mean. We try to maximize G T C which is by definition equal to the arithmetic mean of the variances of the transform coefficients to the geometric mean of the variances of the transform coefficient. Now, this ratio attains its minimum value equal to 1.

When all this variances are equal for all k and this ratio takes on much larger values, when the difference between the variances is large or we could say the dynamic range of the variances is large. So, although the K L T is an optimal transform in the sense that it maximizes G T C, it is not practical in most circumstances, so what is the problem with K L T. So, the problem with the K L T is as follows.

(Refer Slide Time: 45:24)



The K L T maximizes the G T C that is transform coding gain, but is the function of the input statistics, so what it implies that specifically the K L T equals to Eigen vector metric of the input autocorrelation matrix. Now, unfortunately realistic signal are non stationary, so what this implies is that the autocorrelation function will be function of time. This implies that the autocorrelation matrix will change with time and this will result into continual K L T redesign, if optimality is to be preserved. Now, Eigen vector computation is computationally intensive especially for large N.

Now, another problem is that the autocorrelation is computed on the source output and this is not available to the receiver. So, either the autocorrelation matrix or the transform itself has to be send to the receiver, now this overhead can be significant and it itself may negate any gain obtained by using the optimum transform. So, the question is, are there fixed orthogonal, or orthonormal transforms that do a good job of maximizing the transform coding gain for typical input signals.

Fortunately the answer to this question is yes many transforms have been developed and discussed in the literature in the context of different application. For example, there are popular transforms like discrete Fourier transform, Walsh-Hadamard transform hard transform, slant transform, discrete sine transform and discrete cosine transform and many more. Now, for Markov process with high correlation coefficient between adjacent source output samples, it has been empirically proved that the energy compaction ability of the discrete cosine transform is very close to that of the K L transform.

In practical application, many real sources can be modeled by Markoff sources with high correlation coefficient. This is the reason why discrete cosine transform is so popular transform because its performance is very close to that of K L transform for most of the applications, and discrete cosine transform forms a part of many international standards including J P E G, M P E G and H.261.

Now, except for the scaling factors Walsh-Hadamard transform matrix consist of only plus and minus one and similarly, the hard transform matrix consist of only plus minus 1’s and 0’s. So, from the computational point of view, these transforms are efficient, but where ever sufficient computational are available discrete cosine transform is a transform of the choice.

Now, the choice of a transform is dictated by the following considerations, one is basically is decorrelating property in a particular application, the computational requirement for the forward and the inverse transformation and the size of the transform. Now, as for the size of the transform is concerned it is dictated by practical consideration, in general the complexity of the transform grows more or less linearly with the size of the transform that is N.

Therefore, beyond a certain value of n the computational cause overwhelms the marginal improvements that may be obtained by increasing N. Now, in most real sources the statistical characteristics of the source output can change abruptly like in speech signals, there could be abrupt transitions from the voice period to the silent period or vice versa or images. We could have abrupt transitions between smooth regions and non smooth regions, so if any large that is the size of the transform is large, then the probability that the statistical characteristics change significantly within a block increases. This generally results in a large number of the transform coefficients with large values, which in turn leads to a reduction in the compression ratio.

So, let us summarize our study of transform coding transform coding consist of three main steps. First step is to divide the data sequence into block of size N. This size is dictated by practical considerations and in image coding application usually the size chosen is 8 by 8. Now, each block is mapped into a transform sequence using a reversible mapping in the form of an orthonormal transform though K L T is an optimum transform in the sense that it maximizes transform coding gain. It is not practical in most of the circumstances and discrete cosine transform performance is very close to that of K L T and that is the reason why it is so popular in practical application.

The next step is to quantize this transform sequence according to the statistical characteristics of these individual transform coefficients. Now, the quantization strategy used will depend on the three main factors the design average bit rate, the statistical characteristics of the transform coefficient, and the effect of distortion in the transform coefficient. On the reconstructed sequence and the final step is to deploy entropy coding on this quantized transform coefficients in the form of some binary encoding techniques, like run line coding Hoffman coding and arithmetic coding, now at the receiver.

We have the reverse operations we have first entropy decode followed by an inverse transformation. Now, for a forward transformation since we have chosen a orthonormal matrix the reverse transformation matrix is nothing but the transpose of the forward orthonormal matrix. So, with this we come to the end of our study of transform coding and also, we have reached to the end of our course on information theory and coding.