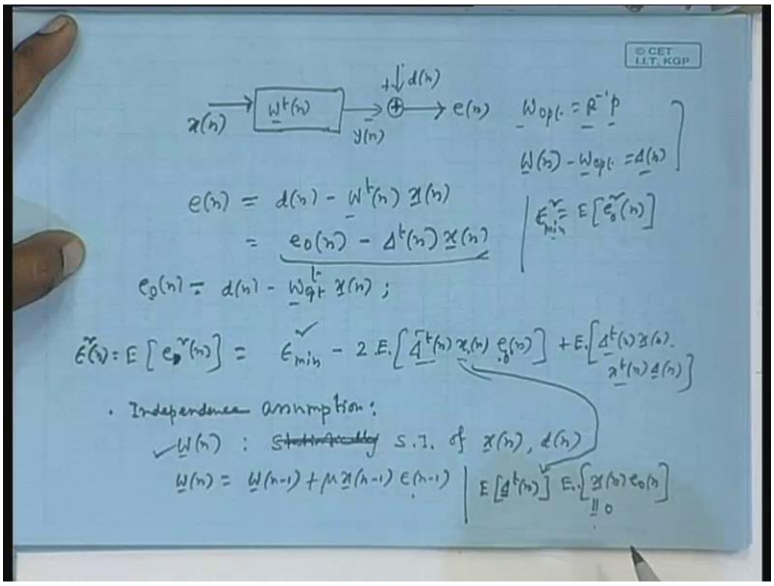
Adaptive Signal Processing
Prof. M. Chakraborty
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 9
Convergence Analysis (Mean Square)

(Refer Slide Time: 01:28)



So, it is like this let us come back to the real case everyday you know I begin the class with the diagram like this because this is my reference. But, today the weights will be it will be in a vector form you can either write w n or w transpose n we have so far proved convergence in mean. So, I am not going to repeat that part because that we have considered extensively we have first you know dealt with the real valued filter weight case then extended it to the complex valued case.

So, we have seen that as time tends to infinity n tends to infinity it has n index tends to infinity, I mean the mean of the weight vector. So, because when vector is you know every component is fluctuating is random variable, but the mean of that coincides with the optimal 1 as n tends to infinity.

So, if provided you choose the steps size within a limit from within a limit that is mu should be greater than 0 or less than 2 by lambda max. So, lambda max is the maximum magnitude Eigen value of the input autocorrelation matrix. So, then we derive the easier, but stricter ground from that 2 by trace R lambda should be mu should be less than 2 by trace R that is where I mean that is true for both real valued as well as complex valued cases. But, as I told you that if it had it been a proper pure steepest descent such that weight vector indeed would have coincided with the optimal 1 exactly.

But, I represented I replaced R matrix by a weird estimate just x n into x transpose n or x n x Hermitian n. So, did I for p vector that is just x n into d n or x n into d star n depending on the case, so obviously you are paying a price you are not finding the weight. Now, even at n equal to infinity or very large n weight is not stationary at the optimal 1 it is still fluctuating and as a result if you find out that filter error even in the steady state. So, that is called steady state when n tends to infinity actually not infinity when n takes very large values we called it steady state.

But, even in the steady state that is when indeed it has converged that is the steady statements when indeed the mean of the tap weight vector has coincided its giving you the optimal 1 that is a steady state thing. So, it normally takes 500, 400, 300 iterations you know I mean depending on the case, so that is a steady state, so even in the steady state you are not getting exact optimal weight vector. Here, you are getting still a fluctuating quantity whose mean is optimal 1 as a result the en that you have here will never be the 1 which has a minimum variance because we wanted to find out the first.

So, the optimal filter weight by minimizing the variance of this and we go the Wiener filter and all that, now under the present situation if you take the variance of E n you are obviously you will not get that minimum mean square quantity. So, a minimum mean square error that is the error which gives f rise to minimum variance because you are not putting the optimal weight exactly you are giving a fluctuating quantity whose mean is the optimal 1. But, w transpose n or w n is not equal to optimal 1 always, so then the question is how much is the deviation how much do we lose in terms of the error variance.

So, because you can intuitively see that if the tap weights are such the filter weights are such that around in the steady state around that optimal value they are fluctuating. But, fluctuating in a very narrow range that is they are spread or variance of each tap weight around the mean optimal value it is not much. So, obviously this en will be a better one that is its variance will be closer to the minimum variance that is attainable.

Now, in fact if the tap weights are such that they are fluctuating in a very narrow range around the mean, this will be indeed almost at least asymptotically the optimal error in the sense that its variance will be nearly minimum. Now, on the other hand, if the tap weights are such that the mean is fluctuating in steady state, but mean is as guaranteed by a convergence analysis mean is this optimal weight vector. But, its range of variation spread or variance around that mean is really very large, so obviously that will reflect on en if you compute the variance of E n.

So, that will be that will have you know I mean much deviation from the minimum variance that was attainable, so this 2 are related of course. So, we will just do an exact analysis because we want to keep that extra component the spread of the weight error in the steady state for each weight each filter coefficient or each weight under bound. So, then let us start with this, in fact this is a bit of repetition because towards the end of the previous lecture I did this. So, I you know as you know my practice is to always take up from the last 5 minutes of previous of this lecture and then continue.

So, E n is as usual that I am doing this analysis will be done purely for a real valued case for complex, it is more complicated we can, but very clumsy. So, we will we will not do that, so d n minus w trans in this definitions you all know what is x n vector I do not have to tell you, it is a tap weight vector, it is tap vector that goes from 0 to N-th. So, n plus 1 components and you all know this we know that w opt is R inverse p and we defined the deviation w n minus w opt was our delta n is not it, these are all known.

So, if you replace w n by w opt plus delta if you replace w n by w opt plus delta n then you can write this error as E o n minus where E o n is what the 1 which you would get with minimum variance. So, that is when you really put optimal filter weight here w opt here then the output is w opt transpose x n vector error is d n minus that which is E o n that is eon is d n minus w opt transpose times x n vector.

So, ideally I would have wanted this, but this deviation is causing this problem, so let us see effect and as I will tell you that as we go through this lengthy analysis. So, you will also learn some techniques of how to do manipulations you know how to simplify things in statistical analysis that is one major objective of this course. So, it is not just adaptive filter understanding and derivation and all that this course will give a solid training on statistical analysis of signals and systems. So, that if you find you know this kind of things in other context say communication and control system analysis you will be, you will not find difficultly.

So, you know you are used to this kind of analysis whether you are reading a paper and all that it is not just adaptive filter theory that you want to emphasize on. So, if you take the variance, variance is square of this is a scalar, so a minus b whole square equal to that you know a, a square minus 2 a b plus b square and apply expected over this. So, this is another definition epsilon square was the minimum variance is it or I was going beyond or what.

Student: No, it is using this term.

Thank you, so this we know then that was the minimum variance attainable epsilon square. In fact, you can write epsilon mean square error and obviously we are not getting this. So, what we are getting I am calling it epsilon square n because after all this is a function of n as n tends to infinity it can be independent of n, but in general a function of n. So, that will be what square it up square of this quantity that a minus b whole square thing, obviously one term with be coming from this which will give rise to epsilon. So, I am not doing those steps square it up means square it up, a square a minus b whole square.

So, a square expected value of that that will give you epsilon mean square this square is part of the notation it indicates that the quantity is of sec min is a second order quantity. So, when I put a square here just to this 2 is nothing, it is just part of the entire notation it just reminds you that the quantity that you represent is a second order quantity this is square, pardon me epsilon square n, this is e, sorry this is e, I am sorry this is e, thank you very much. So, this will be epsilon square mean minus twice E a b or b a both are same, so delta transpose n I write x n.

So, this is a scalar row vector column vector scalar this is a scalar, so minus 2 a b, a and b or 2 b a both are same because a and b are scalar. So, I write this way plus E of this quantity square up this quantities squaring up means I can repeat it and why repeat you can you know the transpose of a scalar is same as the scalar. So, I can repeat this, so I am not showing those steps just here me out I have to square it up, square it up means this multiplied by itself or this multiply by the transpose of itself because scalar and its transpose they are same.

So, that means first is delta transpose n and x n and the transpose of that, so that will bring back x transpose n here and delta n. Now, look at this here, I made that weird assumption which was very well in practice that is independence assumption, independence assumption was this I repeat again see this is a very key thing.

So, you can key weakness in the l m s algorithm analysis is that we assume that is independence assumption there are some recent research where people try to avoid this assumption make some other analysis. So, there will be some progress in that, but not to this extent that is called deterministic convergence analysis. Now, very recent work actually last 3, 4 years because this assumption is kind of you know I mean means I mean it makes us fill bit uneasy where it assumes that w n is stat is when I say independent.

So, I am writing here statistically independent, but in future when I say independent, I would mean statistically independent unless I specifically say they are linearly independent or that is another index independent difference diff different context. But, when I talk of random variables being independent, I will in usually imply statistically independence that is the joint probability density is a product of individual probability densities.

So, statistically spelling has been wrong, so I am writing s I statistically independent statistically independent of what you assume current weight vector and d n.
Now, you see I have seen that this is not a very good assumption technically because what is w n if you see, if you apply the l m s algorithm instead of n plus 1 on the left hand side if you put it n. So, it becomes n minus 1 here plus mu into x n minus 1 vector E n minus 1, this is your l m s algorithm you see w n depends on x n minus 1 vector.

But, that vector and x n vector has got overlap in, in fact almost all the components except for 1 or 2, so that means w n has some dependence on x n and this e, E n minus 1 depends on d n minus 1, d n minus 1 minus the corresponding filter output. So, these depends on d n minus 1, but d n minus 1 and d n they may be correlated it is not that there is a wide process there will be correlated. So, this also has got some correlation with d n, so it is not a very good assumption, but then it possibly works because they say that you know its product is less and then mu use is also less.

So, this contribution is less and all that that is how they try to build up an argument, but this works very well in practice. So, how it works we do not know experimentally it has been simulated it has been tried, so many times we have done it. So, many times we will give and take we do it regularly, in fact here in various contexts, so it works. So, there is something I do not know which is inside that is why people are still doing research to find out more effective co I mean convergence analysis and all. So, this assumption I make if I make the assumption then that means, why only w n delta n what is delta n it is w n minus w opt w opt is a constant quantity.

So, if w n is statistically independent of these 2, so is delta n, so is delta n, so delta n is independent of x n d n and E o n again E o n depends on what d n and x bar n this is constant. So, I can say this guy delta n is independent of both these because this is x n and this again depends on d n and x n only. So, this is independent of these two I did not use the assumption of uncorrelatedness, I told you uncorrelatedness does not work here because uncorrelatedness of 2 random variables x and y means only means E of x y is equal to E x into E y.

But, you cannot say if they are uncorrelated then E of x square into y equal to E of x square into E of y, that kind of thing will be here you see x n and this also has x n component. So, it will be product square terms of x will come up that times this, so you cannot just simply assume make use of uncorrelatedness assumption. But, you have to bring in statistically independence then only you can separate out then you can separate out this part E over that this part E over that. But, what does that mean that is this term if I write this way this term gives rise to E of and what is E o n, E o n is the error for the optimal filter E o n is the error for the optimal filter.

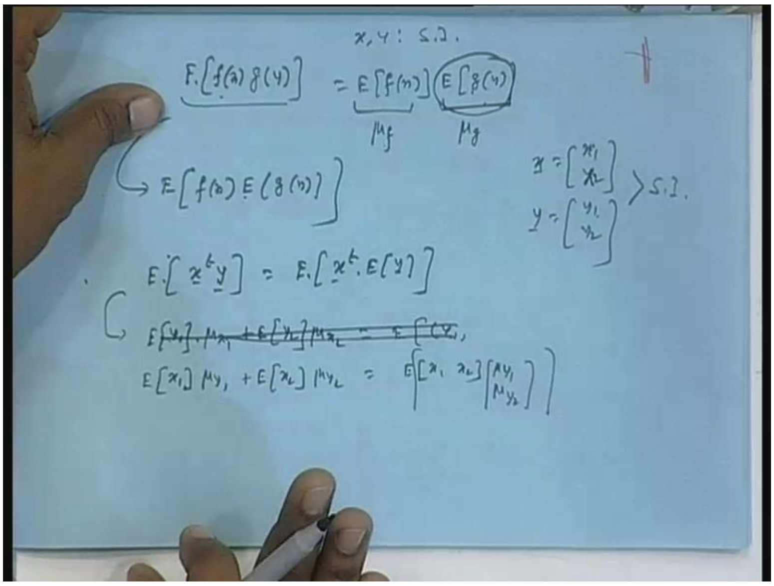
Then E o n is orthogonal to each component of x n, we have we have proved it, you remember we also proved it we not only talked based on analogy or do you want me to do it again. So, E o n whether you put it before or after it does not matter because this is scalar that times each component if you take the product and expected value there is a correlation which was that dot product or inner product. So, that it is 0 because this is a optimal 1 this corresponds to the optimal filter that time this error corresponds to orthogonal projection.

Now, that is d n is orthogonally projected on the space spend by x n, x n minus 1 up to say x n minus N which first term this one on that. But, I am not, I have not pushed good that for few seconds I thought, so earlier and then I realized that I have not pushed n to infinity in this analysis. So, it is finite, I mean I am not into steady state n a I am trying to find out the expression at any general n not necessarily 1 convergence and all and then I will slowly let n tend to infinity. So, what you are doing is you are going before you are substitute the limit at the beginning only you understand.

So, I cannot do that only in the steady state after convergence this mean is 0 that is why I am not touching that you do not, cannot assume that limit at any finite n or it before I take the limit n tends to infinity you cannot put the limit here. But, this is 0 then right this gives rise to 0 we have proved it in some of the lectures earlier, so we are left only this term and this term you see this is positive term.

So, you can see a square minus 2 a b plus b square, so b square is a positive term a square, so that is why this extra contribution comes. Now, I would have been very happy if I had only this much, but you can easily see some extra terms is coming, so let us see how much is this term. So, you can see one thing, please look at this vector delta transpose n, x n x transpose n delta n and E before that how to analyze it only thing I know delta n is independent of x n only thing I know. So, before I do that, here I will you learn some trick in statistical analysis involving matrices and all specially, but before I do minimum think of this.

(Refer Slide Time: 19:50)



Suppose, I have got two quantities x and y x and y are say statically independent x y psi and you are finding out E of F x is a function y g y is a function obviously F x also and g y. So, they are also independent it is coming purely from x this come purely from y, so they are also statically independent. So, we know this will be simply this E of F x into because you multiply by joint density joint density separable two different integrals, two different means and product. But, this is a scalar quantity this is, sorry this is, this is a constant quantity nonrandom quantity, so suppose I take it this way this is say this I called say mu g this I called mu F.

So, it is mu F into mu g, do not I get the same thing if here I bring in E on this, since they are separable are this two same mu f into mu g E, g y is mu g mu g constant it goes out. Then I have left with E, F x which is mu F since they are separable, here I can do that you cannot do otherwise, since they are statically independent separable please see this trick this is the basic. Then I will extend it to matrices are you following this E, F x g y ordinarily you cannot do that, but if they are separable because the moment you apply E over it no longer remains random it goes out and that is a only possible only if this 2 are separable.

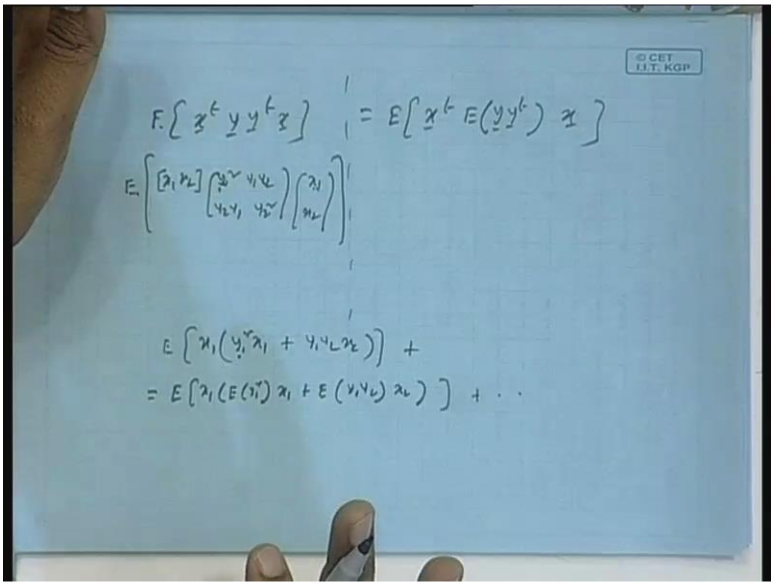
So, that is after application of E directly also they could be separated that is this has ,no I mean you see this can be done separately this can be done separately in that case you can

do that inside the outer E operation. Now, suppose you have got 2 vectors instead of x and y say 2 vectors x and again x and y are 2 mutually statically independent vectors.
So, they may be these 2 may be not independent these two may be not independent to each other, but x and y are 2 different things they are statically independent. So, any x 1 with y 1 or x 1 with y 2 x 2 with y 1 x 2 with y 2, they are all statically independent pairs suppose this is given to you they are si and suppose you have got a things like this. In fact, instead of going directly to the result, let me do some exercise this my claim is can be written as y after all what is x transpose y please see I am taking you step by step to some main result.

So, what is x transpose y if you want to do it directly x 1 y 1 plus x 2 y 2 on that you will apply E, so since x 1 y 1 there are ind min independent you will E x 1 into E y 1, E x 2, E x 2 E y 2 like that. But, if you take that E y 1 or E y 2 some constant call it mu y 1, mu y 2 it is like mu y 1 into x 1 mu y 2 into x 2 and then you apply expectation it is the same thing, are you following me. So, what do you get here E of y 1, I am writing is as it is not mu y 1 into say mu x 1 plus E of y 2 mu x 2. So, it is like E of y 1 or may be the other way E of x 1 mu y 1, so It is like E of this vector times mu over mu y 1 mu y 2 which is again E of y 1 y 2.

Now, you will get the same thing you understand the please understand the trick that is you do not, do not do this I will do, so much and then verify no please see the core of it what is the thing is they are separable that is why it is working these 2 are separable. So, whether you do this expression before hand and then do this multiplication and expectation they will then appear as constants or you first multiply and then do expectation you will get the same thing because they become separable. Now, I am just showing a result why do not you this is only an intermediate step this is true or not you have you have understood the trick here, now I make it slightly more complicated.

(Refer Slide Time: 26:04)

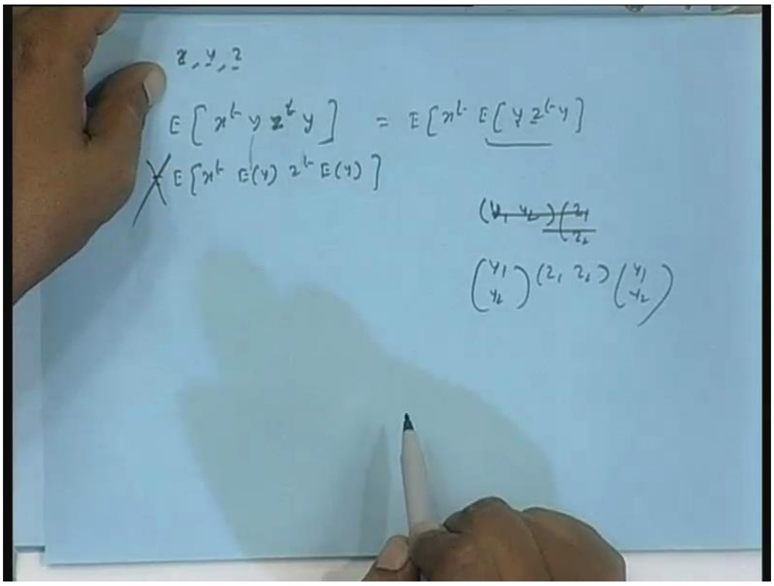


Now, suppose it is x Trans min x transpose y, y transpose x then my claim is this is same as E, this will remain as it is you can apply this one, this to get this result I made you prepared through that previous result. Now, you see this is I mean suppose you want to do like the conventional way, so this is y, y transpose is a matrix y was y had y 1 y 2. So, this will be y 1 y 1 square y 1 y 2, y 2 y 1 and y 2 squares this is this matrix 2 by 2 matrix and then you have got x 1 x 2.

So, please see that actually please try to understand what is happening inside, so that you know I mean you come across similar cases and you yourself you can tell me what will be the result. But, either you do the entire product and then apply E separately, but if you do the entire product say 1 term y 1 square after all this into x 1 plus this into x 2 that times x 1. So, there if you apply E y 1 square that can be separated out may be I work out from 1 term say y 1 square into x 1 and y 1 y 2 times x 2 that will be the first component of this product vector that with x 1.

So, that with x 1, one term plus another term E over this is separable, so I can take E on this is as good as E x 1 E y 1 square x 1 by previous arguments plus E y 1 y 2 x 2 plus the other term can I do that. So, what is this after all its like this if you apply E over this matrix E y, y 1 square times x 1 into x 1, E y 1 y 2 times x 2. Now, please see what I am doing I mean in case you want if you want I can repeat have you understood this, so this is simple technique is there do I have to explain further. So, this is, this I can do though, this is enough for me, but just to teach you further suppose I have got a situation like this x vector y vector z vector.

(Refer Slide Time: 29:20)



So, you have got a situation like this say y z transpose, so this is a matrix say y something like this, y z transpose is a matrix y z to a matrix that times a column vector. So, column row column something like this, here for god sake do not make it E of x transpose E of y z transpose E of y this is not correct in general not correct, can you see this or not. So, if y itself is again uncorrelated what you what this is equal to by our previous logic is E is x transpose into E of this part this is always true if provided z also is statically independent with x.

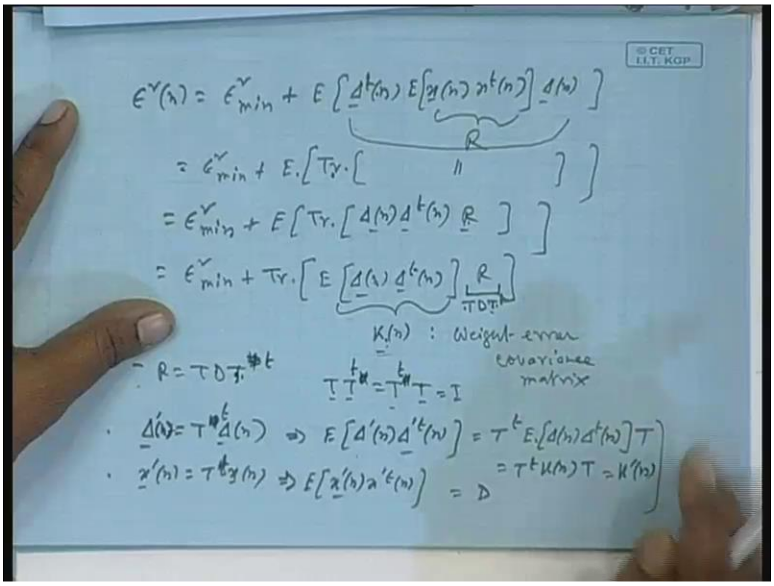
So, that the entire thing can be separated out when z also is statically independent with x then this can be done, but then this you cannot separate out even if suppose z is statically independent. So, suppose z is statically independent with y you cannot write it is like y 1 y 2, please see this z 1 z, sorry y 1 y 2, z 1 z 2 and again y 1 y 2.

So, understand there will be some square terms z 1 y 1 times y 1 square and all that and expected value of that even if z 1 and y 1 as statically independent. So, you will get E into E of z 1 into E of y 1 square E of y 1 square is not square of E y 1 that is where you will get if apply this here follow me.

So, one E of y 1 y 2 would come E of y 1 y 2 would come and their product will come up not of not that of E of y square. But, E of y into E of y that kind of thing, so please whenever you are in doubt just take 2 by 2 case and all that, so now I come back to that what was my contention, I was here this fellow become 0, I am here and here. Then obviously you see delta n I made the assumption that this delta n is statically independent of x n that was my independent assumption in terms of w n or delta n both mean same.

But, delta n is w n minus w opt w n minus w opt, so delta n also is statically independent of x n that is the assumption, however where it may appear to be I have made it. So, I can apply that logic E of this entire thing means E of same with E coming between x n and x trans min coming before x n into x transpose n that is what is that I am repeating again.

(Refer Slide Time: 32:39)



Now, what we got here is epsilon square is epsilon square mean plus E of delta transpose n then we had x n, x transpose n delta n, let us do it I can say. Since, they are statically independent, since they are statically independent by our assumption you can apply the E over this and this is your R matrix u m. So, E of delta transpose n R delta n R is matrix delta n is a column vector, so product is a column vector row into column scalar.

Now, scalar and it trace they are same please see the tricks we apply you know to simplify things. So, this is same as E of an expected value of a trace and trace of an expected value they would be the same take a square matrix, take trace and then E or take E and then trace after all trace is sum of diagonal values.

So, we apply the expectation then take sum up the diagonal once or sum of the diagonal once and then E you will get the same thing. So, suppose E of trace I take the trace of this guy to start with trace of this entire thing this thing and you have seen trace of a b is trace of b a. Now, suppose this delta transpose n into R is your a and delta n is b, suppose delta transpose into R this part I call a matrix and delta is b. So that means it is same as trace of delta transpose into R and this trace, so this is trace of still this is scalar quantity only delta transpose is a matrix into matrix trace of that trace is scalar.

So, an expected value of trace or trace of expected value they are same I told you, so now what I would do I take trace out put E in and remember R is not random. So, R can be taken outside the E operator R is constant no longer random, so matrix into matrix, you multiply and take E or take E on this. Then do the multiplication of the matrix you will get the same product this is the matrix which is not random which is constant not random this is a matrix. Now, you first multiply and then take E over each element or you better take E over each element then multiply by you will get the same thing please use your imagination.

So, that means I get into R this matrix I give it name k n what is this matrix what is delta n delta, n consist of the tap weights the will be tap weight error, tap weight error, vector tap means filter coefficient. Now, tap weight or coefficient they are the equivalent terms sometimes we say tap weight sometimes we say coefficient is filter coefficient sometimes just weight. So, the books called the tap weight error vector that is error from the optimal 1 which is again a random quantity.

So, this quantity if you take this into its transpose, so what then what will this be called weight error covariance actually it should not be called covariance. But, as n tends to infinity that time at least delta n has got 0 mean, so correlation and covariance they are same. So, keep that in mind they call it tap weight error or weight error weight error covariance matrix, so you understand as n tends to infinity we have to see what becomes this. But, essentially whatever it is you will get an extra component which will get added to this you will never get this alone.

Now, you will never get this alone because delta n is I mean this is not becoming 0, this is because delta n its mean is 0 weights only fluctuate it does not become exactly equal to the optimal weight always. So, you will have this is a non zero matrix that times R and then take trace, so some quantity it will come up question is this quantity should be kept under some bound and that bound should be our under our control by some parameter. So, that is why we will and analyze this quantity later, but you understand if this quantity becomes larger naturally it is a bad one we are deviating much from here and vice versa.

So, what is the trace into k n into r because then I could make use of the fact trace of a b equal to trace of b a, E and this is a scalar quantity scalar quantity and its trace they are same. So, you will permit me to bring in trace there, trace of 2 is same as 2 trace of 5 is same as 5, I am got this result you through this process. So, you learn some trick that how suddenly trace can be brought in because then I wanted to bring in delta from right to from right most to leftmost. Then delta into delta transpose they come together and I apply E over it R get separated out because it is constant.

So, how, no good question think of it you have got something in the, this side delta transpose n R delta if you apply E over it all terms get mixed up how will you take push R to the right side. So, that is not possible please think well that you are thinking, so these are the things I want you people to think it is not global there. So, you want to take R out and delta transpose into delta is it even dimensionally that is not such, no that is a separate question you understood or please do little bit by yourself taking. But, I brought in trace, so the delta could be brought in the beginning using the fact the trace of a b equal to tract of b a.

Then I can happily push R out and I get a covariance term here this covariance matrix what is its significance and if you take diagonal elements I am primarily interest in the diagonal what are the diagonal elements. So, diagonal elements will be, will be the variances any covariance matrix cross terms are the correlations not correlation covariance’s between various components. But, the diagonal entries they are real here positive they are the variance power that I have, I should look into that a tap weight error for each tap coefficient its variance around the mean.

Now, actually variance that should be under control, so diagonal entries are very important in this case you can see not only diagonal all the entries are important. But, that is how it turns out to be because this entire matrix to be multiplied by R and then trace has to be taken. So, not just diagonal entries because this gives an exact dependence on k n of actually epsilon square n, so in fact its value in the steady state means limit n tend to infinity. Here, this remains as it is limit n tend to infinity this you have to analyze how it behaves as n becomes larger and larger.

Now, to make life further simple instead of dealing with this what we do we replace, now R by t d, t hermitian where you know t consist of the Eigen vectors ortho normal. So, that is orthogonal and each has norm unity d consist of Eigen values which are real positive because I am assuming positive definite matrix. Then that means, t is unitary t, t h is that is I and what was and instead of delta n, you define this quantity and instead of x n. So, you define this quantity, now what is covariance of this, what is this instead of hermitian I am writing hermitian, here you understand why I am writing hermitian because even the Eigen is there.

But, in fact there is no need I can write transpose because the Eigen values are real guaranteed to be real and R matrix also has real terms then there Eigen vector components also have to be real. So, in general any real matrix does not mean that Eigen values are real they can be complex in this case, since this Eigen values are given to be your real because this matrix is hermitian. But, real and positive matrix itself consist of real elements because I am considering real value data only, so if R is real and d is real then you know how to compute Eigen vectors given.

So, you will get a solution in terms of real valued co numbers only, so this will become simply transpose, but in general they are complex . So, transpose what is this quantity then you replace this t transpose delta n t transpose delta n and delta transpose t take t out t these things you should be able to do. So, you know I mean by now you should be pretty conversant with this kind of manipulations R replaced by t, t transpose this around this transpose t will go to the right side. So, t transpose comes out t comes out delta transpose and delta transpose E over that is k n, so this is t transpose k n t 1 thing and what is the variance of this.

So, I would not tell you this is something very simple in fact it will be a diagonal matrix t transpose x, x transpose t, t transpose x, x transpose t, E over that. So, E comes in x, x transpose is R, so t transpose R t that is equal to d if you multiply t transpose here t transpose t cancels if you multiply t here t. So, this is a very standard tricks you decompose R like this and then take the t transpose takes apply t is over that x, t transpose x that becomes x prime that has got diagonal c or that becomes uncorrelated.

So, its correlation matrix is diagonal given by d this is, this you should not forget I will do it frequently this kind of decomposition immediately multiply t transpose. Now, take the x pre multiply by t transpose you call it x prime that will have correlation matrix equal to d. So, components will be uncorrelated this is very standard in communication and control signal processing this kind of whitening operation. So, if that components become uncorrelated means like a white signal white signal has got uncorrelated signal then only the power spectral density is flat.

So, this is called a process of whitening that is termed I do not know whether you come across this term whitening filter and all that in fact this is have you come across this term karhunen loeve transformation k l. So, anyway in compression and all the first transform that teach you is this is that take x this correlation matrix. Now, go get this Eigen decomposition t transpose, t transpose times x that will give you what at x prime which will have which will have uncorrelated components. So, earlier you have got components which have lot of correlation, so lot of redundancy you remove the redundancy make them uncorrelated.

Now, see who has higher variances take only those through others who do not have high variances that is how you achieve compression. Now, the problem with this transform is that you have to do this Eigen value, Eigen vector, Eigen analysis which is very computation you know expensive you can do in real time. So, they approximate it by D C T that is why all this J P E G and all that where you use D C T come in they try to asymptotically approach this. But, this is called k l transform this decompose get into this

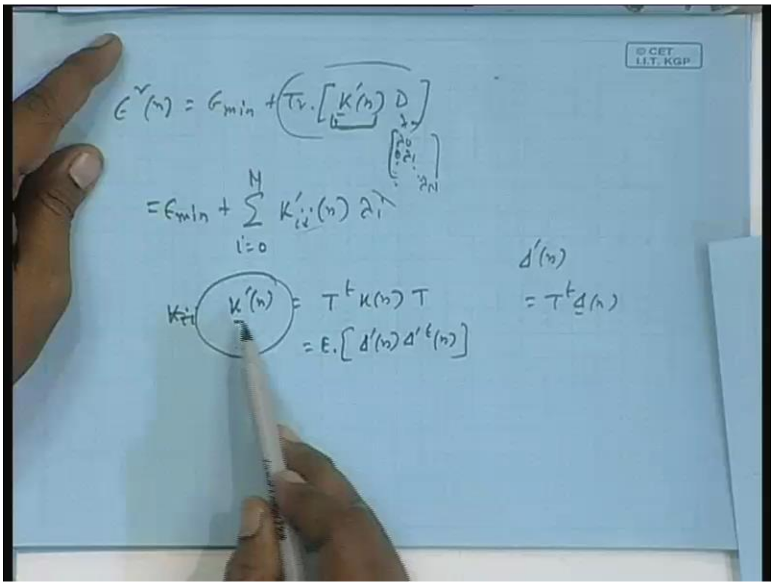
Eigen decomposition take t transpose multiply your data vector with that that is the transpose.

But, in fact t transpose consist of rows which are orthogonal then each row times the data vector will give you 1 component it is like you know I mean the those rows or if you see them as columns they become new orthogonal axis. So, new orthogonal axis in a n dimensional space you try to find give in a data vector you have to project on each component. Now, find the components those components are uncorrelated that is k l transform I am going, I am deviating. So, I cannot I am not speaking much on that k l transform karhunen loeve all compression books will start with this first you must have come across k l transform this is k l transformation nothing else.

So, this will have I can write d you are agreeing or not I am not write a doing all those steps this is d, d consist of the positive real Eigen values of R matrix. Now, if I apply this definitions here, now trace of k n R please see R, I can I want to write as t d t transpose and then apply that theory. Now, do not see it look at it as k n, k n into t d t transpose trace of that, so trace of a b is trace of b I will apply again this would be t transpose it is my b.

Now, t transpose I will take as b the remaining once k n t d as a, so obviously t transpose will come here next time t transpose k n t and what is that. So, the variance of this guy which we call k prime n are you following me I will apply that thing trace a b is trace b a. So, this is my b t transpose k n t d is m t a, so trace of a b is same as trace of b a, b a means t transpose you see t transpose k n t this part I can call k prime n and left with d, so entire thing will be trace of k prime n into d.

(Refer Slide Time: 50:45)



Now, what is k prime n if you multiply please see this is a diagonal matrix you are multiplying a square matrix by a diagonal matrix post multiplication what will be the resulting matrix. So, first column of this will be multiplied by lambda 0, second column by lambda 1, I told you how to do matrix multiplications please remember that linearly combined in the columns by the respective elements row elements. So, this is a diagonal matrix, this 1 lambda 0 lambda 1 dot say lambda n this has 0, take this first column by these elements. So, you have to linearly combine the columns of this, but this other elements are zeros other elements are zeros.

So, only the first column will be multiplied by lambda 0 second column by lambda 1 like that after that you are taking trace. So, 0-th element means lambda 0 has come there, 1 1-th element lambda 1 has come there 2-th element lambda 2 has come there. So, that means this is nothing but summation this matrix i, i-th diagonal element times lambda i, so even here you can see if this quantity is bounded if each term is bounded as n tends infinity. Then this will be under our control you cannot make this 0, Eigen values are not all zeros then there is a peculiar process and process which has all 0, I mean which you which can take only 0 values always because variances are 0.

So, i that is, that is not the case there will be this term always you have to see how behaves as a n tends to infinity what happens to this quantity that is the something you have to bear with. Now, how to keep that minimum that is what you have to see; that means, I have to study this quantities. So, what is k i, I just repeat this will be the in a most very elaborate analysis spending possibly 2 days or 1 and half days. So, what are this quantities they are coming from this k prime n what is k prime n, k prime n was your t transpose k n t that is E of delta prime n where delta prime n was t transpose delta n.

So, this quantity figures here in terms of its diagonal entries we will study this quantity how it behaves this matrix as n tends to infinity. So, that means I have to substitute all this delta prime by t term delta n and all those things here and do some analysis. So, that is a very lengthy though interesting by now you are prepared for it some other tricks I have shown there will be repeatedly applied here there here there. So, please sharpen your mind before coming on next class, thank you very much, yes anything yeah you want this. So, the class is over if you have any question you can ask me because I see they have not switched off yet.

Thank you.