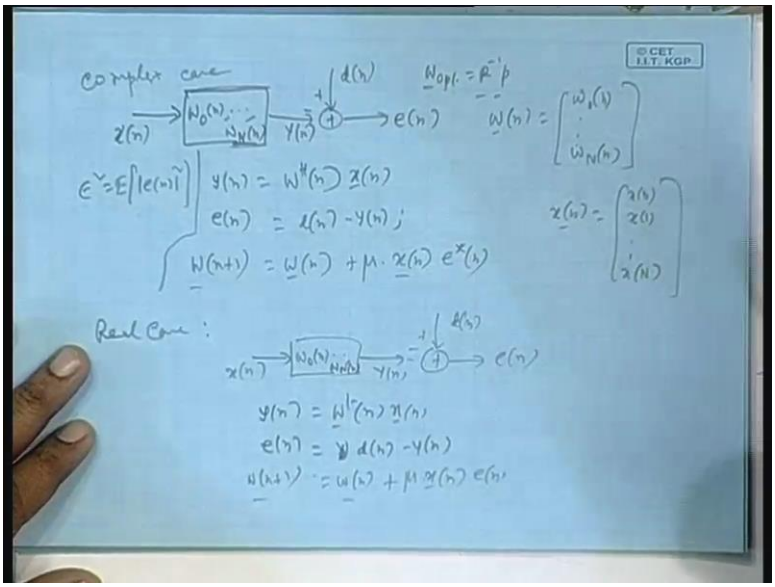
Adaptive Signal Processing
Prof. M.Chakraborty
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 8
Convergence Analysis (in Mean)

(Refer Slide Time: 01:10)



We had just concluded with LMS algorithm for the complex case. So, let me just quickly go through those steps there we had this thing; x n was a zero mean zero mean random process complex value random process. The filtering problem was this w zero n dot dot dot say w capital N n d n y n; dn also is a zero mean complex value random process if you subtract the error this is e n you define w vector as w zero n dot dot dot w capital N n. Then, the departure is the filter output we are talking in terms of w n there is w zero to w capital N as a coefficient, but when you filter if filter with their conjugate values. Alternatively, if you filter with some complex valued coefficients we evaluate their conjugate and then get back the filter coefficients where the conjugation.

So, then our equations were like this y n was w H n x n vector x n as usual. Then, you find out e n as d n minus y n and then update as mu into x n vector e star n. You can write e H n also because e n is a scalar. So, star of H there mean the same there is a complex valued case. Here, the optimal filter would have been what optimal filter that is w opt is still R inverse p, but then you use w opt Hermitian multiplied with x n, but then

you get y n. And this w opt minimizes epsilon square, which is in this case E of mod e n square. This was opt that is a complex LMS case is the real this is a complex case.

For the real case, you can view real case as a special case of this because at Hermitian or normal transposition means the same thing originally has means the same thing. So, in that case again you have got x n the filter weights w zero n dot dot w N n y n d n plus minus this e n; w n is and x n they remains same in this case. So, this is a special case of this you can see now; because y n is simply w transpose n you can write w w transpose n x n you can write w Hermitian here also, because under real case there is no question of conjugation.

Say e n as y d n minus y n and w n plus one is w n plus mu x n vector as it is and we write e n e n or e star n they are same because they are real. So, this is the special case of this. Remember, how you obtain this algorithm we first started with the steepest descent, which was on offline procedure iterative offline procedure. We took that plot of epsilon square versus all the tape weights that was a quadratic function, which has unique minima. So, we started with one point and then went in the opposite direction of gradient multiplying the gradient value is sub suitable constant steepest mu and then when back and forth around that optimal point and then finally, converged on that.

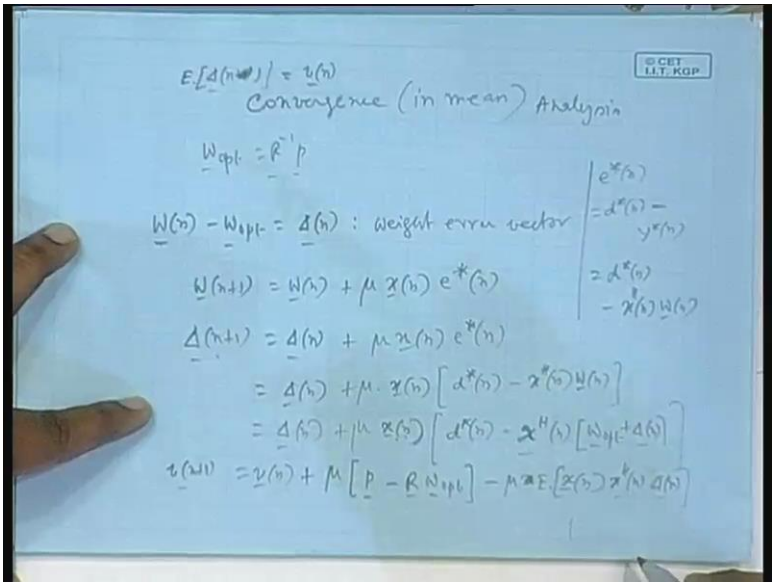
That was steepest descent on steepest descent what you need we replace R and p by some wiener estimate; R by simply x n into x Hermitian n and p by x n into e star n, which is the very wild estimate, but we say that still algorithm will work. But then since you are not giving the current value of R and p definitely this will not converge exactly on w opt. If you had given it exactly I mean the correct value of R and p,, it is a simple steepest descent exact steepest descent procedure and it should we can show it will directly converge on this thing the minimum point, but since you are not doing that. You are going only very approximate value for R and p necessarily you have to lose something.

So, we will not your filter weights will not converge directly on the optimal weights, but it will converge actually in some other way and that convergence is its mean. There is a mean of the tape weight this filter weight vector as time tends to infinity see everywhere it along the time axis it is fluctuating, but as time tends to infinity it will be fluctuating

around the optimal values that time; if you see the mean that mean will be optimum value. It will still not be optimum, but it its mean will be optimum then we will see that how to keep the variants around the mean or spread around the mean under check under some control.

In fact, as well as possible that will be a better kind of convergence that is that is how things will evolve. So, we now do this convergence analysis as I said that the complex analysis is more generalized include this special case also real case as you can see by comparing the two algorithms now. We will do the convergence analysis for the complex case. So, what I will I do?

(Refer Slide Time: 06:48)



So, this is convergence in mean convergence analysis. You know what is w opt? That is R inverse p. Now, your w n minus w opt that is the error vector error in filter weight we call it weight error vector. Ideally, it should go to zero as n tends to infinity, but because you used approximate values of R and p it will not go to zero , but this will have a zero mean will be fluctuating. We call it I denote it by delta n is a vector it is called weight error vector, weight error vector. Now, your equations were this is the relevance equation you subtract w opt from both side.

What you get here is delta n plus one delta n plus same thing, but within this e star n dot e n. So, I want to write both left hand side and right hand side in terms of delta n w. I want to remove w and bring in delta apparently it might appear to that you know I mean we have done our business delta here delta, here what actually when this in this e also a w is z n. What is e n? After all d n minus filter output y n and y n is w Hermitian n into x n vector. So, there w lies. So, I have to expand it and again replace that w d star n minus y star n please see d star n minus y star n. I am writing separately e star n is d star n minus y star n same as d Hermitian n minus y Hermitian n because they are scalars.

Then, if you replace this y star by y Hermitian then this becomes what y n is w Hermitian n x n. So, y Hermitian is x n Hermitian w n that is what will come here; w n and this w n I will replace yes w opt plus delta n from here, w n is what w opt plus a deviation is delta n so far so good. Now, I apply expectation operator on both side. E of i define E of delta n plus n delta n expected value of that the deviation it is not zero some expected values. So, there is delta n is fluctuating around something. If that something is zero; that means, w n is fluctuating around w opt. So, suppose e delta n we say it is v n.

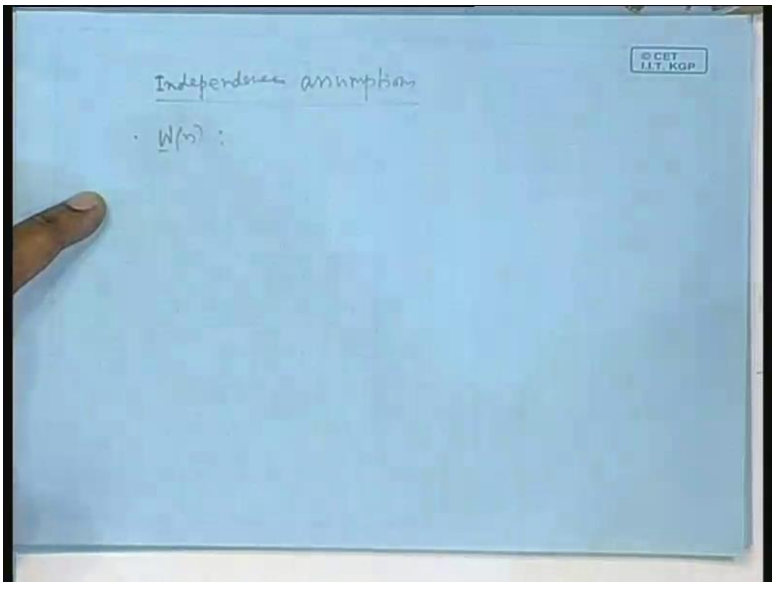
So, v n is not random after you apply expectation operator randomness goes. But remember I am putting a n here because it is not really stationary because filter weight the weight vector what from delta n come delta n is nothing but w n minus w opt; w 1 is constant, so delta n is equivalent to w n, but w n is not stationary processes is changing now with time by a recursive equation LMS of that equation. LMS of that equation it is not like a pure high purely random process, what there is no bias on particular time access whether you observe mean here or there or elsewhere or variants you will get the same thing it is not.

There is a recursive equation by which w n is generated from its passed value. So, it is not you cannot say that is stationary that way because there is a relation. That is why I am keeping an x n. Now, if you do that now apply e on both side; obviously, you get v n plus one as v n plus mu into x n d star n if you apply what that what you get x n into d star n p vector cross correlation vector mu into x n d star n cross correlation between x n and so you get p minus x n x Hermitian w opt; w opt is constant. So, R will it remains outside expectation operator; so only x n x Hermitian n that is R that is R.

So, we get R w opt and another term that term is very important. So, that term I write separately mu x n E x n x Hermitian n delta n. It is product three random quantities one vector row column vectors row vector another column vector. So, it is actually a scalar, but there are three random quantities. This is also random this is random and this random I am coming to that later. Now, consider this w opt is R inverse p. That is p is R w opt. So, this kind tells p equal to R w opt. So, this is equal to 0. So, I am left with only this or of course this. This term is crucial term and is very difficult to analyze things further and it is here the inventor of this algorithm Widrow made some assumptions.

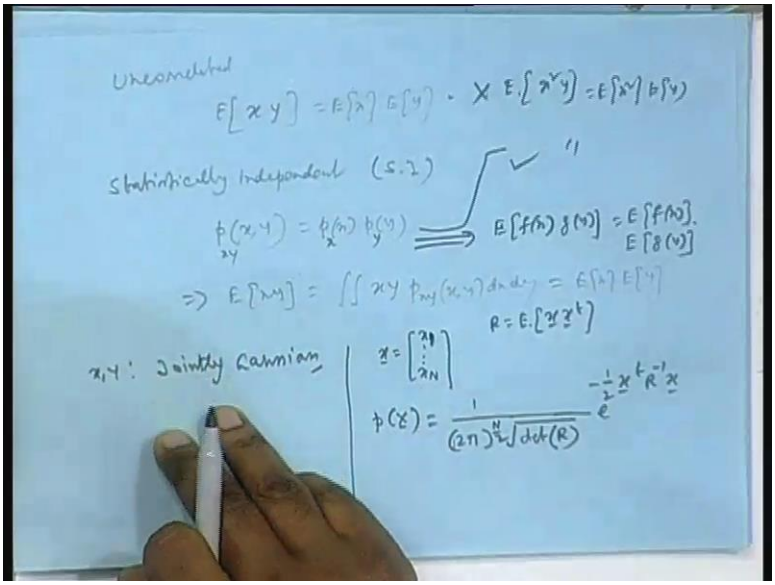
Those assumptions, I mean you can say where they are questionable assumptions because we can always ask question we can all then of course, he has a way to justify also the assumptions, but those assumptions work in practice. I am just this assumptions are called independent assumptions, independents assumptions. What are the independent assumptions? Let me write here.

(Refer Slide Time: 14:34)



We assume w n vector, it is statistically independent not just uncorrelated. You know that statistically dependence is stronger condition than uncorrelated. Do you know this or may be for your safe I have to say.

(Refer Slide Time: 15:05)



Suppose, x and y x and y just see this this is a departure from the discussion. Suppose, x and y to start with assume they are real valued. So, if you take x and y and if they are uncorrelated, uncorrelated means is E x E y if they are complex valued would have been E x y star equal to E x into E y star, but just lets confined to real case this is one uncorrelated and statistically independent if the two variables are not just uncorrelated.
If I say they are statistically independent and loosely I will just called independent; that means, statistical independent only not linearly independent, there is something called linearly independent that I told earlier.

That one is has a linear relation with other there is still a stronger condition, because you are as assuming a specific form not just dependence specific form of dependence. Statistical independent, statistical independent if they then; that means, there joint probability density is individual density product of two individual densities. You can say p x y for the joint case is p x x p y y. If I otherwise write p x into p y mathematicians will found they assume I am using the same function as though x and y have the same probability function, which may not be the case just to differentiate I put p x of x p y of y just for mathematic I mean to say ourselves from mathematicians.

Now, if this is given my point is if they are statically independent that is SI statistical independent then they are uncorrelated not the vice versa. If they are statistically

independent, what is E x y? E x y is nothing but what is E x y E x y is nothing but x into y into p x y x y dx dy and the rest is very obvious; p x into x one integral p y into y. So, it is E x E y same thing applies for complex valued case also, but given this you cannot get back this. given that E x y equal to E x into E y in general you cannot get back this. But there is one case you see this is not I may not between dealing with that right now, but just for knowledge sake and also one purpose of this course that you not only than adaptive filter through this process, you develop lot of techniques to handle statistical quantities or statistical analysis.

For that purpose, I tell you that is there is one case there is one case, where one means the other and vice versa. That is E if x y are jointly Gaussian zero mean I am talking at a talking of zero mean cases. So, if they are jointly Gaussian if they are jointly Gaussian, then this thing then this will happen why you know. If x y jointly Gaussian you know what is the joint Gaussian formula joint Gaussian distribution of a vector. Suppose, I have got a zero mean vector this case I handle here, in general can you I we can extend these further if E x y z E x E y E z like that you can extend further. Now, we suppose consider in general case why even x y general multi variant Gaussian or vector Gaussian thing Gaussian vector or multi variant Gaussian case.

There, if I give a vector x which has x one dot dot say x n or may be x zero to x n and I say they are zero mean, but jointly Gaussian. That means what is that density to probability density of the vector is this; one by 2 pi may be I take x 1 to n you know others we have to write n plus one unnecessarily one to n here. So, to number of variable is n. So, 2 pi into n by 2 then determinant of if I say R and square root of the positive square root of that. What is R? R is I am not composite I have some more terms are there R is E of that is a correlation or covariance they are in the same here is zero zero mean case; this thing into e to the power minus these are thing.

In fact, if it is a you can now verify if it is only one variable if it is only one variable then what happens then 2 pi into 1 by 2 1 by root 2 pi determinant R means R is simply the variance of a single variable square root by sigma. 1 by root 2 pi sigma e to the power minus x; x transpose is only x only one element that is x one call it x x one square takes one square. So, x 1 square by R is only a scalar sigma square and there will be a two here

one by two. No, I am telling a zero mean I am dealing with zero mean there just to make life simple. This is a zero mean multivariable Gaussian distribution.

Now, suppose here it is said that this variables are uncorrelated. Zero mean and uncorrelated means the correlation will be zero E of x 1 into x 2 will be E x 1 into E x 2 as which is zero. Only diagonal elements will be non-zero positive number because they will imply variance. That means R will be what a diagonal matrix consist of sigma one square sigma two square dot dot dot sigma n square inverse of that one by sigma one square one by sigma two dot dot dot. Then, you can easily see you can you can break it up as product form may be you can take as an very this is very simple exercise.

What is determinant of R? Sigma one square sigma two square dot dot sigma n square; square root means sigma one sigma two sigma dot dot. So, you can spread it as to root two pi root pi root two pi n times, sigma one sigma two up to dot dot sigma n. You will have e to the power minus x one square by sigma one square into e to the power minus x two square by sigma two square of course, with the half dot dot dot. So, it is a product of individual Gaussian densities. I am not showing that I think is pretty obvious. So, in that case this happens if they are jointly Gaussian then uncorrelated means independent and vice versa.

This is always true it dependents means SI means uncorrelated, but not the other way except for the Gaussian case. Another thing for our analysis as we will see; I may specifically mention that its independent assumption we will assume statistical independents of something. Why I mean you could have lift with uncorrelatedness you know when this will not work, when you have to go beyond this assumption because we have doing some analysis. You want to make minimum assumption. So, instead of using uncorrelatedness you are finding that you cannot still solve go for that then you have to go for statistical independence.

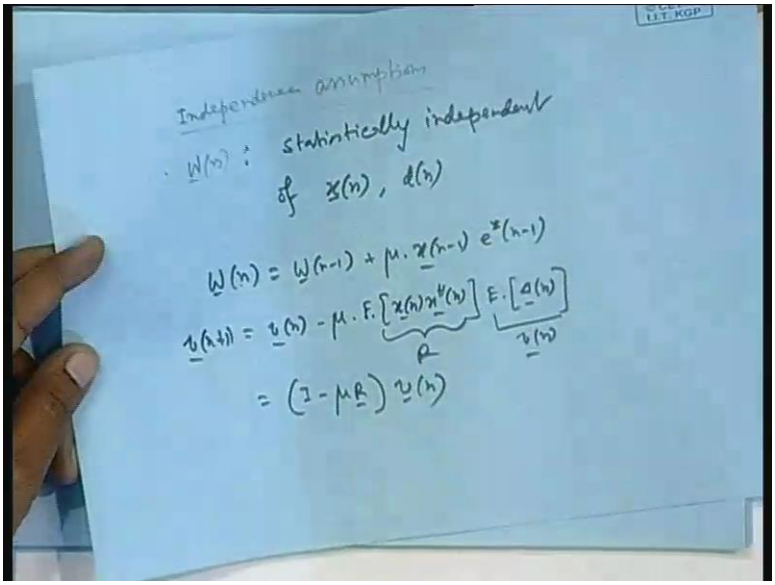
What is that this gives and which uncorrelated does not give? Suppose this is given, but it will never mean E of say x square y is E x square into E y. It will never mean, but if this is given then this will be true; this will be true because x square y joint density will be still satisfying this. If x and y are statically independent x square and y also a statistical independent. So, you multiply with the respective joint density and then you

can separate out you get this. If pi only this then E of if this is given this will imply E of any f x g y will be E of f x into E of g; these are not given in books, which is not written you have to understand.

But this does not apply does not follow from uncorrelatedness, if you handling with kind of cases, where x and y directly does not occur, but some function of x and some function of x and some function of y occur in your expressions you better go for statistical independence not uncorrelatedness. Now, I come to that independence assumption remember these we are doing. Just to give a q this is where we are left with this factor become zero v n plus one is equal to v n plus v n minus mu into this term.

Here, I have got a product of three terms one vector row column vector another row vector another column vector and out of which delta n is what after all it is equivalent to w n in the sense that is w n minus w opt and w opt is a constant not random. So, delta n depends only on w n and nothing else delta n is dependent only on w n find. Now, we make an independence assumption and this assumption might look us but we do.

(Refer Slide Time: 24:50)



It is assumed that is a statically independent of x n vector and d n scalar; x n vector and d n scalar. Now, it was this is why this because you look at weight of that equation w n if I put n on this side it will be w n minus one plus mu into x n minus one vector into e star n minus 1. So, w n depends on x n minus one vector what you see x n and x n minus vector they have a lot of overlap x n vector consist of x n then n minus one n minus two up to n minus capital N. I mean second to n th term here part of the first to n minus one th term common. So, you cannot say so this actually is related to x n vector also even; otherwise there is a correlation in data even if suppose w n dependent not on x n minus one, but on say x n minus ten or twelve.

Even fast data and current data vector they have some correlation; you cannot assume it to be wide. So, that way w n has got correlation with x n some kind of relation, but we make the assumption. So, we got we got an equation we got an equation actually because I have meeting today it has to you know take to it is could not reach you I mean here the video recording to because today is an important lecture anyway. Tell me, this is an equation you obtained LMS equation. So, this assumption I am that is why I am saying this assumption looks weird. Because I am this is by dynamic equation how w evolves with time from its passed value it gives the w dynamics.

Clearly, shows that it depends on data, which is part of current data vector. Then, you know people say that you know this because the mu is very small and this error is small. So, in this equation this has very less contribution compare to this. So, that kind of round about logic some analysis on that people do, but these are this is what assumption works in practice. Now, if that we show for the current analysis; I need only this statistically independence x n vector not d n; but subsequent analysis I will need that. So, that is why I have written the co independence assumption once for all; because I told you right now, we will be proving convergence and mean afterwards.

We will see how the variance around that mean can be kept under bound can be kept under control; that is for more complicated very rigorous analysis just for analysis in mean square this only analysis mean. So, that time I will need independence of w n not only why sub is x n, but also why sub is d n. For those analyses only we have used it and it works, but this is the how it is you know. Then, say if you see this book by Farag and we say that because mu is usually small and dz is small and this product and product for a small quantity its contribution in overall thing is less.

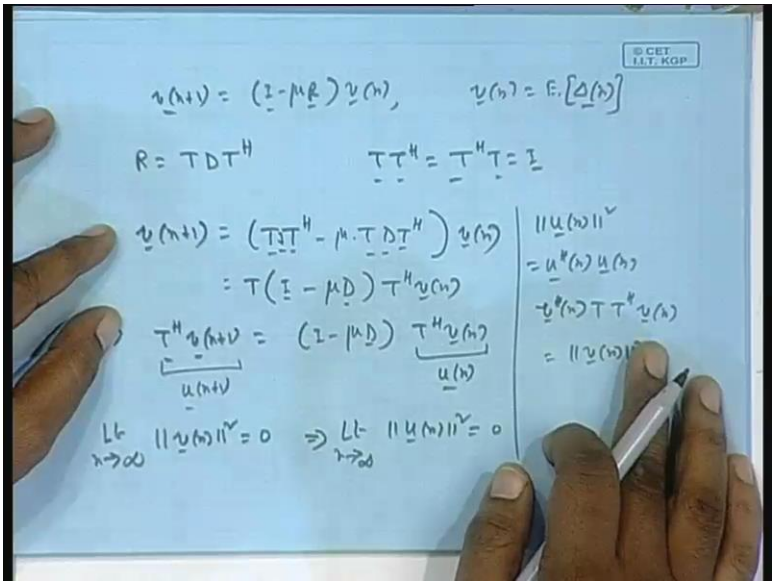
So, that takes down the that lowers the correlation presence of correlation all that people they say, but anyway fact is that this is an assumption that make otherwise you cannot proceed; because what will you do after this if on the other hand if you assume statistically independence. See, why independence and why not uncorrelatedness because delta n it has got w n delta n has got w n vector, but this term this matrix will consist a products x n x Hermitian will consist of what products of data of the same x n vector. So, two terms multiplied of x n multiplied by say a w n component expected value of that typically.

So, they are uncorrelatedness will not work as I told you uncorrelatedness only work if you have single like this E of x y. Then, you can E of x into e of y, but you have got E of x square y. You need statistically independents to write it as x square into E y. So, that kind of situation you have here in this matrix each term has got a product of two data terms coming from x n vector only. That is multiplied by some component of delta n which has one component of w n. Now, you apply e over that each term. So, that e under e there will be three terms.

This is a matrix forward by a vector again it to be a scalar, but scalar component you can write as a summation of many small small scalar terms. What will be the typical scalar term? Scalar term will be a product involving two data samples and one component of w on that e operator. Only if you have statistical independence valid then you can separate them out that is why you bring in statistically independence are not uncorrelatedness. So, under that assumption what you have then you can write this E of x Hermitian x n x Hermitian n separately and E of delta n separately.

That is you can write v mu and this quantity same as v n this quantity same as v n. What is this R so; that means, this becomes equal to v n minus mu R v n and you can write it as I minus mu R into v n R is a Hermitian matrix. In fact, we can assume what to be positive definite. We can assume what to be positive definite. Today the class started at three we could not reach you all, but I thought you cannot come because of your other class today there is a meeting. So, I have starting with follow the video lectures. R is Hermitian matrix. In fact, R is a positive definite matrix we assume. So, we remember in such case, so I just reproduce the equation here for your benefit.

(Refer Slide Time: 31:45)



This is the thing and v n was expected value of delta x this is what we are. You remember, we discussed at length properties of Hermitian matrices and positive definite matrices at that time I said that I can have R as some T D T Hermitian, where T consist

of column vectors, which are the Eigen vectors of R; in general complex valued Hermitian transpose of that. Eigen vectors are mutually orthogonal; one column Hermitian times, another column is zero and you can take the norm of each column to be one; you can assume the norm of each column to be one. You can normalize it you have to remember.

So, that is why T into T Hermitian was identity T Hermitian T was identity we call them unitary matrices. You remember early this was done T T Hermitian was T Hermitian T was I. Do you the D consist of what Eigen values and from positive definite matrix Eigen values are not only real for Hermitian they are real for positive definite; they are also positive this where. So, I replace it here T I we can always write as T T H I we can always write as T T H. In fact, T into I into T H you can put an I here also minus mu T D T H into v n. You can take a T out here I minus mu D T H v n pre multiplied both side by T H T H T will cancel; TH multiply this by T H multiply this by T H.

So, you will get T H v n plus 1 as I minus mu D T H v n T H v n we define as another vector u n. So, that has been this is u n plus 1 essential thing is if you if whenever you have an unitary matrix multiplied a vector the resulting vector has the same norm square as the original one. I will prove it is very simple, but you know unitary matrices like they reflect operations like the rotation by after rotation link does not change translation they are all unitary operations in real life in 3D world. Now, norm square does not change because after all what is norm square of these vectors that is u Hermitian n u n. You remember norm square of vector mod square of this; I think you people did not coordinate you told me you would inform others people are coming now.

Because many are coming I thought I mean we are able to coordinate it today there is a meeting important meeting. So, I will leave at four. So, I will say the class started at three and Jaydeep took the responsibility of informing everybody. So, he needs the kind of blessing from everybody else. Anyway, norm square of u n norm square u n is this you want to understand this. Any vector norm square is what the vector Hermitian into vector itself. So, mod square of first term mod square of second term mod square of third term all added. The norm square if you do if you put that here, what you get u Hermitian means v Hermitian T this u Hermitian if this is u take the Hermitian v Hermitian T.

Then, u n replace T T H v n and this is I I v n is. So, so this is again. So, whenever you have one matrix two two vectors related to each other by a unitary operator operation there norms are same. So, my purpose is to show if I can so now, that limit n tending to infinity equal to zero this is what I want to show. What will it mean? That v n expected value of delta n v n is such that if you take the norm square that will go to zero as n tends to zero, but norm square of a vector zero each component has to become zero. You remember norm square of a vector zero means each component has to be zero; that means, if I can establish this I will establish that has time tends to infinity each component v n goes to zero.

Then is each component of expected value of delta n goes to zero. What was delta n w n minus w opt that will only go in that expected value of w n will go to w opt. This is your delta n if e of this goes to zero; that means, the e of delta w n goes to w opt. So, it will converge in mean that is what I will prove, but you remember I told you just now norm square of v n is same as norm square of mu n. So, it is equivalent to proving limit n tend to infinity if I can prove this I prove this also because norms are same. I have just now proved.

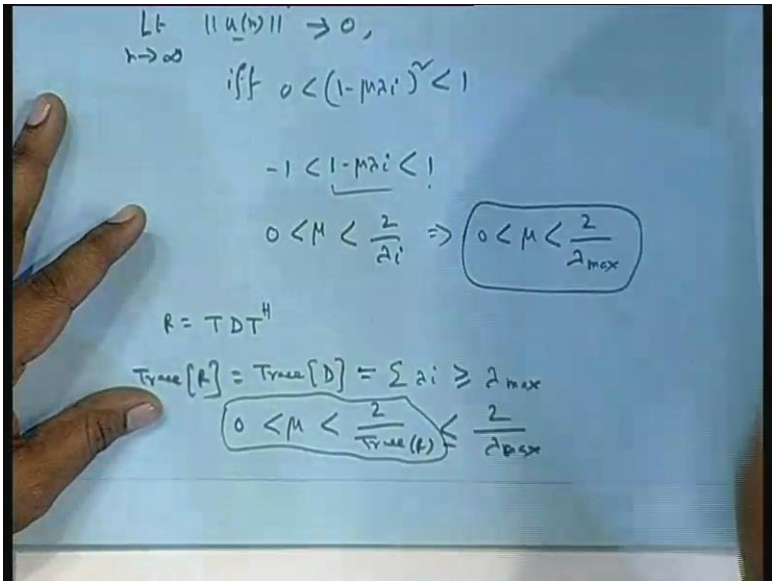
So, now let us look at these equations if I take the norms of after all. What is this? This is a diagonal matrix, I minus mu D is a diagonal matrix. What kind of matrix? I am writing separately this is a rough space I minus mu D this is like one from this I minus mu lambda zero 1 minus mu lambda 1 dot dot dot 1 minus mu lambda n. Some lambda could be same so that times of vector norm square. So, I can write if you take the norm square of this side this side that is same as what this diagonal matrix time the vector is another vector norm square of that.

Norm square that means 1 minus mu lambda zero times the first entry of u i n square of that. So that means, 1 minus mu lambda let me say I square u i n square i equal to zero to n it was simple u n each term u zero u 1 u 2 and they are multiplied by 1 minus mu lambda zero 1 minus mu lambda 1 or 1 minus mu this way. This is simple I am just writing this the diagonal matrix types a vector. This will multiplied first entry this will multiplied second entry like that and square them up. So, I am put an one minus mu lambda could be negative or positive that is why I am putting I should put a mod square of mod square n square they are same because they are real square into this.

In fact, I should put mod here because component of u could be complex. So, this now suppose by hook or crook I can make the coefficient one minus mu lambda i less than one one minus mu lambda square whether one minus mu lambda square to be less than one. It cannot be negative I will not make it zero I cannot make it zero, but suppose this is between one at least there is a one. That means, each component of I mean each component like mod u i n square will be multiplied by a constant less than one.

So, what I will think will be less than the norm of norm square of u n if all were one you get norm square of u n, but every each term is coefficient is less than one then this quantity will be less than the norm square of u n. In that case, we can say that as time marches out the norm square decreases u n has so much norm square, but u n plus one has less; u n plus has. So, much u n plus two has less it will progressively go to zero that is how it will be proved.

(Refer Slide Time: 40:19)



So, I can say that limit n tend to infinity square goes to zero if and only if one minus mu lambda i square less than one greater than zero. It cannot be equal to zero is a square if you equated to zero it has to be equal to zero one minus g. So, this has to be one this is problem actually. It cannot be equal to zero for all lambda. So, this is supposing this so that means, one minus mu lambda i should be between this. So, from this side you get what you take mu lambda e to the right hand side one one cancels mu into lambda i

greater than zero, but each lambda is positive and real because positive definite matrix. So, mu also has to be g greater than zero.

Between these two sides if you take mu on this side one or minus one on this side you get essentially mu then on this side you get mu greater than zero on this side two by lambda i. Each lambda is real; that means, you should have this condition on mu this is the condition two by lambda max because lambda one lambda zero lambda, whichever is the maximum you have to you have to take to that only. It has to be less than two by lambda for each i. So, whether take the maximum one there will give the actual bound.
So, this is the universal bound for mu if you cross this you are finished algorithm will never converge you can just carry out an experiment in lab you see it or not. But sometimes you know computing Eigen value and all this a problem.

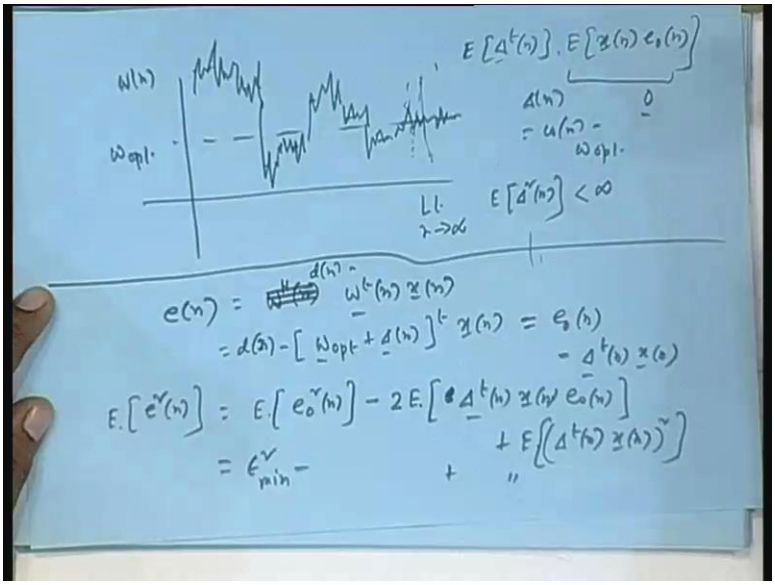
So, from this we derive a slightly stronger bound. You know that R is T D T Hermitian. So, can we say that trace of R is same as trace of D that you have seen already? Trace of R that is summation of lambda i, which is greater than equal to lambda maximum. Now, 2 by lambda 1 2 by 2 by lambda zero lambda 1 lambda 2, which whenever the maximum I have to take that because mu has to be less than that also. So, it will be lambda max only lambda zero lambda one to the maximum one. So, two by lambda will be minimum one I have to satisfy that to; that means, is better restrict to that only that is it very simple logic.

Now, you see trace R e trace D s summation lambda i, which is greater than equal to lambda x of course, because in the summation lambda there is at least one lambda x others can be they are positive or zero can never be negative we know other if it is positive definite of that should be only positive. So, that means 2 by trace R two by if you take reciprocal 2 by trace R is less than equal to 2 by lambda max. So that means, if I keep this term if I keep this two by trace R that will satisfy this. So, that is why in practice you know we stick to this one.

Trace R is not difficult if it is a stationary process trace R after all we will consist of that is variances if it is stationary all the variance terms will be same toeplitz matrix. Toeplitz same variants whether it is sample n th sample where n minus one n minus two all are same variance. So, just matrix number of row times the individual variants where you

have to put it here only. If you choose mu within this zone you will see this converging in mean, but this is not enough you have to ensure that it convergence in mean square also. Mean square means now I understand what is the physical implication of this that if you take just a single weight case because I cannot draw for multiple weight.

(Refer Slide Time: 44:56)



Suppose, this is the w opt; this will be mean that if you take w n w n is fluctuating. So, around this zone its mean is here. There are this zone is mean is here may be mean is here mean is here mean is here like that. Finally, it will be like this, but it can also be like this only it convergence it is mean. So, I understand that as time tends to infinity the w n will have mean around mean on w opt. That is some gain at least some kind of convergence this convergence mean, but next we have to show is this if you take the w n minus w opt delta n here, which is a scalar in this particular case. Then, you have to show that this quantity the variance this also remains bounded.

It will be such quantity it will be quantity then computable finite some closed form value and they are using some parameter we can keep the value as low well as possible. So, it will fluctuating, but within a bound and that time we will see if you take mu high, then what will happen the spread will increase if you take that mu within that range mu one the lesser side this spread will be less. But I told you mean will converse to mean of the weight mean delta n converge to zero delta n delta n mean of delta n converges to zero

that is, what is happening the delta n is the deviation. This is now here delta n whether this much this much this much.

What is the mean of that? Mean of this is here delta n if you subtract w n and that prove I have not done. So, far that is a very rigorous lengthy analysis, but I will take it up. One reason because through that process you will learn lot of tricks of how to analyze you knows statistical things equations and all that. One purpose of the course also impart that thing of statistical signal analysis context is adaptive filter, but as you go through that mass you see you know clever mathematical tricks here and there. So, that is a very lengthy exercise it will take two days. It is a big clumsy also you know there were lot of cross terms this term every one we have to analyze some will cancel some will not.

Finally, we will get a recursive equation of this variants and then we will see the variants will finally, converge to something provided mu is I again chosen from this range from another range. So, that time we will see that this variance will depend on mu the mu is large then the spread will increase large means still within that range then only, it will converge in mean, but the spread will increase if you take mu on the lesser side then spread will be much less. The point is the mu if mu is large then it converges faster it take much less time to heat up on this zone, where the mean will be the actual the actual weight optimal weight a mu is less it will take more time. So, there is a kind of trade off.

No, remember one thing remember one thing that we have deviated from steepest descent. The movement I removed R and p by those we are not doing steepest descent that we are not having that exact equation. That exact equation was governed by R parameter p parameter. That I have taken away R and p are time varying now; x and x transpose n or x n x Hermitian in p is x n x. So, that does not I will cannot directly use that, but still tell me, from that it can be appear to you that will crawl if you are really not doing this I could have use their argument also you; if you are not doing this suppose you are doing only steepest descent pure to steepest descent that time.

You can see that if you take mu small we will be going by small small small steps it will take more time. But it will you are sure to converge, but if you take mu large you will may be jumping to and so much that earlier from quite some time there will be lot of fluctuation you know rapid fluctuation, but finally again it will directly converge.

Unfortunately, here it does not converge directly because R and p have been approximated by some values. So, that analysis is called accesses that mean square error analysis. So, for that we will first see this that suppose this is e n we all know what is e n; e n is now for that analysis I will excuse me I will have to I will stick to again real case for the complex value they are really complex.

So, w transpose n x n, but w transpose n is w opt plus delta n transpose x n. So, we have got two term one term is corresponded the optimal weight e n is d n minus this. So, d n minus w opt transpose x n what does it imply that is when you really put in the optimal filter. That time the error is d n minus w opt transpose x n vector. So, that error is the one which will have the minimum mean square error. That is that is why it is w opt how was w opt obtained by minimizing the mean square error, which is the quadratic function of the weights. So, that error is that I denote as e o n corresponding to the optimal filter, which has the minimum variance.

But you have got since w is not w transpose, but always there is a deviation delta n, whose mean only is zero, but which is never actually zero mean only is zero. You will have another component and that component this delta transpose n delta transpose n x n. As a result if you compute this quantity you take the whole square of this. There is even now finding out the variance of the error if you take the wholes square of that one will be of course. So, this is that minimum mean square there will be another term e 0 n into this a minus b whole square; a square minus 2 a plus b square, but e 0 n is a scalar.

This is a row vector or column vector this is a scalar that times e zero n. So, I can write u n to the right also. First, we can write delta transpose n x n this quantity followed by e 0 n and there is one more term e of square of this guy square of this guy. I am whole squaring a minus b whole square. This quantity is the best one, what is that minimum error variance at a level because that correspondence to the optimal weight. This is the error when you are really putting the optimal weight vector w of then so; that means, this is that minimum mean square error.

We can call it epsilon square mean, you cannot minimize you cannot have any error variance less than this minus this quantity plus this quantity this is as it is now look at this quantity; delta transpose n x n en e 0 n. If I use the delta transpose n depends on w n

x n and e 0 n depends on what, d n and x n e zero n depends on d n and x n. Now, if I apply that independence assumption on this side I have got a quantity depending on w n. These sides I have got quantity depend on x n on which w n is independent by assumption and here also x n and d n. On which also w n is independent so; that means, this delta transpose n is independent of this part.

So that means, this quantity will be just one over minute I am through E delta transpose n into E x n e o n and I can say this is this correlation between the components of x n and u n. Remember correlation between the components of x n and e o n that correlation is zero. You remember orthogonality I said I had proved also that this error for the corresponding to the optimal filter. That is orthogonal to each component of x n. So, the correlation is zero. So, this will become zero vector, which means this terms will be zero. So, we left with this and this. So, I will start from here in the next class.

Thank you very much. Next class is on thursday so that is all. Thank you very much.