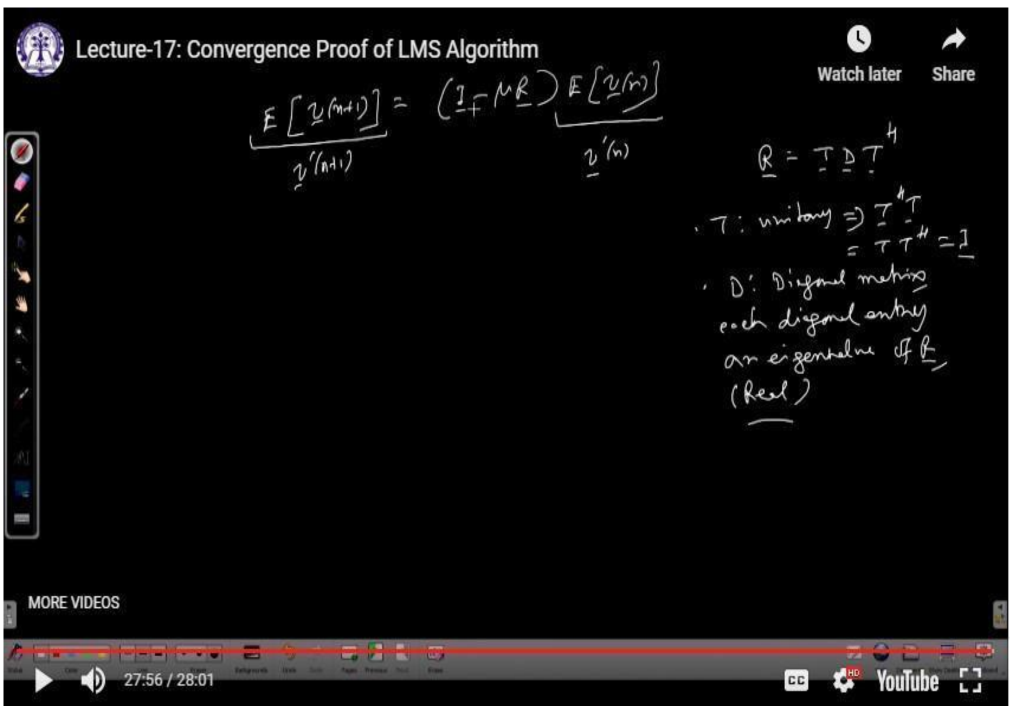
Introduction To Adaptive Signal Processing
Prof. Mrityunjoy Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture No # 18

Convergence Proof of LMS Algorithm (Contd.)

So, last class we did here.



And this E of this vector E of v n I called v prime n which is not random, but a vector which is function of n. So, this evolves like this from v prime n v prime n plus 1 comes all right. Now, I replace R by this. So, I will have identity matrix minus mu TDTH v prime n all right. Identity matrix I can write as TI TH, i into TH is TH, TTH is identity because it is unitary ok.

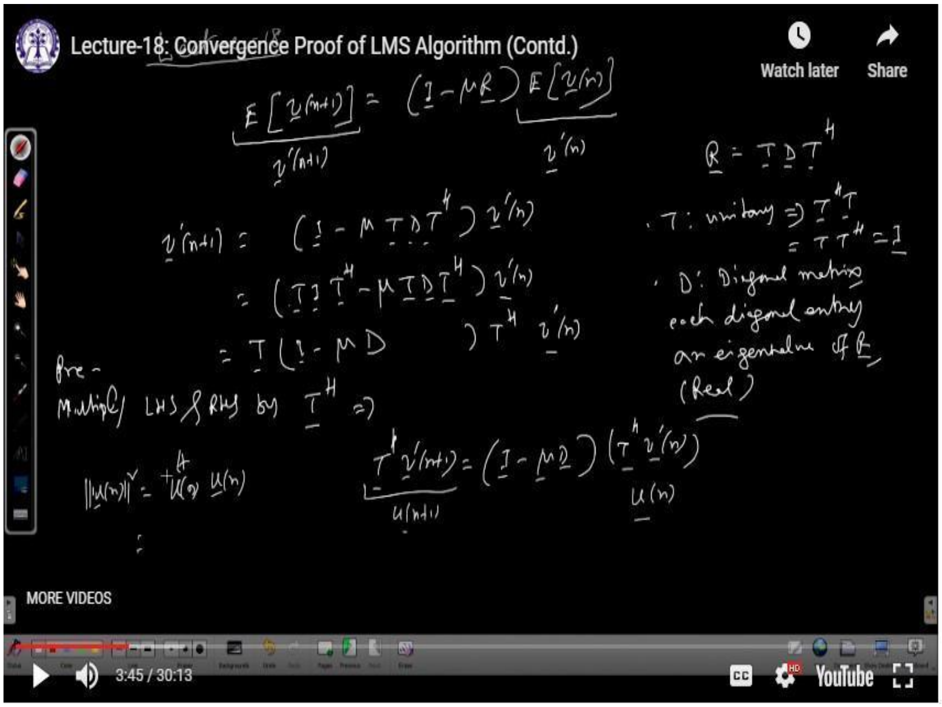
$$\underline{R} = \underline{T}\,\underline{D}\,\underline{T}^H$$

$$\underline{T}^H\underline{T} = \underline{T}\,\underline{T}^H = \underline{I}$$

I take this T common outside T and TH on the right-hand side common, TH TH that goes to the right-hand side, TT comes to the front. So, still i minus mu times d and then v prime n, v prime n. So, I think TI TH that is what I have here then minus mu TD TH all right. And then now I multiply LHS and RHS is called pre multiply this multiply here pre multiply by TH.

So, TH times this is TH times this, but TH T is identity that means, TH is i minus mu d again TH v prime n. So, now, this TH v prime n I call it u n just give it m. So, this is nothing, but TH v prime n plus 1. So, it is u n plus 1. Remember one thing if I see the norm square of u n that is un vector if I take a v element square up each element and then add ok there is a norm square square of the length.

So, that is like u n transpose with u n, but un is T Hermitian v prime n right. So, I have if I put that here and transpose of that, In fact, instead of transpose I can also write it h though they are all real. So, h and t will be same here you know because there is nothing complex. So, Hermitian transpose is ordinary transposition, but since I am following h here let me be consistent.



So, let me put h n u h n un. So, if this is un u hn will be h on this. So, v prime h n v prime hn TH h which is T and then again u n. So, you have directly TH and T TH is identity this identity. So, it is nothing, but v prime h v prime which is nothing, but v prime n norm square.

$$\underline{v}^{'}(n+1) = \left(\underline{I} - \mu \underline{T}\,\underline{D}\,\underline{T}^{H}\right)\underline{v}^{'}(n)$$

$$= \left(\underline{T}\,\underline{I}\,\underline{T}^{H} - \mu\,\underline{T}\,\underline{D}\,\underline{T}^{H}\right)\underline{v}^{'}(n)$$

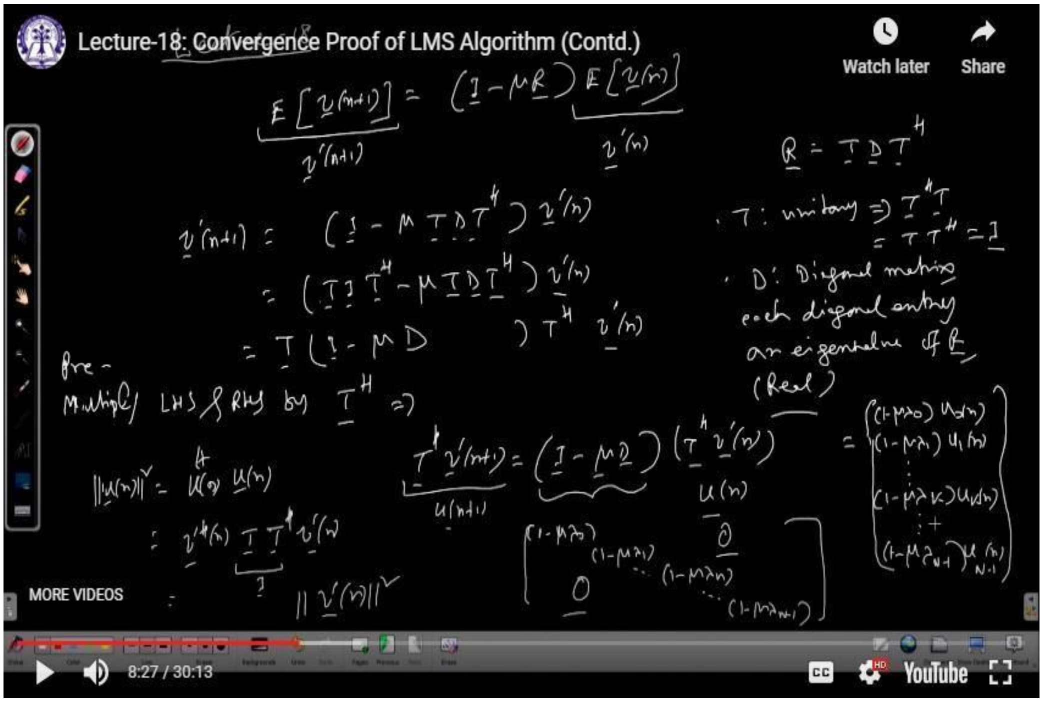
$$= \underline{T}\,(\underline{I} - \mu D)\underline{T}^{H}\,\underline{v}'(n)$$

So, you see un and v prime n they have the same length same power that is same norm square. If one is obtained from the other by pre multiplying by a unitary matrix that is why I say unitary matrix is preserve norms. Norm of the original vector there is length of the original vector and after the transformation by TH, these vectors they are the same. Norm square of un and norm square of v prime n they are same. Therefore, if I can show that with time, this norm square of un this goes down to 0 that will also mean norm square of v prime n will go down to 0 ok that is how I will proceed and norm square of v prime n v prime n is this much.

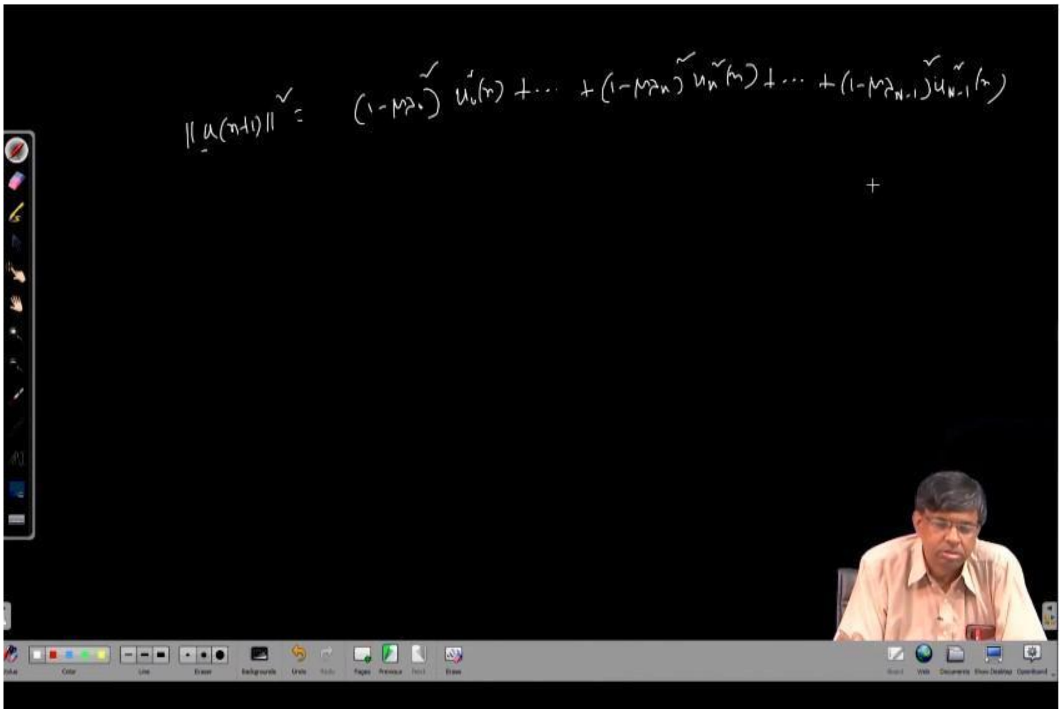
So, and norm square going to 0 means norm square of a vector 0 this vector itself is 0 that I told several times earlier any vector its norm square if it is 0 that means, only possibility vector itself is 0. So, if this vector norm square goes to 0 that means, as n tends to infinity if I can show that that will mean this vector itself goes to 0 as n tends to infinity which will give rise to convergence that is this error convergence in mean there is error vector after expectation that is after mean will become 0 as n tends to infinity. So, and error vector 0 means w opt and w n they will be together they will be same then only the error between them is 0 which is v n this is what I will show ok. So, I will show norm square of un goes down to 0 as n tends to infinity which means norm square of v prime n goes down to 0 as n tends to infinity. Norm square of a vector going down to 0 means vector itself goes down to 0 as n tends to infinity and this is v prime n.

So, E of this goes down to 0 as n tends to infinity and E of this is going down to 0 means v n is w n minus w opt that difference that is E of w n minus E of w opt w opt is constant. So, E of w f w n minus w opt that goes to 0 means E of w n goes to w opt I will show that. This is my game all right. And one more thing you see I minus mu d I is a diagonal matrix with ones in the diagonal places, mu into d d is a diagonal matrix with the eigenvalues ok. So, together is a diagonal matrix. So, it will be like 1 minus mu lambda 0, 1 minus mu lambda 1, dot dot dot 1 minus mu lambda k, dot dot dot 1 minus mu lambda n minus 1, all lambdas are real this side all 0 this side all 0 these are the matrices, this matrix multiplies u n.

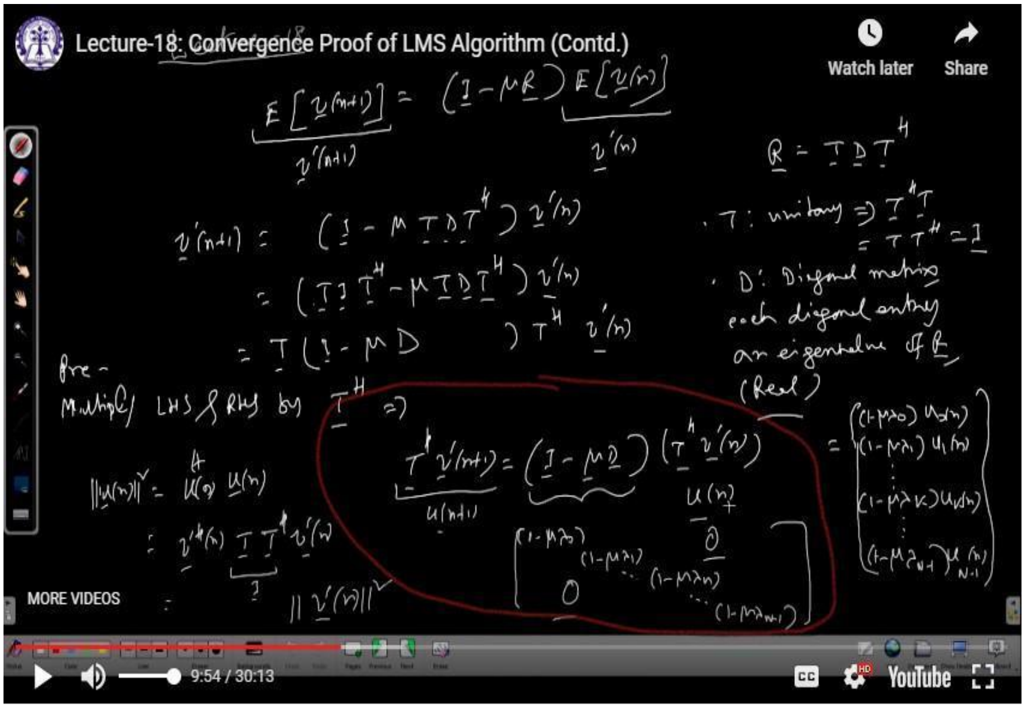
So, top guy of u n call it u 0 n that will be multiplied by this only is a diagonal matrix next guy u 1 n, that will be multiplied by this guy only likewise ok. That means, it will be like 1 minus mu lambda 0 u 0, top guy u 0 n 1 minus mu lambda 1 u 1 n, dot dot dot dot lambda k u k n dot dot dot 1 minus mu lambda n minus 1 u. So, u n plus 1 is just this. So, norm square of u n plus 1 will be norm square of this, but norm square of this is square of this term plus square of this term plus dot dot dot square of this term plus dot dot dot square of this term this is what.



I will write in the next page. So, norm square of will be 1 minus mu lambda 0 square 1 minus mu lambda square u square u 0 square n and likewise plus dot dot dot dot in general 1 minus mu lambda k square u k square n, plus dot dot dot dot 1 minus mu lambda.



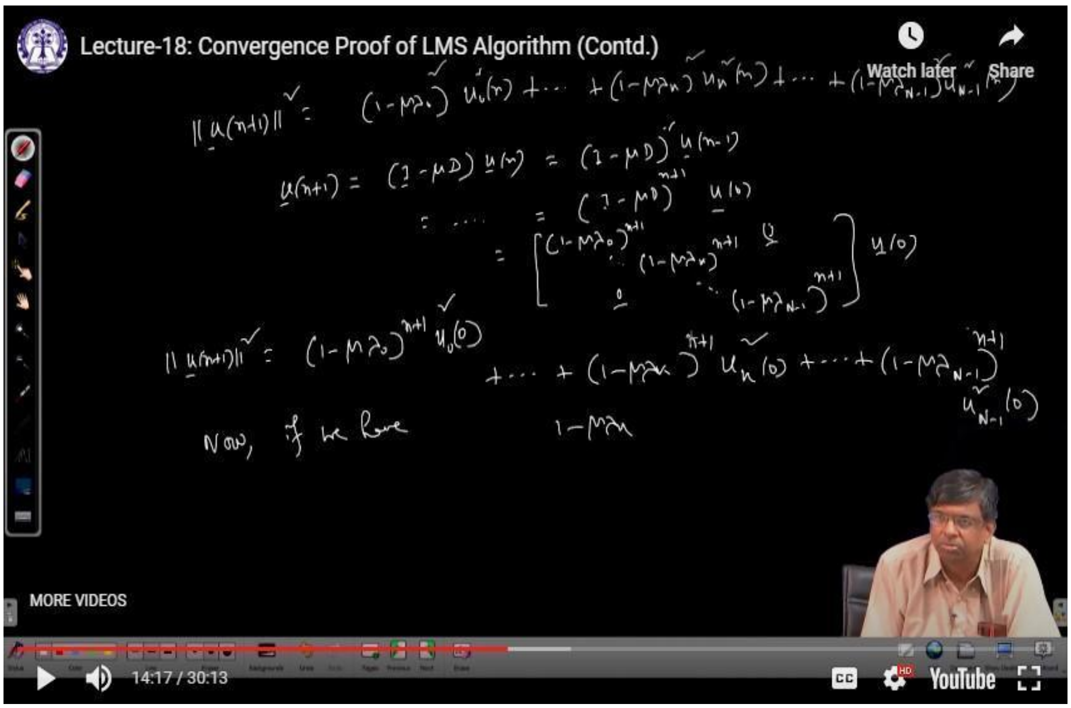
Now ok this also means, but I will use this expression little later let me again sorry. So, this is my i minus mu d into u n this u n itself I can write as u n like instead of n plus 1 if I write n. So, it will be again same matrix i minus mu d into u n minus 1 again u n minus 1 is i minus mu d u n minus 2 and like that.



So, that way what I had is u n you write as because this is independent of n. So, u n again if I replace n plus 1 by n it will be i minus mu d u n minus 1.

So, i minus mu d u n minus 1. So, square u n minus 1 and dot dot dot dot if you go on doing it finally, i minus mu d square cube like that I write here what it is, but it will be u 0 u 0 I will stop. So, if it is n minus 1 it is 2, it is if it is 0 means n minus n. So, it will be n plus 1, n plus 1 this is what and u 0 is a constant u 0 because what is what is u 0, u 0 is what is u n after all this is u n. So, u 0 is T h v prime n and what is v prime n u of v 0 sorry T h v prime 0, u 0 is T h v prime 0 and what is v prime 0 u of v 0, but v 0 is w 0 minus w of this is an initial condition which is not random this is not random.

So, v 0 is not random e of that is itself that is why there is a constant. So, u 0 is a constant this time this this means diagonal matrix into diagonal matrix into diagonal matrix that will be a diagonal matrix right. So, it will be 1 minus mu lambda 0 n plus 1 times we are going on multiplying I minus vd, I minus vd, I minus vd, I mean all diagonal I do with this then 1 minus mu lambda k whole to the power n plus 1, 1 minus mu lambda mu 0. So, now, like this we can write as earlier I had the square now it is this is the only difference n plus 1 times the 0th component of u 0 square, dot dot dot dot lambda k to the power n plus 1 again kth component square and dot dot dot dot 1 minus mu lambda n minus 1 to the power n plus 1 the last component that is competitive with you there is this. Now if we have 1 minus mu lambda k its value if its magnitude lies between 0 to 1, then as we power it up n to n plus 1 to n plus 2 to n plus 3 n plus 4 this goes down, because it is less than 1 in magnitude suppose it is half, but one-third.



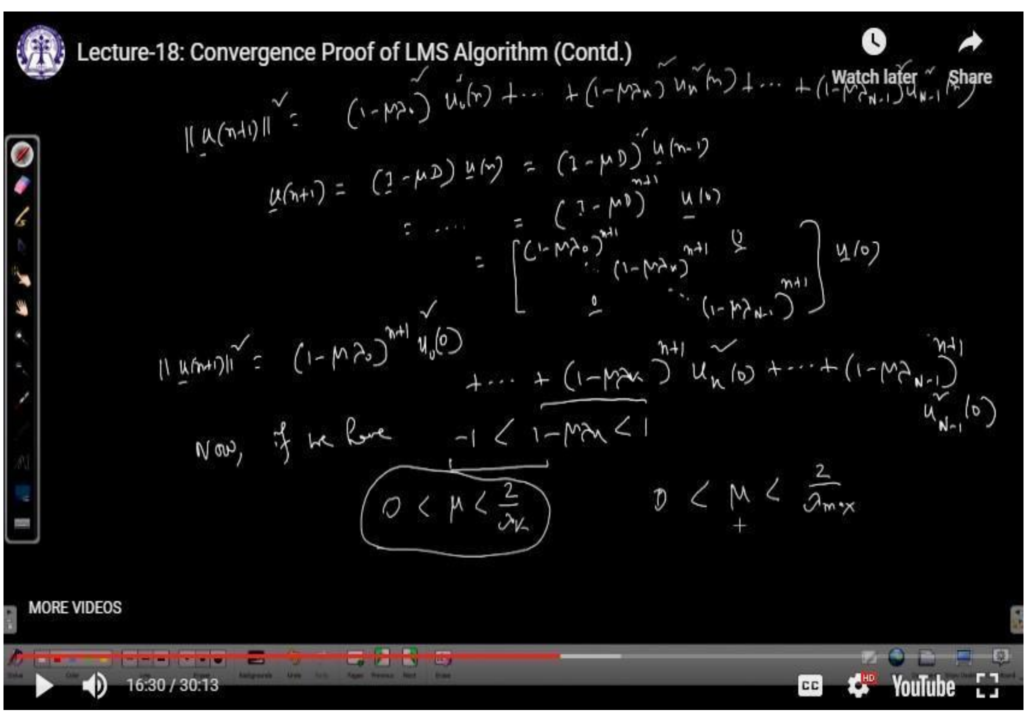
So, one-third to the power n plus 1 then one-third to the power n plus 2 then one-third to the power n plus 4 as we go on iteration value is decreasing decreasing and finally, it will go down to 0 and if all goes down to 0 this also will go down to 0. So, strong condition sufficient condition for these to go down to 0 will be if everybody has a magnitude between 0 to 1 that means, its value is between 1 to minus 1 ok. It can be minus 0.3 also or plus 0.3 also does not matter magnitude should be less than 1, then only as we go on doing it its magnitude finally, will go down to 0 ok. So, this should be true for all lambda k. This means if I take this part then you take this to this side minus 1 to this side so that means, mu less than 2 by lambda k and if I take this inequality 1 1 cancels.

$$\underline{u}(n+1) = (\underline{I} - \mu D)\underline{u}(n) = (I - \mu D)^{r}\underline{u}(n-1)$$

$$= (I - \mu D)^{n+1}\underline{u}(n)$$

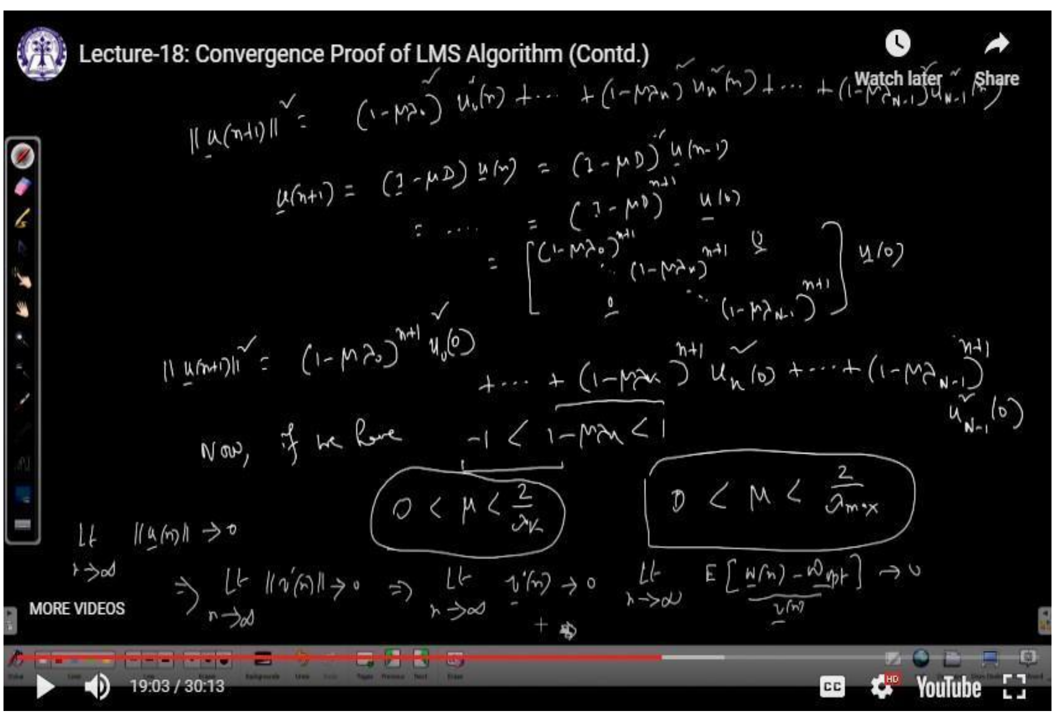
So, minus mu lambda k less than 0 mu lambda k greater than 0 lambda k is not only real it is positive because these are eigenvalue of the input autocorrelation matrix which is assumed to be positive definite. So, if that is positive mu must be positive. So, that is what then only this greater than 0. So, these are range now 2 by lambda 0, 2 by lambda 1. So, minimum is 2 by lambda max, lambda max means the eigenvalue is the highest magnitude all are positive.

So, highest positive value then 2 by lambda max is the minimum. So, mu should satisfy that also then it will be satisfied mu should be will be satisfied mu less than 2 by lambda k will be satisfied for other eigenvalues also if this is satisfied because this is lambda max. So, 2 by lambda max is the minimum value of this ok. For any other lambda k 2 by lambda k will be larger than this. So, if mu is less than 2 by lambda max mu is automatically less than 2 by lambda k.



So, this is the condition we have for convergence ok. This is a condition of convergence, but convergence means convergence in mean ok convergence is mean that is what I find from here limit n tending to infinity u n vector is norm or norm square goes down to 0 which means since norm 0 means vector itself is 0 it means limit n tending to infinity u n vector itself goes to 0. 0 vector there is a scalar norm the vector, but what was u n, u n from the previous page we can see what was u n, u n is th v prime n ok. So, one minute u n is th. So, both have same norm u n and v prime n they have the same norm and that is what. So, I should use that it is v prime you see here un and v prime n they have the same norm.

So, if this norm goes to 0 this norm goes to 0 this implies limit goes to 0 and then I say if the norm is 0 vector itself goes to 0 ok. There is only way a norm can be 0. So, this means limit n tending to infinity v prime n goes to 0, but v prime n is in the previous page you have seen this v prime n e of v n and v n is w n minus w opt. So, limit e of v n and v n is w n minus w opt this is your v n this goes to 0 which means e of w n minus e of w opt w opt is constant c of w opt itself.



That means next page I go e of w n limit n tending to infinity is w opt.

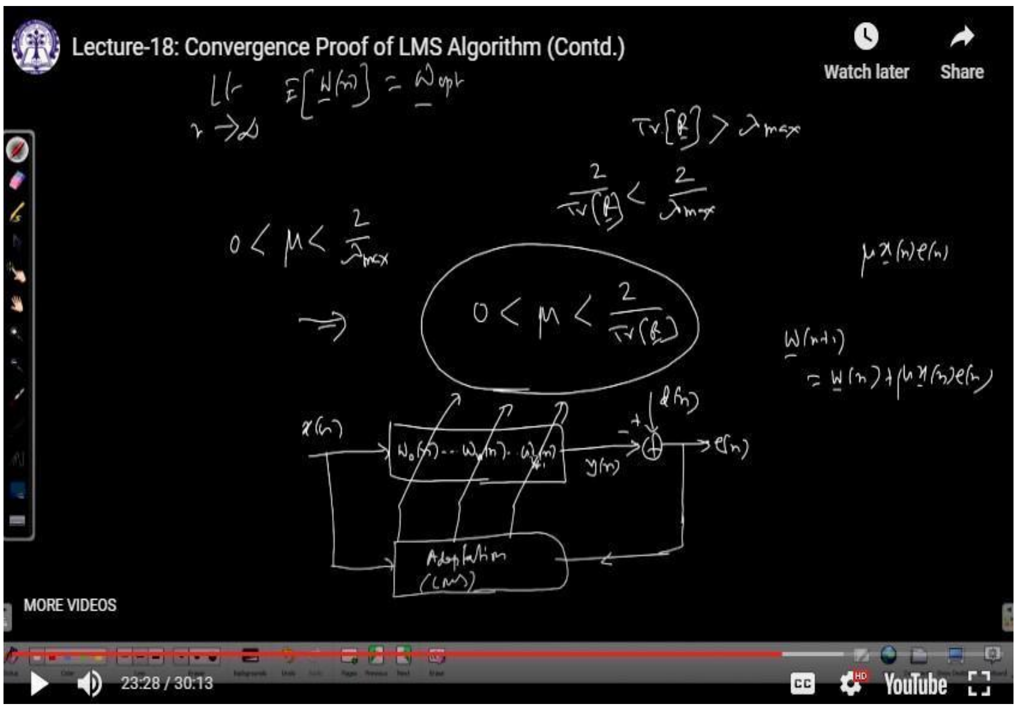
That means expected value of each weight converges on the corresponding optimal value. There is a that is what I had shown mean of the fluctuation converges on the corresponding optimal value, but not the coefficient itself ok. One more thing we know trace, trace is a sum of diagonal elements we have seen we also know if I give you r which is a positive definite matrix, trace of r is a summation of all the eigenvalues right. So, trace of r this should be greater than or equal to lambda max, because trace of r means lambda max plus summation of other eigenvalues ok. So, it is greater than equal to lambda max. In fact, if it is strictly positive definite other eigenvalues also are positive, lambda max plus other nobody can be 0.

So, it is actually strictly greater than all right. So, under that assumption let me make it strictly greater than which means 2 by lambda max is greater than 2 by trace r. So, if I our condition was this 2 by lambda max, but you know finding eigenvalue and all that is cumbersome in practice trace of r is very easy the summation of the diagonal elements no computation of eigenvalue and all that. So, if mu is less than 2 by trace r mu is automatically less than 2 by lambda max. So, that is why we modify it to this mu less than 2 by trace r.

If this happens if this happens it will always converge in mean convergence in mean is guaranteed ok. This was also fine, but this requires a bit impractical because it requires computation of eigenvalues and all that that is for big matrices is very difficult cumbersome actually. So, this you can do very quickly. So, this is what is usually followed mu greater than 0 less than 2 by trace r all right. So, this is how I have given the proof of LMS algorithm and LMS algorithm you know I mean schematically is shown like this xn W0n dot dot dot Wkn is yn.

So, there is an adaptation part LMS that takes the input data because you know the adaptive part the update part was mu xn vector en. So, xn vector has x of small n plus pass terms pass terms are already in the system. So, current term has to be taken. So, adaptation takes that vector then has that vector and it needs en, this is my en this also goes in and it calculates that update formula update equation by that it calculates Wn plus 1. So, once it is updated it changes them for use in the next clock cycle and this is how it is indicated tuning training.

So, this is a schematic of LMS based adaptive filter all right.



Now going instead of going further on the theory and other forms and all that I will take up examples several examples you know application examples all right of adaptive filters. This application examples can be classified into two type for first is forward modeling. Suppose I have got an unknown system modeled as LTI linear time invariant system linear shift invariant system such modeling of a requires the filter order even if I have filter to be large by a large order if I have filter you can model an unknown system ok. So, the coefficients are h0 h1 dot dot dot maybe you know h n minus 1.

Then I have to identify them estimate them and they may be varying with time from time to time. So, I should be able to track them that is a model forward modeling ok. So, when you observe I mean you give a WSS input to the plant that you can generate in the lab it is not a problem WSS input to this ok. Output when you observe as this input goes through the system at various points some noise gets added. So, all the noise effects, you can combine together add as a separate noise component z n and this is y n. I should not use the term y n let me use it let me use something maybe v n and this is what you observe.

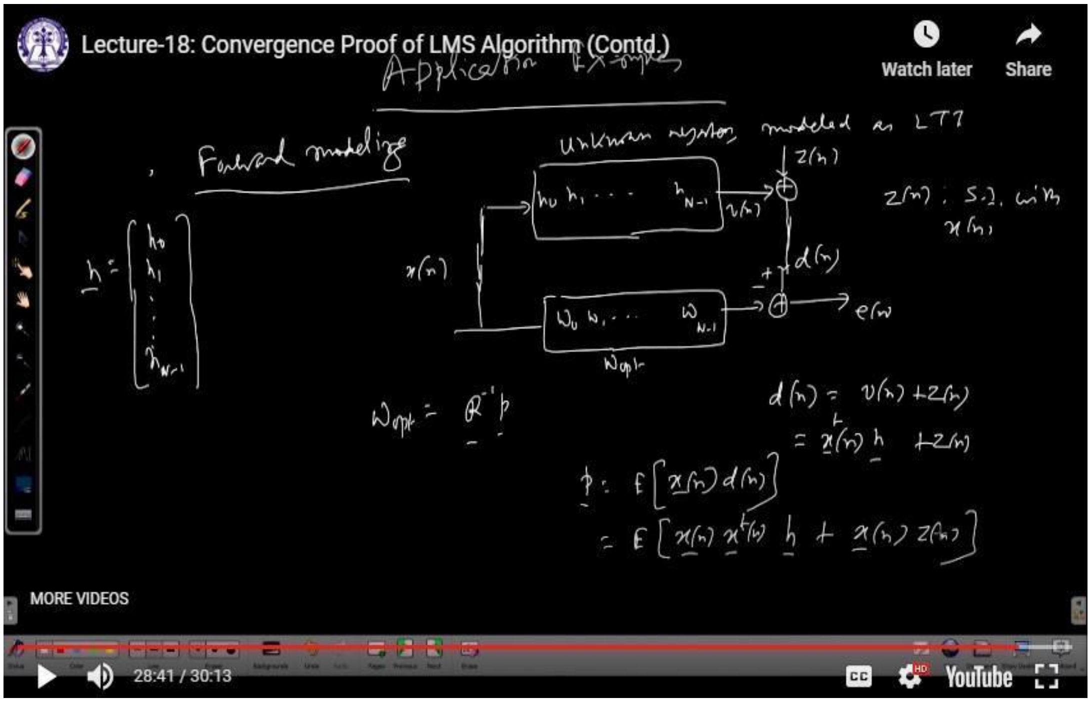
So, plant output when you observe you do not get the pure raw v n you get a noisy version v n ok. Now, suppose my question is you want an optimal filter W op it has W0 W1 dot dot dot W n minus 1. It is assumed x n is 0 mean and therefore, v n is 0 mean z n is I statistically independent with input very natural is a noise generated in inside the system it has nothing to do with the input. So, you can assume that to be independent of each other all right. So, if you want it to be if you want this to be an optimal filter.

So, that output error minimum I mean variance is minimized the W op will be what we do the formula R inverse p right R inverse p, p is the autocorrelation between x n vector and d n and R is the input autocorrelation. Now, what is d n? d n is v n plus z n and what is v n? v n is x n transpose H and this plus z n what is H? H is the true system coefficient vector I do not know them I have to find them true system coefficient vectors. So, linear convolution so, either x n transpose H n x n transpose H or H transpose x n they are same. So, I write it as x n transpose H that is v n plus z n that is my d n and z n is statistical also 0 mean and statistical in the with x n. So, p will be what? E of x n d n and now replace d n by this.

$$\underline{p} = E[\underline{x}(n)d(n)]$$

$$= E[\underline{x}(n)\underline{x}^{t}(n)\underline{h} + \underline{x}(n)z(n)]$$

So, x n x transpose n H plus x n there it is a scalar. So, I should not put this. So, x n into this part I have done x n into this part and x n into z n all right.



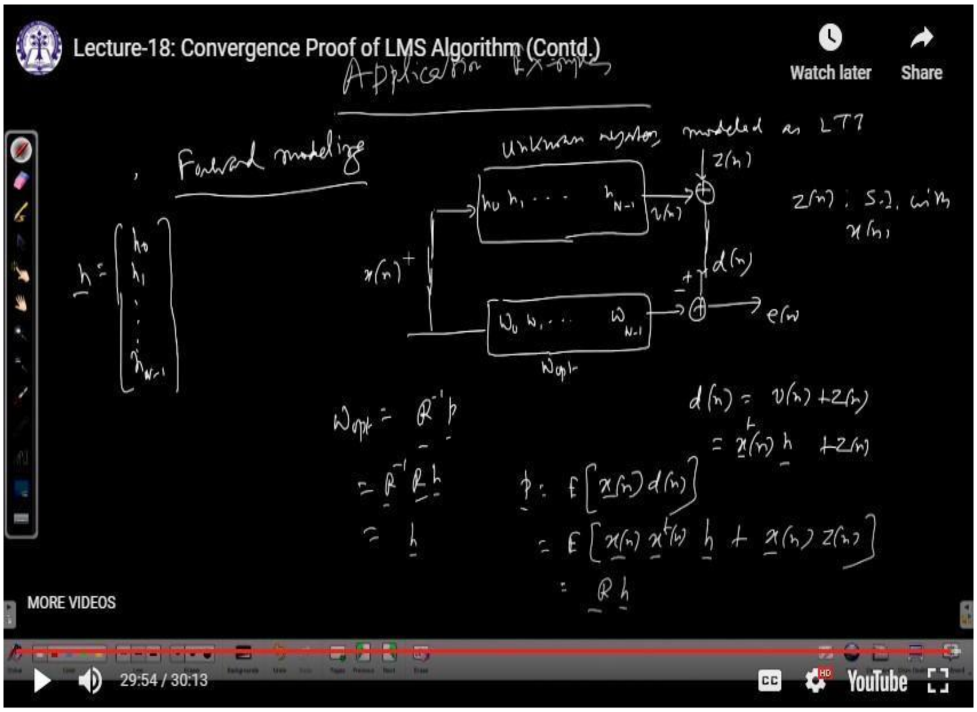
Now, all the elements of x n and elements of element of z n to z n they are statistically independent. So, expected value of a product will be expected value of this into expected value of this each is 0 mean.

So, if you take expected value of this vector times z n every element will be after expectation will be 0 because they are statistically independent. So, E of E on a product of any term of x n into z n will be E of that x n term into E of z n term and each is 0 because they are 0. So, that goes to 0 and E on this part and H is not random. So, it will be E on this into H this is R. So, this is your RH R is the autocorrelation matrix which means in this case the optimal filter will be R inverse RH and it can say H which means if I know the value of capital N correctly and calculate the optimal filter in this case using this as d n plant output noisy plant output as d n and input x n that optimal filter will be the actual true system parameters.

$$w_{opt} = \underline{R}^{-1}\underline{p}$$

$$= \underline{R}^{-1}\underline{R}\,\underline{h}$$

$$= \underline{h}$$



This is called forward modeling ok, because I am going, I am trying I mean a input is going in the forward direction and that is what I am modeling ok. This is the basic scheme and this scheme has been used in several context which we will consider ok, forward modeling and some applications. So, that will be in the next class. Thank you very much.