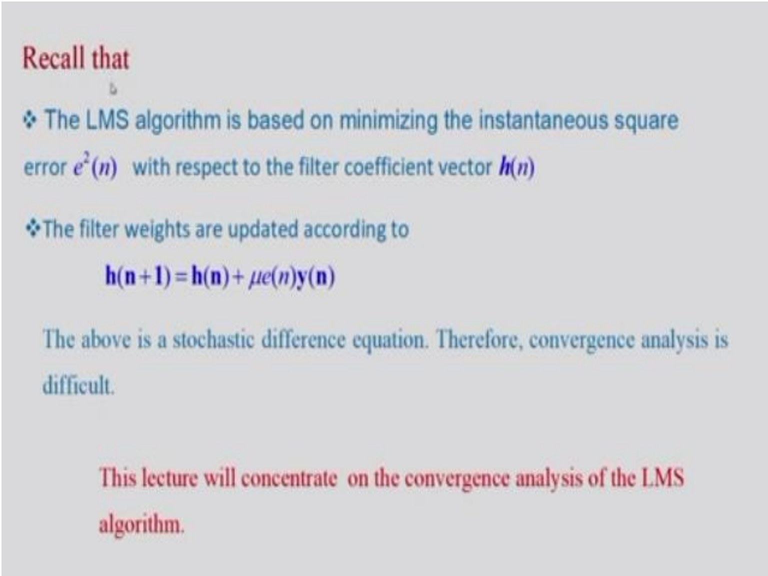
Statistical Signal Processing
Prof. Prabin Kumar Bora
Department of Electronics and Electrical Engineering
Indian Institute of Technology – Guwahati

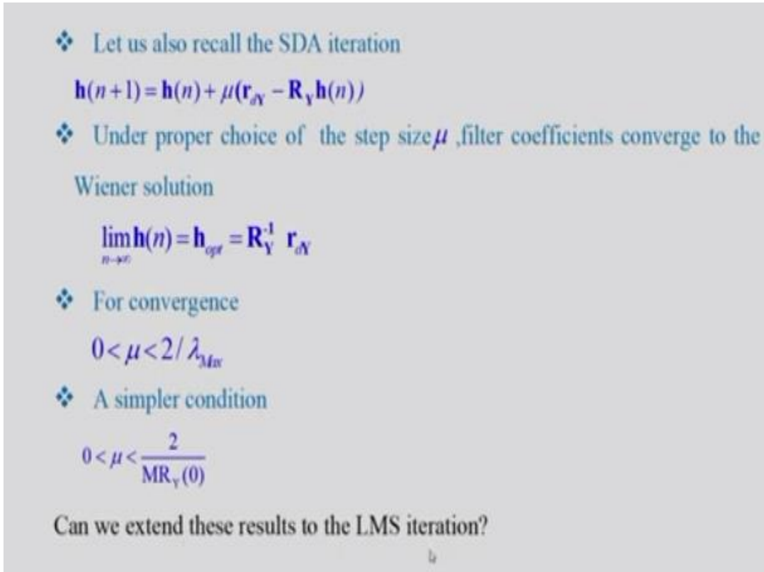
Lecture - 27
Adaptive Filters 3

(Refer Slide Time: 00:45)



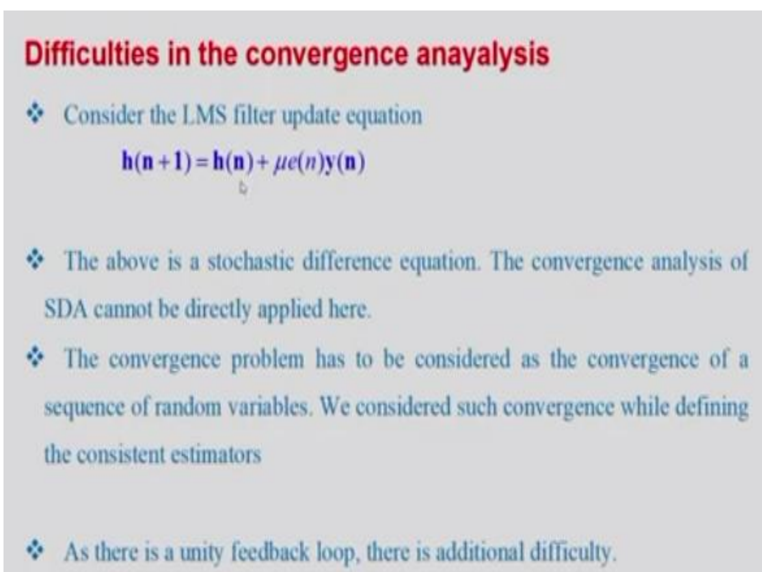
Hello students. Welcome to this lecture on adaptive filters. Recall that the LMS algorithm is based on minimizing the instantaneous square error e square n with respect to the filter coefficient vector hn. The filter weights are updated according to this equation h of n + 1 is equal to hn plus mu times en into yn, yn is the data vector, mu is this step length parameter. The above is a stochastic difference equation. Therefore, convergence analysis is difficult. Our knowledge of deterministic convergence cannot be applied here. This lecture will concentrate on the convergence analysis of the LMS algorithm.

(Refer Slide Time: 01:42)



Let us recall the SDA iteration h of n + 1 is equal to hn + mu times Rdy – Ry into hn vector, where Rdy is the cross-correlation vector and Ry is the autocorrelation matrix. Under proper choice of step size mu, the filter coefficients converge to the Wiener solution, that is as n tends to infinity limit of hn is equal to h optimum and that is given by Ry inverse into cross correlation vector and for convergence is given by mu lies between 0 and 2 by lambda max and a simpler condition was mu lies between 0 and 2 by M into Ry0, where m is the length of the filter. Can we extend these results to the LMS situation? Let us try to answer this.

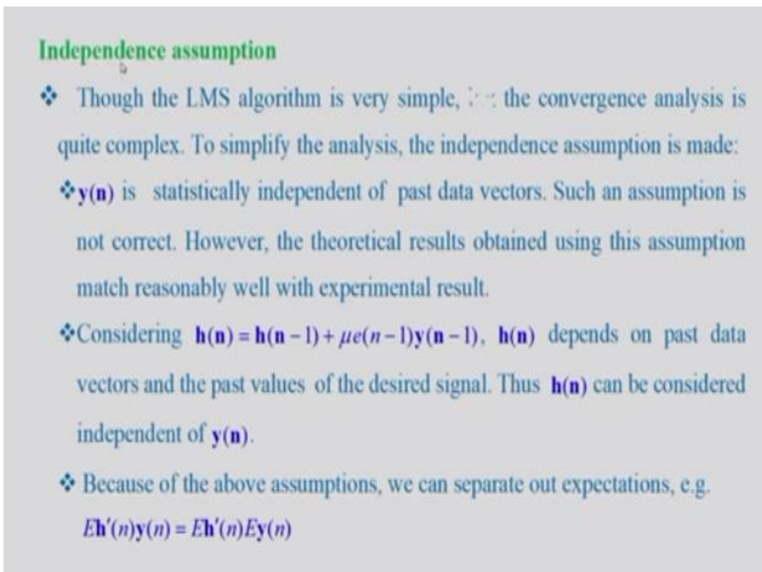
(Refer Slide Time: 02:55)



Consider the LMS filter update equation hn + 1 is equal to hn + mu times en into yn. As it is a stochastic difference equation the convergence analysis of SDA cannot be directly applied. So the answer to this question is that, the convergence analysis of SDA cannot be directly applied. The convergence problem has to be considered as the convergence of a sequence of random variable. This hn sequence is a sequence of random variables.

We consider such convergence while defining the consistent estimators. There are different modes of convergence for the sequence of random variables. So particularly we considered convergence in probability and mean square convergence. Here also, we will try to apply mean square convergence. As there is a unity feedback loop here, this coefficient here is 1. There is additional difficulty in the convergence analysis.

(Refer Slide Time: 04:20)



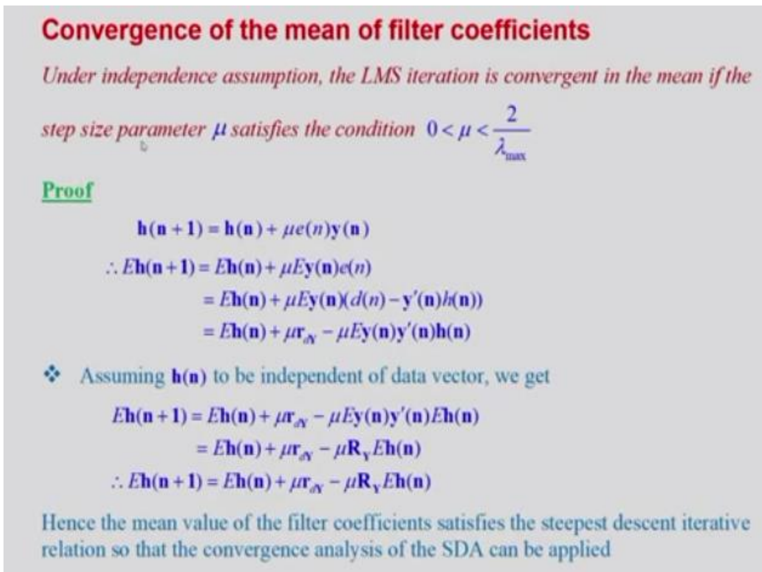
We have to make one important assumption, that is independence assumption. Though the LMS algorithm is very simple, the convergence analysis is quite complex. To simplify the analysis, the independence assumption is made. The data vector yn is statistically independent of past data vectors. This is the assumption. Such an assumption is not correct, because all data vectors will be dependent only.

However, the theoretical results obtained using this assumption match reasonably well with the experimental results. Now consider the filter update equation hn is equal to hn - 1 + mu En – 1 into yn – 1. Here hn depends on hn – 1, which is function of past data and it depends on E of n –

1 and yn – 1. So that way it is independent of yn, because yn is anyway independent of yn - 1 according to independence assumption.

Because of the above assumptions, we can separate out expectation. For example, we can write E of h transpose n into yn will be equal to E of h transpose n into E of yn. So this type of separation will be needed in our analysis.

(Refer Slide Time: 06:14)



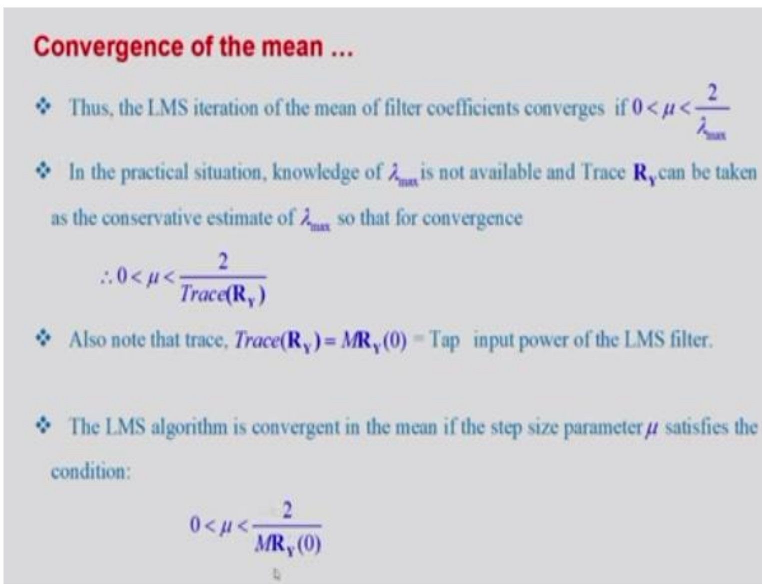
First, let us see the convergence of the mean of the filter coefficients. When iterations proceed, whether the mean of the filter coefficient is converged? Under the independence assumption, the LMS iteration is convergent in the mean, if the step size parameter mu satisfies the condition that is mu lies between 0 and 2 by lambda max. This is the same as the SDA condition. We will prove this h of n + 1 is equal to hn + mu times en into yn.

Therefore, expected value of e of hn + 1 is equal to E of hn + mu times expected value of yn into en. That is equal to E of hn + mu times E of yn into error is dn - y transpose n into hn. Now, I can separately take the expectation here. So therefore, the expression will be equal to E of hn plus this quantity will be Rdy, that is the cross correlation matrix and this part will be mu times E of yn into yn transpose into hn. Now, we will make the independence assumption.

Assuming hn to be independent of the data vector, we will get E of h sub n + 1 is equal to E of hn + mu into Rdy. Now here, we will write this entire expectation is product of two expectations, that is mu times E of yn, yn transpose into E of hn and this quantity is now E of yn, yn transpose is the Ry matrix. Therefore, this will be E of hn + mu into Rdy – mu times Ry into E of hn. Therefore, the average value of h of n + 1 is equal to average value of hn + mu Rdy - mu times Ry into E of hn, average value of hn.

So this is same as the iteration for the SDA algorithm. Hence, the mean value of the filter coefficients satisfies this steepest descent iterative relation, so that convergence analysis of the SDA can be applied.

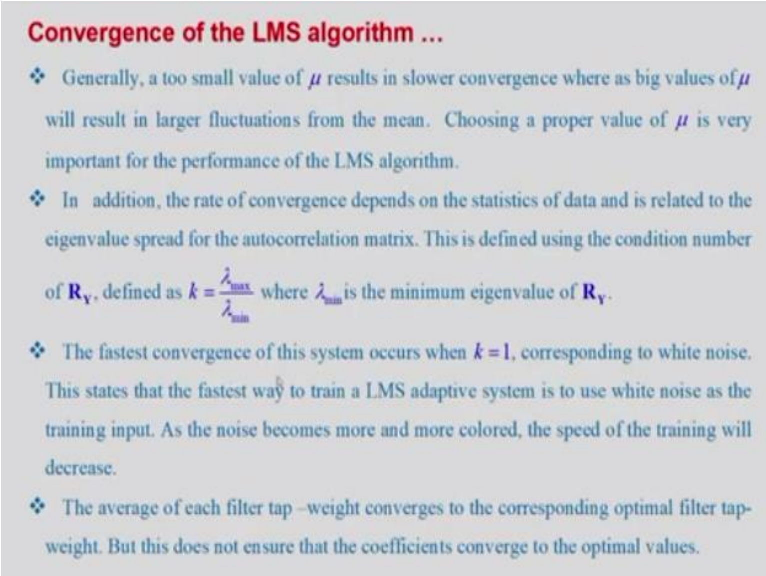
(Refer Slide Time: 09:19)



Thus, the LMS iteration of the mean of the filter coefficients converge, if mu lies between 0 and 2 by lambda max. So if we consider in terms of the mean of the filter coefficient, then the iterations will converge as long as mu is between 0 and 2 by lambda max. In practical situation, knowledge of lambda max is not available or it is difficult to obtain. Therefore, we can consider trace of Ry instead of lambda max, like in the SDA method.

So therefore, the conditions for convergence is mu should lie between 0 and 2 by trace of Ry and trace of Ry is equal to M times Ry0, that is the tap input power of the LMS filter. The LMS algorithm is convergent in the mean, if the step size parameter mu satisfies the condition mu lies between 0 and 2 by M times Ry0. So that way, we have to choose a value of mu, which is less than this quantity.

(Refer Slide Time: 10:41)

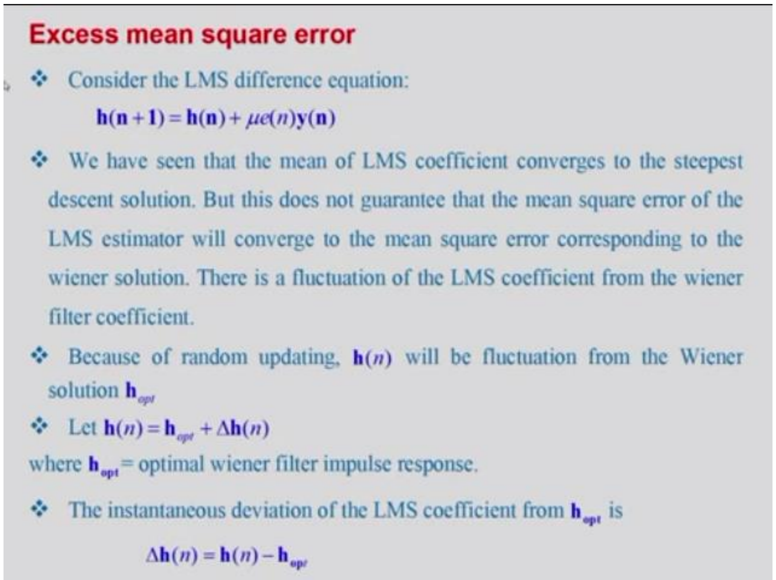


If we select a too small value of mu, then convergence will be slow; whereas a big value of mu will result in larger fluctuation from the mean. Choosing a proper value of mu is very important for the performance of the LMS algorithm. In addition to the above, the rate of convergence depends on the statistics of data and is related to the Eigenvalue spread of the autocorrelation matrix, that we have discussed earlier.

This is defined by the condition number of Ry, which is lambda max by lambda min. Therefore, rate of convergence will depend on k, it should be small. The fastest convergence of the system occurs when k is equal to 1 corresponding to the white noise. So this condition also implies that the fastest way to train an LMS adaptive system is to use white noise as the training input, because in the case of white noise all eigenvalues are equal.

Therefore, lambda max by lambda min will be equal to 1. As the noise becomes more and more coloured, the speed of training will decrease. The average of each filter tap weight converges to the corresponding optimal filter tap weight, but this does not ensure that the coefficients converge to the optimal value, because only average part is converging; we have to consider the fluctuating part also.

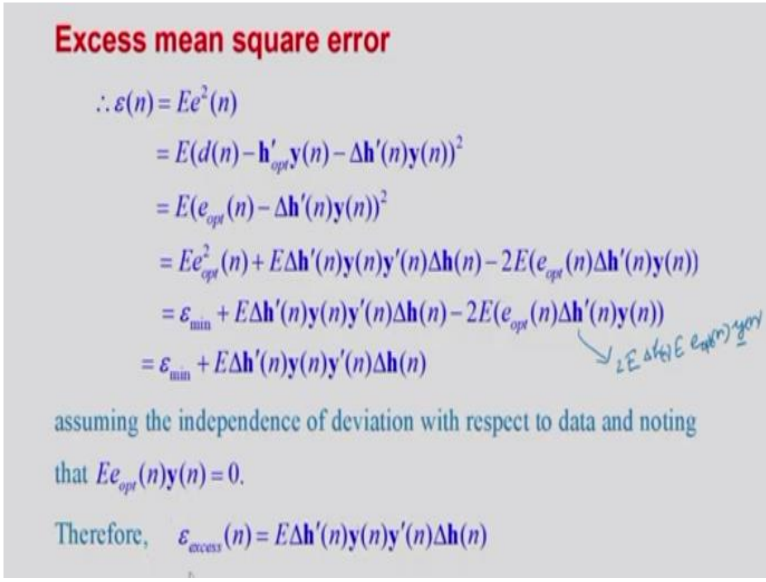
(Refer Slide Time: 12:38)



Let us consider the LMS difference equation h of n + 1 is equal to hn + mu en into yn. We have seen that the mean of LMS coefficient converges to this steepest descent solution. The mean will converge to the steepest descent solution. But this does not guarantee that the mean square error of the LMS estimator will converge to the mean square error corresponding to the Wiener solution. There is a fluctuation of the LMS coefficient from the Wiener filter coefficient.

Because of random updating, hn will be fluctuating from the Wiener solution h opt. Let hn be equal to h opt + delta hn, where delta hn is the fluctuation and h opt is the optimal Wiener filter impulse response. That is h opt is the optimal Wiener filter vector. The instantaneous deviation of the LMS coefficient from h opt is given by delta hn is equal to hn – h optimum.

(Refer Slide Time: 14:03)



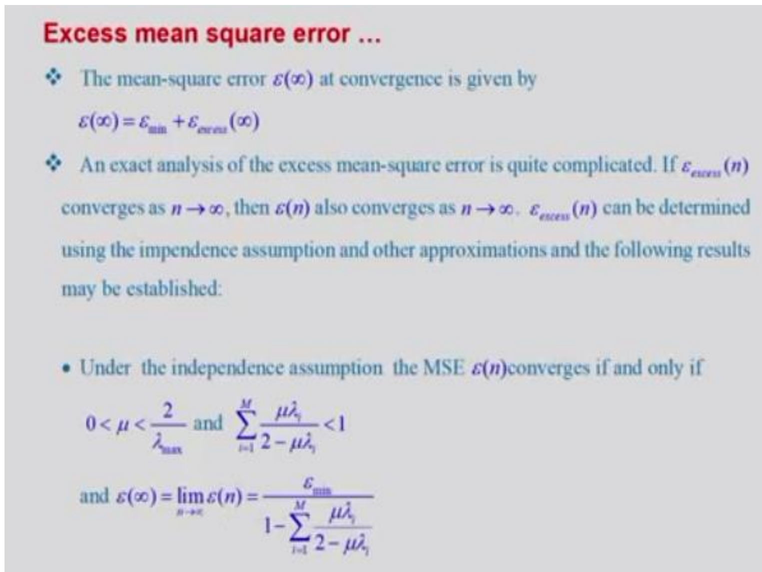
Therefore, mean square error, this is the mean square error, MSE is equal to E of h square and is equal to E of dn – h sub transpose into yn – delta hn transpose into yn whole square. This is the mean square error. Now this part is the error due to optimum Wiener filter. We denote it by e opt n. Therefore, mean square error will be e opt n – delta h transpose n into yn whole square and this we can write like E of square n, that is the Wiener mean square error epsilon min plus E of square of this quantity delta hn into yn, yn transpose into delta hn.

Because these are matrix, so this square term will come like this, minus the cost terms, that is twice E of E opt n and into delta h transpose n into yn. So that we will get this as that is mean square error of Wiener filter plus E of delta hn into yn, yn transpose into delta hn – 2 times e opt n into delta hn transpose into yn. So that way, we can write that the MSE is equal to Wiener filter mean square error plus E of delta hn yn into yn transpose into delta hn.

Because this term will become zero because of the independence assumption, we can write that this deviation. First, we can write this expression is equal to twice E of delta hn transpose into E of E opt n into yn vector. We can write like this because of the independence assumption, we can say that this is the deviation and this deviation is independent of data therefore we separate out the expectations. Now this part error is orthogonal to data.

Now this expression will become zero. Now mean square error will be simply epsilon min that is the mean square error corresponding to the Wiener filter plus this term. This is the excess mean square error. Therefore, there is an excess in the mean square error that is given by E of delta hn transpose into yn into yn transpose into delta hn. So how to analyze this excess mean square error.

(Refer Slide Time: 17:46)

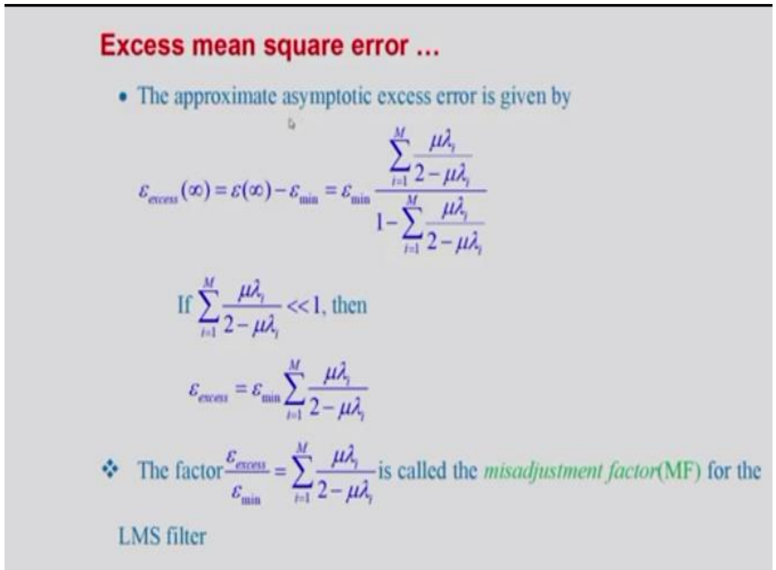


The mean square error that is epsilon infinity at convergence is given by epsilon infinity that is the mean square error as n tends to infinity is equal to epsilon mean, that is the mean square error corresponding to the Wiener filter plus epsilon access infinity. This is the excess mean square error and exact analysis of the excess mean square error is quite complicated. If excess mean square error converges as n tends to infinity, then n will also converge.

Because if this part converges, this part will be also converging, that excess mean square error can be determined using the independence assumption and other approximation and the following results may be established. We will not establish it, but this is a very important result. Under the independence assumption, the MSE epsilon n converges if and only if mu lies between 0 and 2 by lambda max and summation of, this is the additional condition summation mu lambda i divided by 2 – mu lambda i, i going from 1 to M is less than 1.

Not only this condition is to be satisfied, this condition also is to be satisfied and under that situation the mean square error that is equal to limit of epsilon n as n tends to infinity is equal to epsilon min that is the mean square error corresponding to Wiener filter divided by 1 minus summation mu lambda i – 2 – mu lambda i, i going from 1 to M. Therefore, this is the mean square error corresponding to the LMS Wiener filter.

(Refer Slide Time: 19:58)

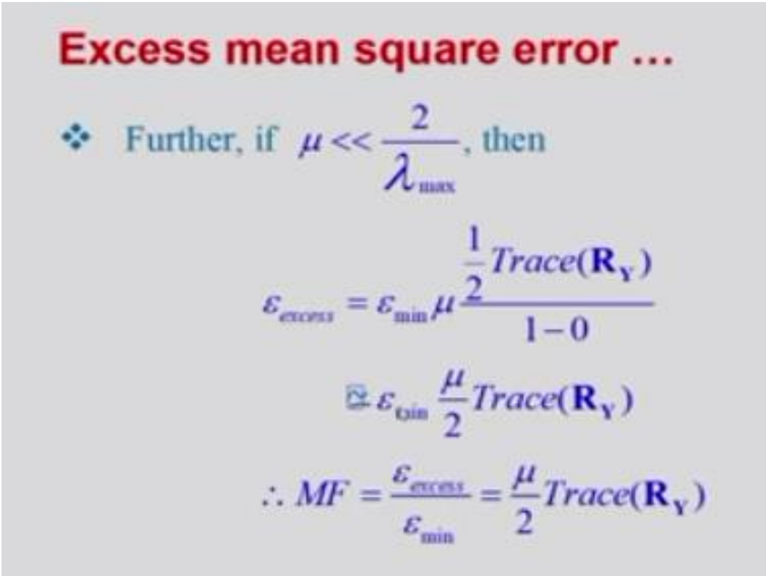


Now we can find out the approximate value of excess mean square error and that is epsilon excess infinity that is the excess mean square error at convergence is equal to epsilon infinity that is the mean square error minus the Wiener mean square error equal to epsilon min that is the Wiener filter mean square error multiplied by summation mu lambda i by 2 – mu lambda i, i going from 1 to M divided by 1 – summation mu lambda i divided by 2 – mu lambda i, i going from 1 to M.

We will make now some assumption, some simpler approximation. Suppose this quantity is much smaller than 1, then denominator will be simply 1, so that we can write epsilon excess that is the excess mean square error is equal to epsilon mean times summation mu lambda i divided by 2 – mu lambda i, i going from 1 to M. This epsilon excess divided by epsilon min, this factor which is equal to summation mu lambda i divided by 2 – mu lambda i, i going from 1 to M is called the misadjustment factor, MF for the LMS filter.

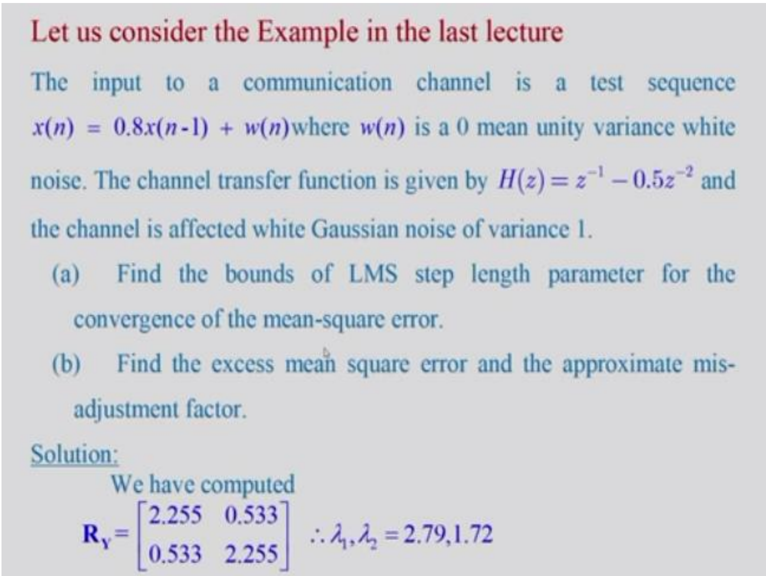
So there will be an excess amount in the mean square error by this factor. Further if mu is much less than 2 by lambda max, then excess mean square error can be approximated by this relationship, epsilon min into mu by 2 into trace of Ry, because here this term becomes 1 and this becomes summation lambda i and which is equal to trace of Ry matrix.

(Refer Slide Time: 22:15)



So that way, misadjustment factor will be now mu by 2 into trace of Ry. This is the misadjustment factor for the LMS filter.

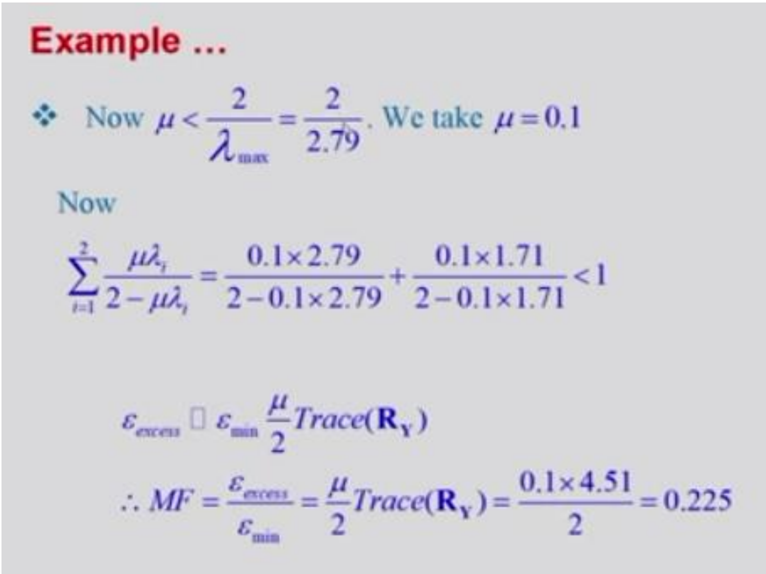
(Refer Slide Time: 22:27)



Let us consider the example in the last lecture. The input to a communication channel is a test sequence given by xn is equal to 0.8 xn – 1 + wn, where wn is a 0 mean unity variance white noise. The s channel transfer function is given by Hz is equal to z to the power -1 - 0.5 into z to the power -2 and the channel is affected by white Gaussian noise of variance 1. Now new questions are.

In addition to last time questions, find the bounds of LMS step length parameter for the convergence of the mean square error. Find the excess mean square error and the approximate misadjustment factor. Here Ry we have already found out that is equal to this matrix 2.255 here diagonal is 2.255 and up diagonal is 0.533, 0.533. Therefore, its Eigenvalue we can find out lambda 1 and lambda 2 are 2.79 and 1.72 respectively. This is the maximum Eigenvalue.

(Refer Slide Time: 23:49)

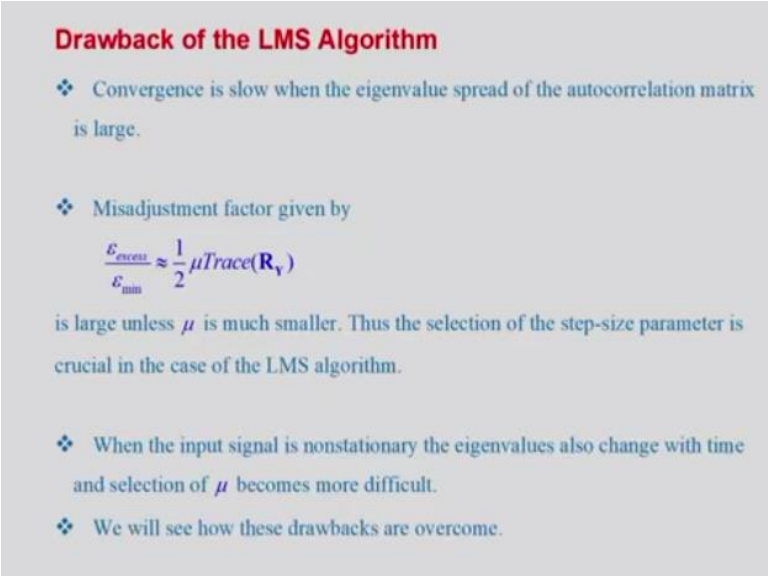


Mu should be less than 2 by lambda max and that is less than 2 by 2.79, whose mu should be small for excess mean square error. So we take mu is equal to 0.01. This is less than this quantity. Then, we see this factor that is we have to determine the excess mean square error. Therefore, this factor we have to determine summation mu lambda divided by 2 – mu lambda i, i going from 1 to 2. Here M is 2 and this we can find like this with mu we know.

Mu is 0.01, lambda 1 is 2.79 divided by 2 – 0.1 into 2.79. Similarly, mu lambda 2 is 1.71, therefore 0.1 into 1.71 divided by 2 – 0.1 into 1.71 and this factor is less than 1. Therefore, our choice of mu is right and excess mean square error, we can now determine. So excess mean square error will be determined by this expression. So epsilon mean, Wiener filter mean square error that we can determine from the Wiener filter solution.

And this is given trace of Ry we can determine, so we can determine the excess mean square error. Now the misadjustment factor is given by excess mean square error divided by the optimal mean square error and that is equal to mu by 2 into trace of Ry and when we put mu is equal to 0.1 and trace of Ry is equal to sum of the diagonal elements that is equal to 4.51 divided by 2 and that is equal to 0.225, that is there is a misadjustment of 22.5% from the optimal mean square error.

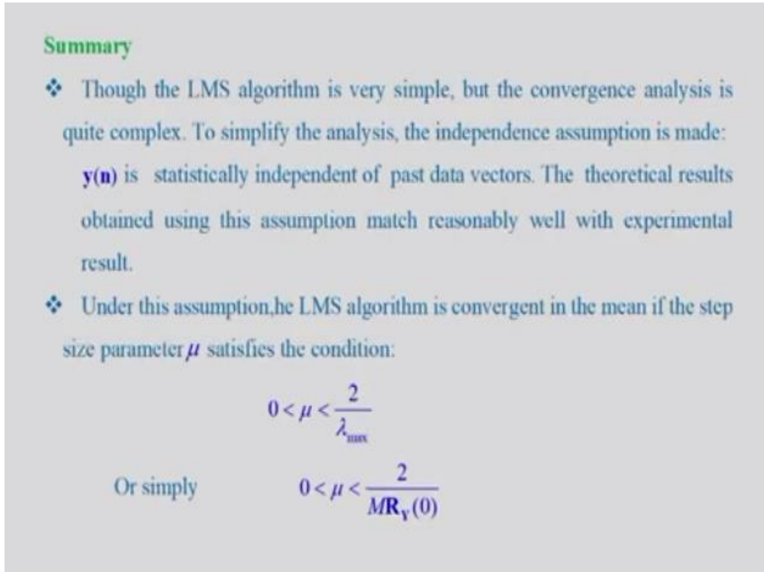
(Refer Slide Time: 26:12)



Let us discuss the drawback of the LMS algorithm. Here convergence is slow when the Eigenvalue spread of the autocorrelation matrix is large. That is one important factor. Misadjustment factor is given by that is excess mean square error divided by the Wiener mean square error, that is approximately equal to half of mu times trace of Ry and if mu is large, this quantity will be large. This error will be large unless mu is much smaller.

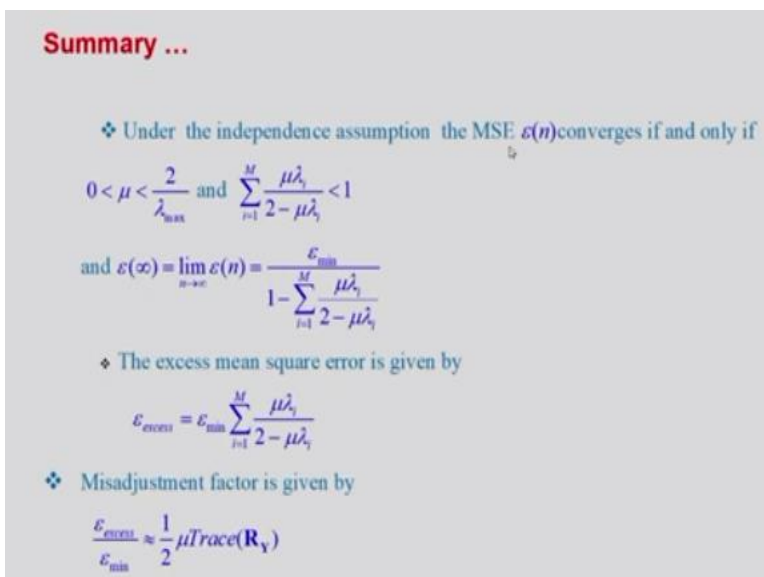
Thus the selection of step size parameter is crucial in the case of LMS algorithm. When the input signal is non-stationary the Eigenvalues also change with time and selection of mu becomes more difficult. We will see how these drawbacks are overcome in the next lecture.

(Refer Slide Time: 27:23)



Let us summarize the lecture. Though the LMS algorithm is very simple, the convergence analysis is quite complex. To simplify the analysis, the independence assumption is made, that is yn is statistically independent of past data vectors. The theoretical results obtained using this assumption match reasonably well with the experimental results. Under this assumption, the LMS algorithm is convergent in the mean if the step size parameter mu satisfies this condition, zero is less than mu is less than 2 where lambda max or simply mu lies between 0 and 2 by M into Ry0.

(Refer Slide Time: 28:18)



Under the independence assumption, the MSE epsilon n converges if and only if these two conditions are satisfied mu lies between 0 and 2 by lambda max and this additional condition is summation mu lambda i divided by 2 – mu lambda i, i going from 1 to M must be less than 1 and under this situation, the mean square error at infinity is given by epsilon min divided by 1 minus summation mu lambda i divided by 2 – mu lambda i, i going from 1 to M.

This is the Wiener filter mean square error. Now the excess mean square error is given by epsilon excess, this is the excess mean square error is equal to epsilon min times summation mu lambda i divided by 2 – mu lambda i, i going from 1 to M and the misadjustment factor is given by that is excess mean square error divided by epsilon min that is approximately equal to half of mu times traces of Ry. Thank you.