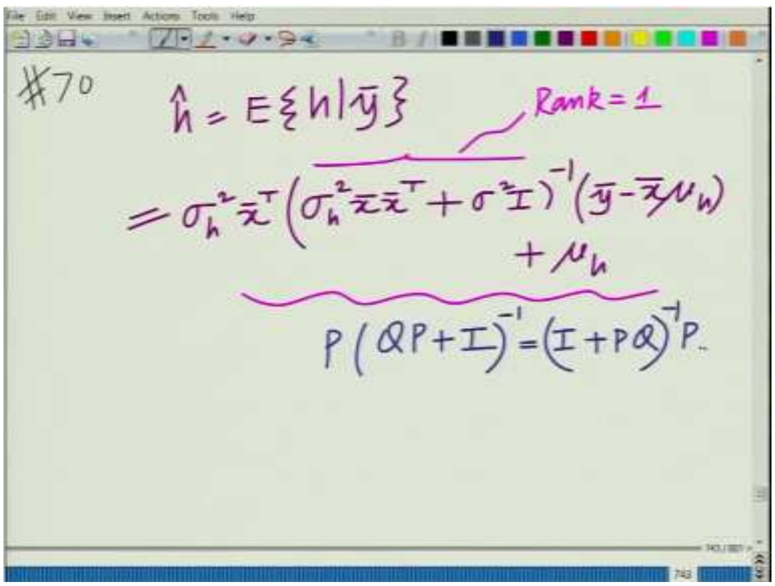
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Lecture No. 70
MMSE Estimate and Covariance for the Scalar Linear Model

Hello, welcome to another module in this massive open online course, so we are looking at the MMSE estimate of h given y bar, and we have written this as follows.

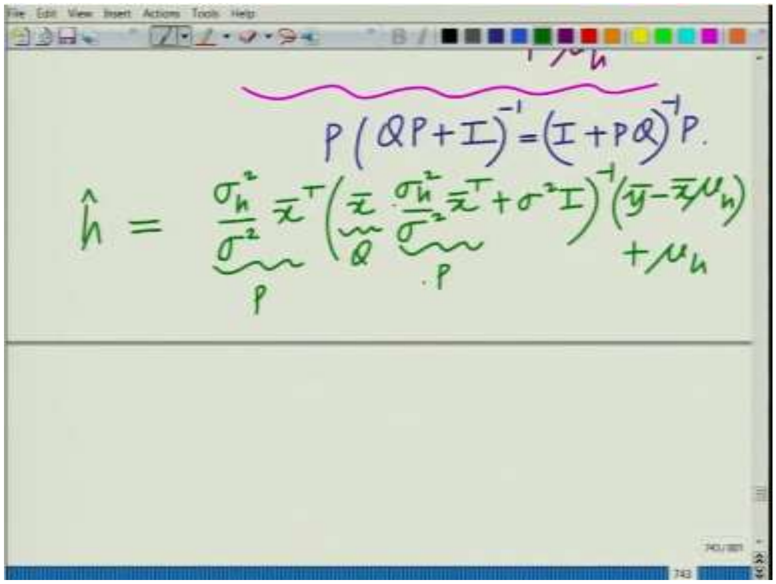
(Refer Slide Time: 0:25)



You have h hat equals the expected value of h given y bar and we have seen that the expression for this is sigma h square x bar transpose times sigma x h square x bar x bar transpose plus sigma square identity inverse into y bar minus x bar mu h plus mu h. So, this is the expression for the expected value of h given y bar or the what we are calling as the MMSE estimate of h and now I am going to use this, I am going to simplify this.

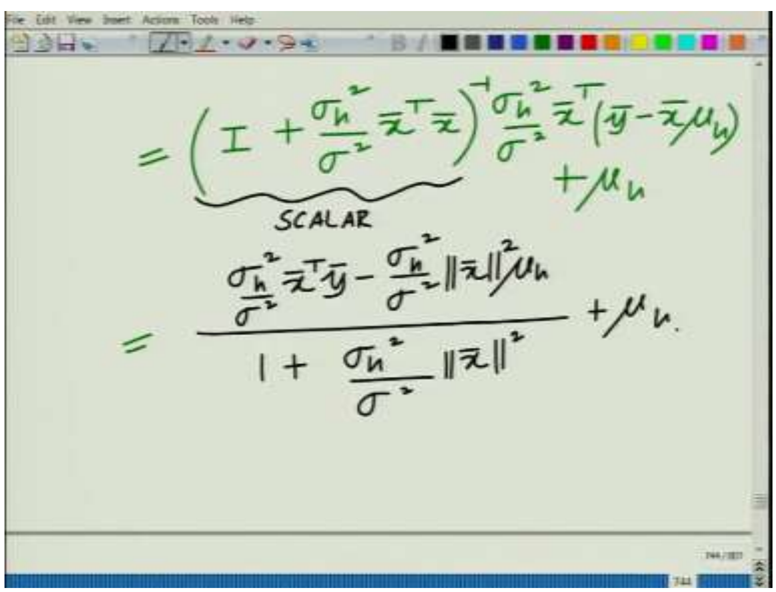
So, I am going to, because remember we said this is a rank 1 matrix, I think this is what we already noted that the rank of this matrix, this is a rank 1 matrix and I can use the following property, I am going to use the property P into QP plus identity, this is the property that we proved during the matrix inversion lemma P into QP plus I inverse equals I plus PQ inverse into P.

(Refer Slide Time: 2:02)



And this basically implies now that and now I can write this, bring the sigma square outside so I can write this as sigma h square over sigma square times x bar transpose times sigma or x bar into sigma h square by sigma square times x bar transpose plus sigma square identity inverse times y bar minus x bar mu h plus mu h and if you look at this now this is your matrix P and so this is your matrix Q and this is your matrix P.

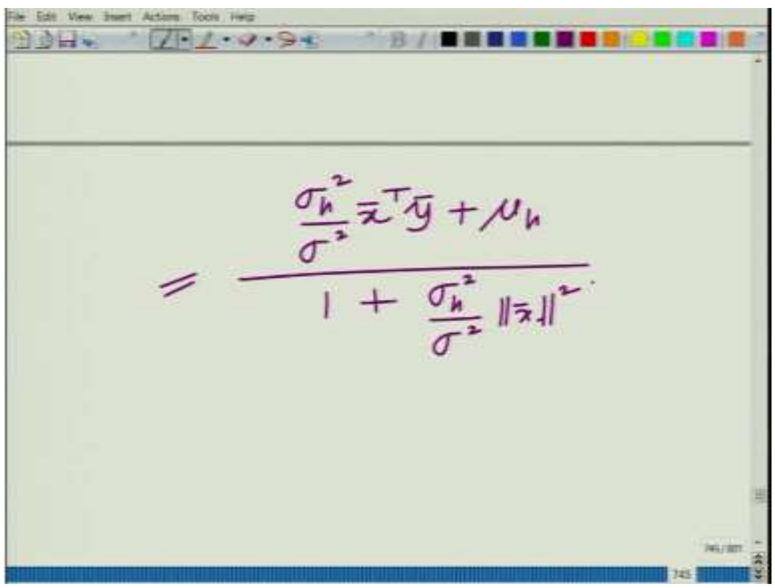
(Refer Slide Time: 2:56)

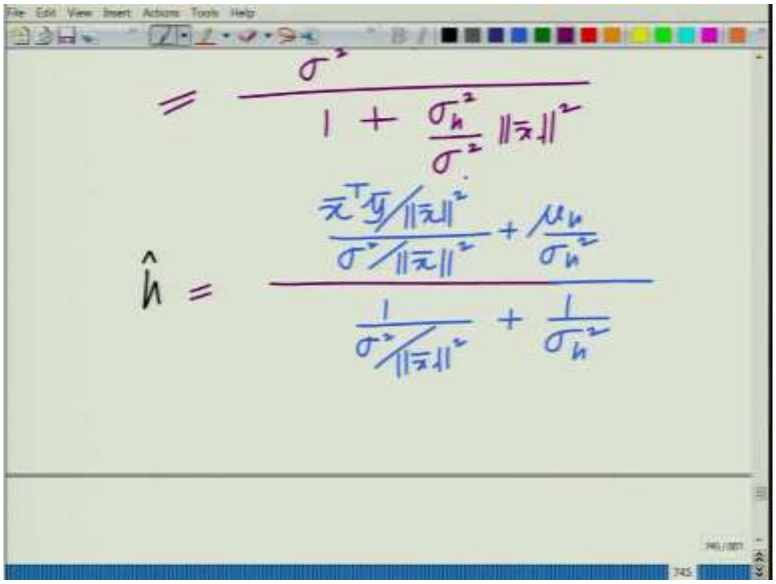


So, I can write this as P into QP plus I inverse equals I plus PQ inverse into P, so I can write this as I plus PQ inverse I plus PQ so P is remember this quantity here sigma h square by, so I write sigma h square by sigma square x bar transpose times x bar inverse into P, that is sigma h square divided by sigma x square, sigma square times x bar transpose into y bar minus x bar mu h plus mu h.

And this is therefore, now if you see I can write this as, this is now you can see this is simply a scalar quantity, it is not very difficult to see that this is simply a scalar quantity. So, this would be sigma h square divided by sigma square x bar transpose y bar minus sigma h square divided by sigma square into x bar transpose x bar that will be norm of x bar square times mu h divided by 1 plus sigma square, I mean sigma h square divided by sigma square norm of x bar square plus mu h.

(Refer Slide Time: 4:46)

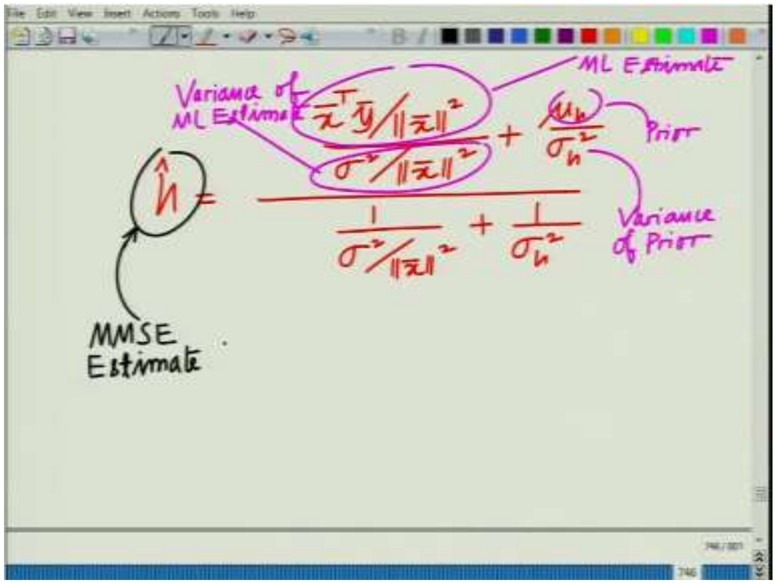




And now if you simplify this you can clearly see this reduces to the following expression, this is going to be given as sigma h square divided by sigma square x bar transpose y bar y bar plus mu h divided by 1 plus sigma h square divided by sigma square times norm x bar square. Which is now if you go one step further, you can simply write it as x bar transpose y bar divided by norm x bar square divided by sigma square divided by norm x bar square plus mu h divided by sigma h square plus 1 divided by sigma square divided by norm x bar square plus 1 divided by sigma h square.

So, if you look at this interestingly what you observe is this is a very interesting thing, this quantity now this is h hat, this is the MMSE estimate h hat. And what we observe is let us now derive some insights into this estimate is in fact very interesting, what you will observe is this if you look at the numerator.

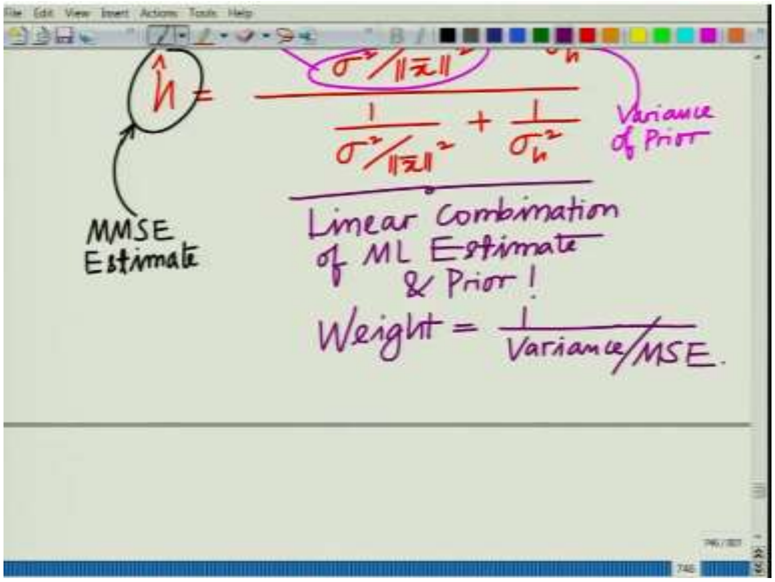
(Refer Slide Time: 6:32)



Let me just write it separately because I am going to now probably write a little bit over the expression so I just want do not want to disturb the original expression, let me just write this separately x bar transpose y bar divided by norm x bar square divided by sigma square divided by norm x bar square plus mu h divided by sigma h square divided by 1 by sigma square divided by norm x bar square plus 1 over sigma h square, and if you look at this, this is very interesting because this is the ML estimate.

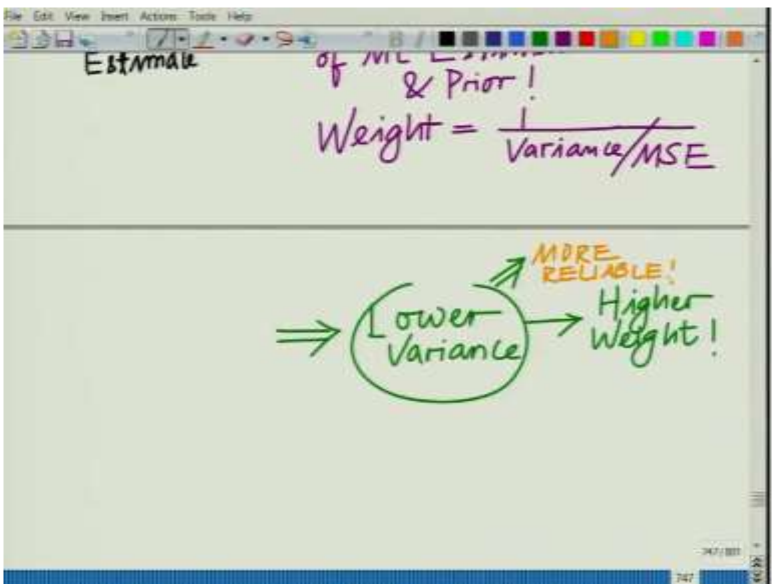
So, this quantity is the ML estimate and this quantity is the variance of the ML estimate or MSE of ML estimate and this quantity is the prior and this is variance of the prior or the prior mean, and now if you look at it what it is doing is this and this is basically your MMSE estimate, this is essentially what we are calling as the MMSE estimate. So, what this MMSE estimate is doing is taking the ML estimate and it is taking the prior and it is performing a linear combination of the ML estimate and the prior and what are the weights in the linear combination that is the inverse of the MSE.

(Refer Slide Time: 8:42)



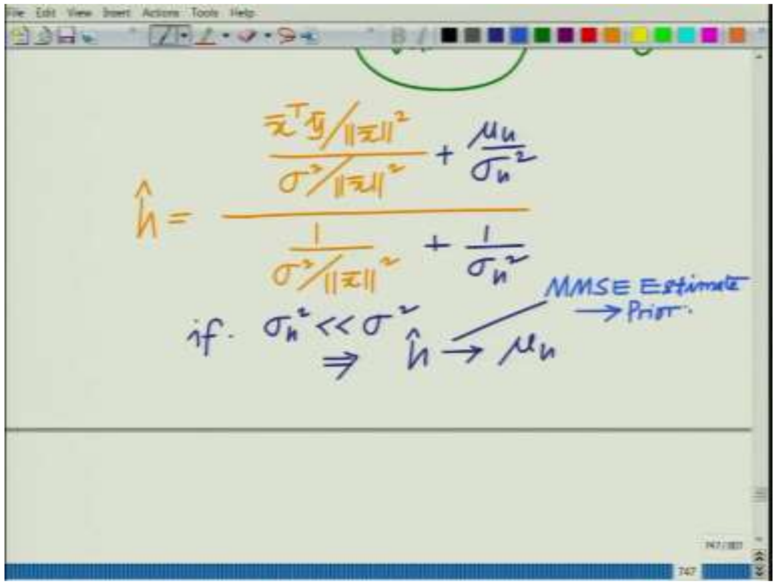
So, lower the MSE or lower the variance, it means higher the weight, so if you look at this, this is a very beautiful interpretation this is basically a linear combination of ML estimate and the prior, weight equals 1 over the variance slash MSE.

(Refer Slide Time: 9:29)



Implies, what does this implies? Lower variance leads to higher weight, naturally that makes sense because if the variance or MSE is lower the weightage of that should be higher, because it is more reliable, more accurate, so lower variance this implies essentially it is more reliable, so that is what it means. So, implies that it is more reliable, so the weight should be higher.

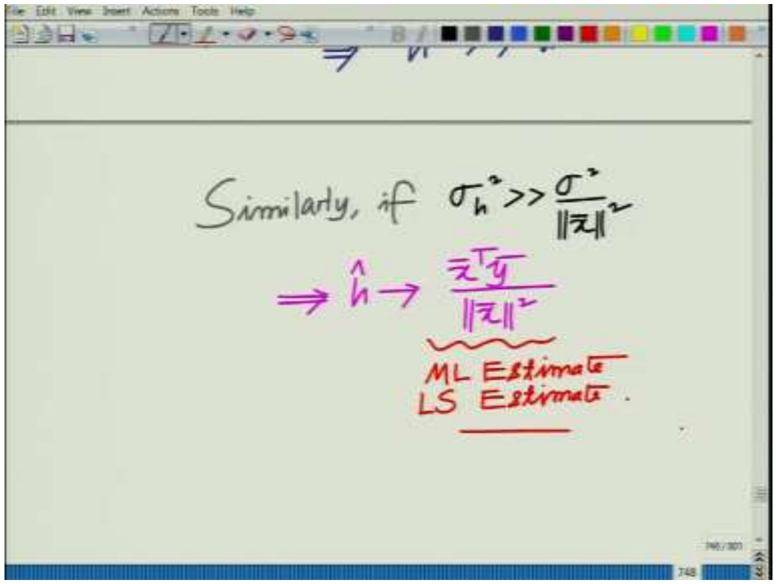
(Refer Slide Time: 10:24)



And further now you can also see from this expression if you take a look at this expression h hat equals x bar transpose y bar divided by norm of x bar square divided by sigma square divided by norm of x bar square 1 over sigma square divided by norm of x bar square, let me write these two with different colors, plus mu h divided by sigma h square plus 1 over sigma x square, now as sigma h square tends to infinity or sigma h square.

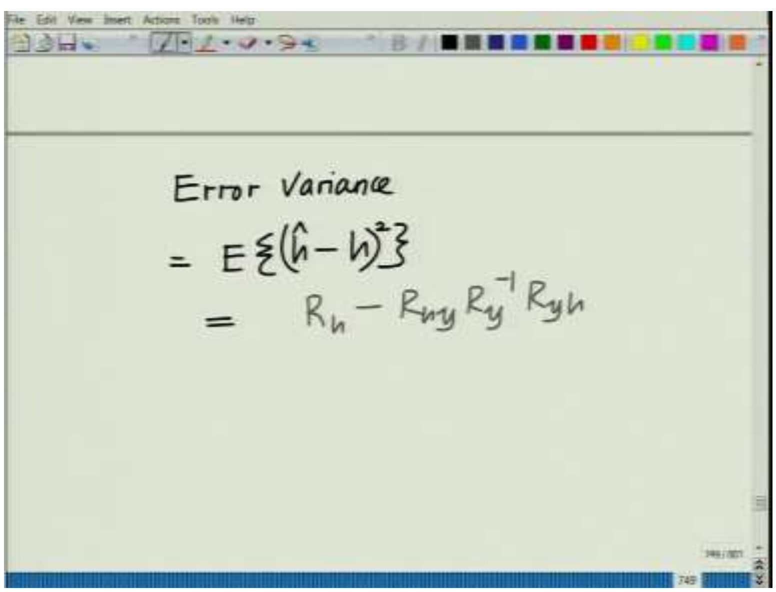
Now if sigma h square much less than your sigma square, sigma h square is very small this implies you can see this term will be large, that is your second term that is mu h by sigma h square 1 over sigma square these will be large which implies h hat tends to your mu h, so variance of prior is very small implies the mean MMSE estimate tends to the prior, this implies MMSE estimate tends to prior.

(Refer Slide Time: 12:00)



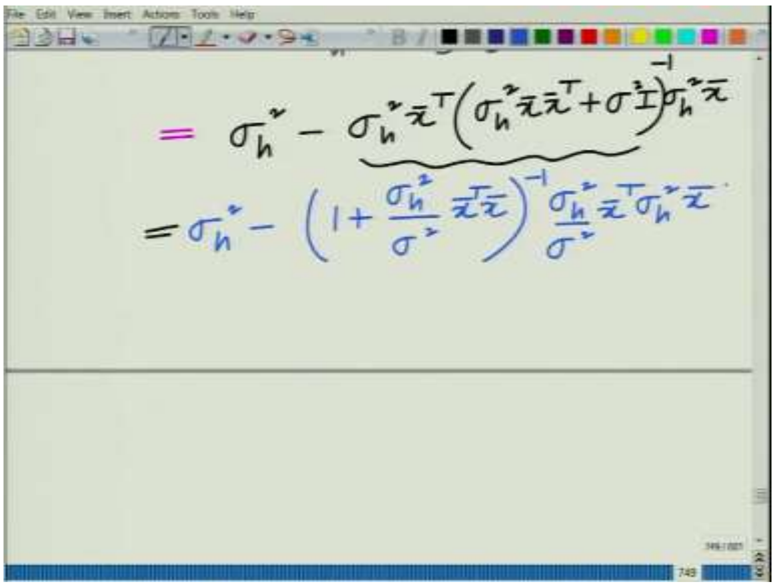
Similarly, by the same token if sigma x square is much greater than sigma square by norm x bar square this implies h hat tends to your ML estimate, that is your x bar transpose y bar divided by norm of x bar square which is basically the ML estimate. So, if you look at this, this is essentially the maximum likelihood estimate or basically what we are also calling as the least squares estimate, another name for this is for our Gaussian estimation problem, this is essentially the least squares estimate.

(Refer Slide Time: 13:09)



Now, further the error variance, now we calculate the error variance, we ask what is the error variance of this estimator. Let us calculate the error variance, error variance is given as follows. The error covariance or the error variance in fact, error variance this is equal to expected value of h hat minus h whole square, this is a simply a variance because this is a scalar quantity which we already know the expression that is Rh in this case, Rh minus Rhy into Ry inverse into Ryh.

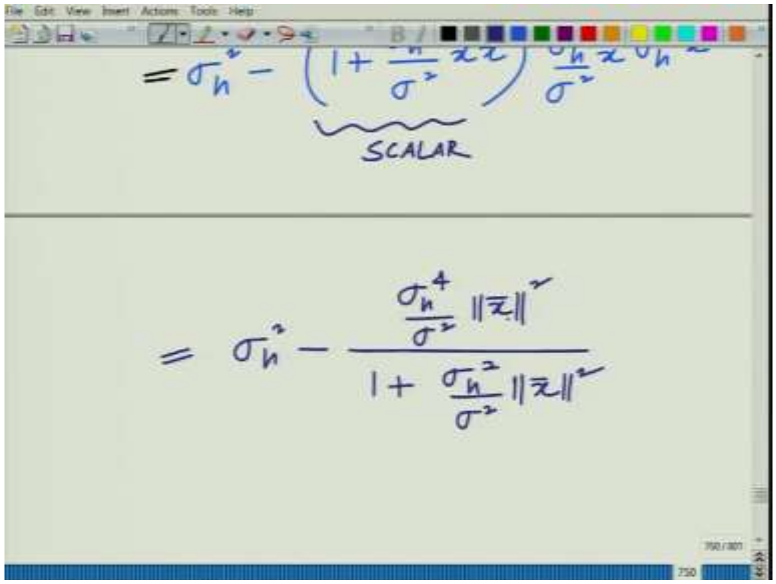
(Refer Slide Time: 13:59)



Which if you look at this, this is given as sigma h square minus sigma h square x bar transpose sigma h square x bar x bar transpose plus sigma square identity inverse into sigma h square x bar, so this is your Rhy Ry inverse into Ryh and if you look at this quantity, this quantity is exactly what we have over here, that is what we had simplified earlier this quantity is exactly what we had written as this thing that is your P into QP plus I inverse.

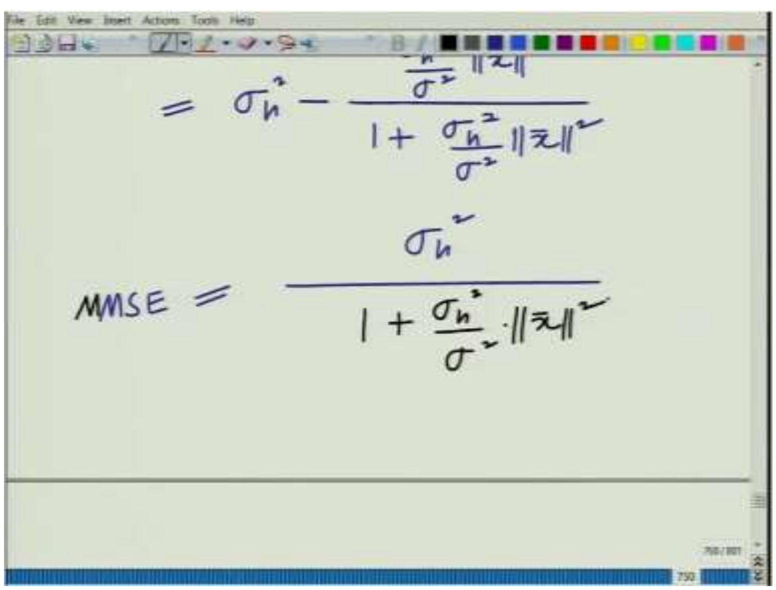
So, this is exactly if you look at it this exactly this quantity, so I can replace this by so using the same result again essentially using the same property I can write this as sigma h square minus this quantity, I can write this as 1 plus sigma h square over sigma square x bar transpose x bar inverse times sigma h square divided by sigma square x bar transpose sigma h square x bar.

(Refer Slide Time: 15:41)



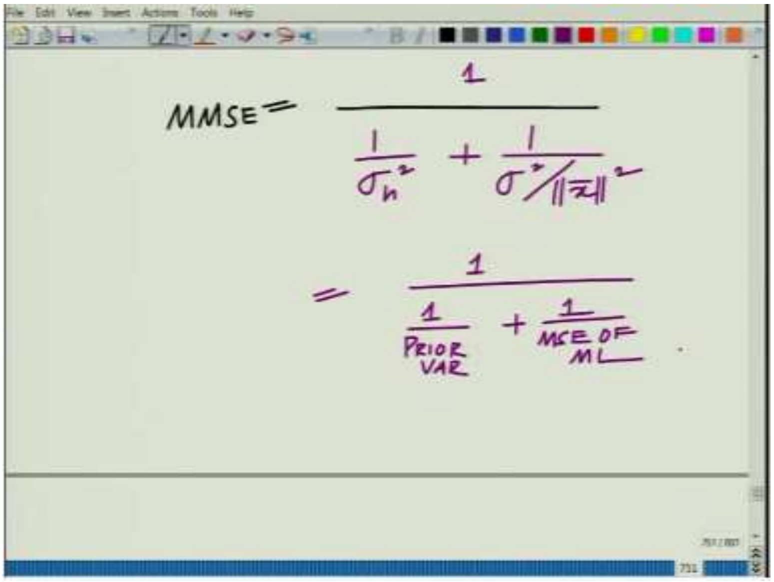
Which is essentially if you look at this, this is essentially sigma h square minus sigma x, I guess you can write this as sigma h raise to the power of 4, so once again the idea is that this is a scalar quantity, if you look at this is x bar transpose x bar so this quantity is a scalar quantity. So, this quantity is a scalar, so I can write this as sigma h raise to the power of 4 divided by sigma square norm x bar square divided by 1 plus sigma h raised to the square divided by sigma square into norm x bar square.

(Refer Slide Time: 16:34)



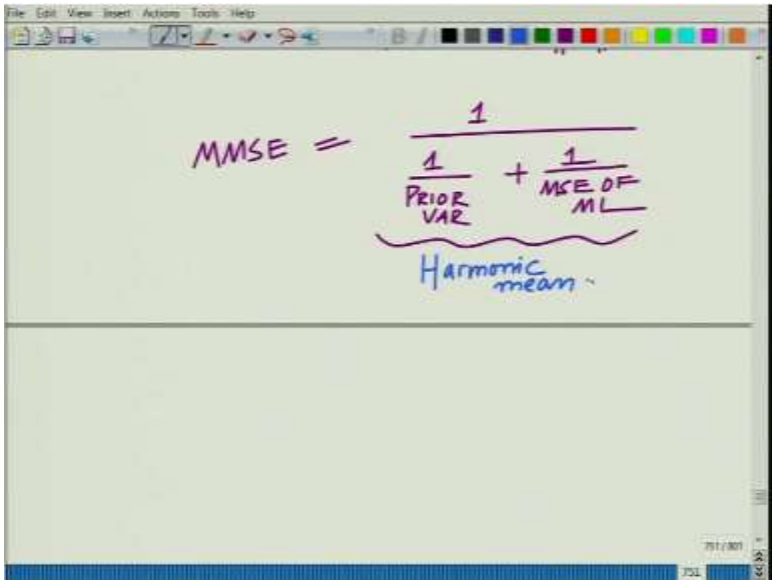
And if you simplify this again you will get something very interesting. So, this would be sigma h square, so the MSE the mean square error or the minimum mean squared error right this would be 1 over sigma h square divided by sigma square into norm x bar square.

(Refer Slide Time: 16:58)



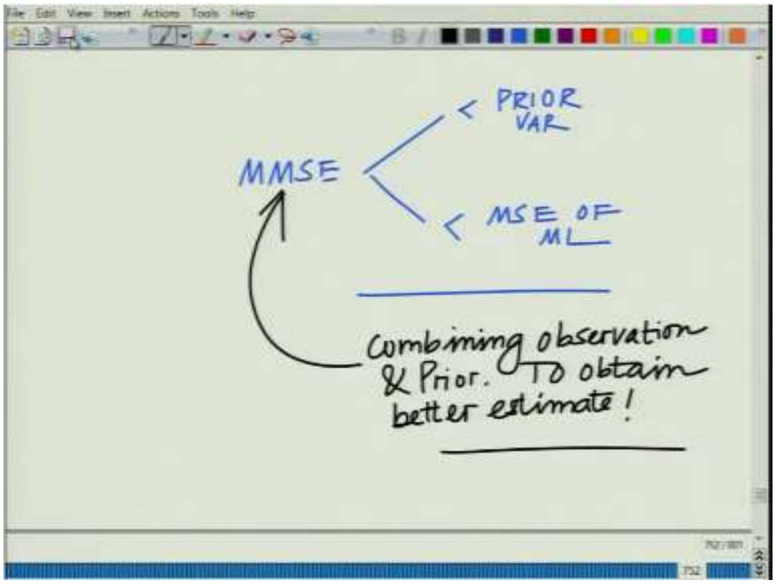
And if you look at this, this is equal to the MMSE minimum mean squared error, this is equal to 1 over bring the sigma h square down, so that would be 1 over sigma h square plus 1 over sigma square divided by norm of x bar square. So, this is essentially if you look at this 1 over the inverse of the variance of the prior inverse of the variance of the ML, so this is 1 over prior variance plus 1 over MSE of ML.

(Refer Slide Time: 17:57)



So, essentially this is the harmonic mean, so this you can see this is the harmonic mean, so which means it is like a parallel combination of resistances, so it is a harmonic mean of the prior variance and the ML MSC which means is less than both.

(Refer Slide Time: 18:22)



So, the MMSE because of the property this is less than the prior variance, so it is less than the prior variance and it is less than the MSE of the ML estimator, so it is less than the prior variance and less than the MSE of the ML. So, using the prior and ML you are combining them to obtain a better estimate lower MSE than each. So, essentially what this is doing is the MMSE principle combining the observation or the likelihood from the likelihood, combining observation plus prior to obtain a better estimate, so better estimate that both which has a lower MSE than both.

That is essentially the MMSE principle or the minimum mean squared error, so these are some of the interesting insights and of course, the analysis is endless, so I again once again urge you to go into all these aspects, the various aspects of linear algebra that we have looked at, the various principles look at the various applications and derive such beautiful applications as well as insights. So, we will stop this discussion here and thank you very much.