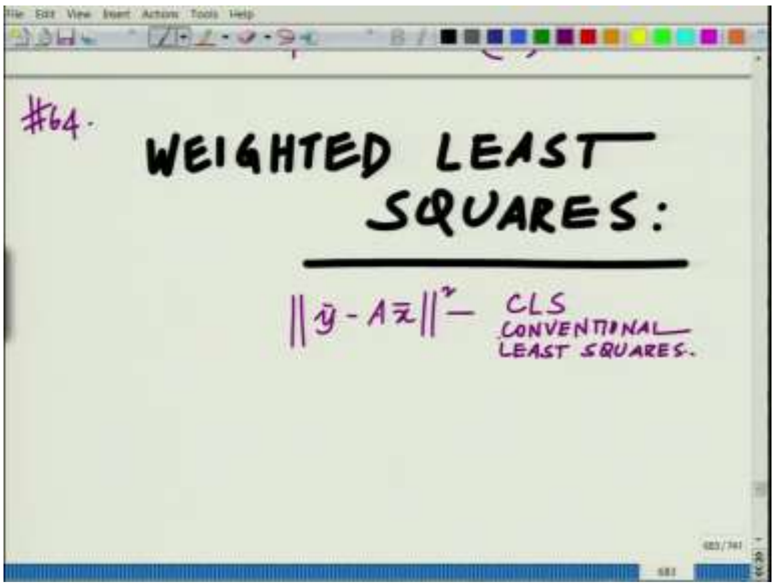
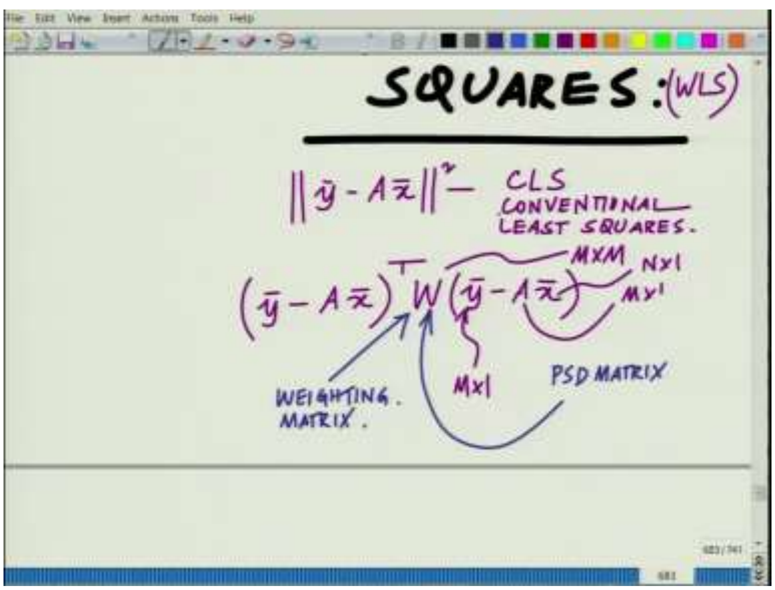
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Lecture No. 64
Weighted Least Squares Examples

Hello, welcome to another module in this massive open online course. So, we are looking at the concept of weighted least squares, remember where you have a weighting matrix. So, the weighted least squares.

(Refer Slide Time: 00:27)



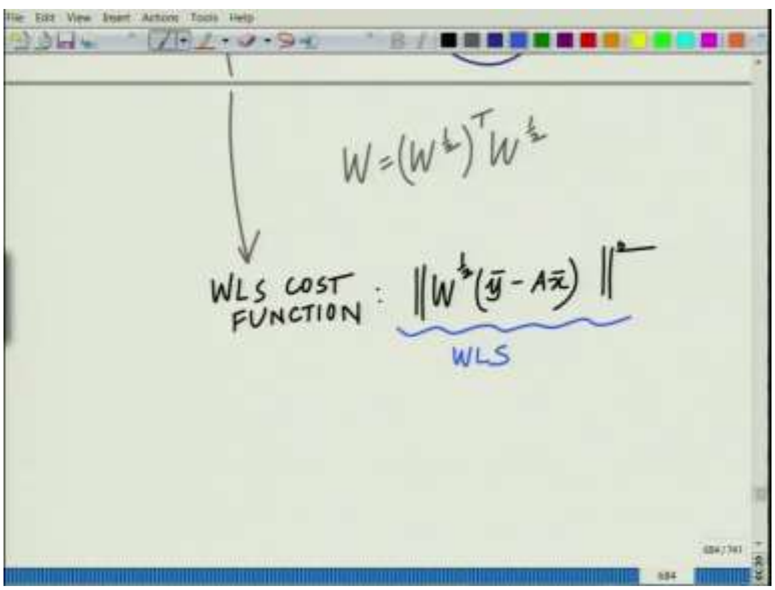


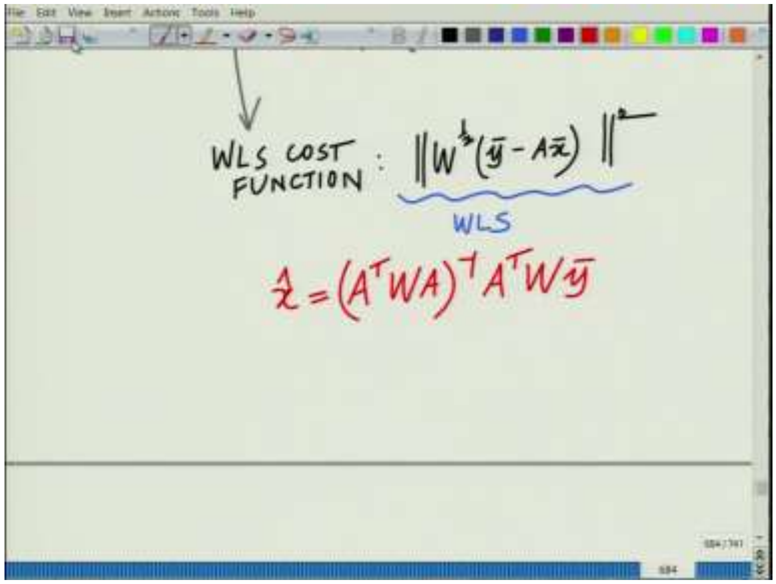
So, let us continue our discussion on this paradigm of, continue our discussion on this paradigm of weighted least squares. What happens in the weighted least squares? Well, you have the least squares cost function y bar minus Ax bar squares this we are calling as your CLS or Conventional Least Squares. And your Weighted Least Squares or what you can also call as your WLS.

I can write this as y bar minus Ax bar transpose y bar. In fact, you have a weighting matrix W into y bar minus Ax bar. And we have already seen this, this is an m cross m weighing matrix A is m cross 1 x bar is your n cross 1 parameter vector you can think of y bar as m cross 1. And this is essentially your weight matrix or matrix of weights, your weighing matrix.

This is your weighing matrix or your weighting matrix. I think sometimes it is also known as the weighting matrix I think that might probably be a better nomenclature, it is your weighting matrix. And this W is a positive semi definite matrix the other attributes are this so, you have all the properties of a positive semi definite matrix implies that its eigen vectors are orthogonal, eigen values are non negative greater than or equal to 0, so on and so forth.

(Refer Slide Time: 02:54)

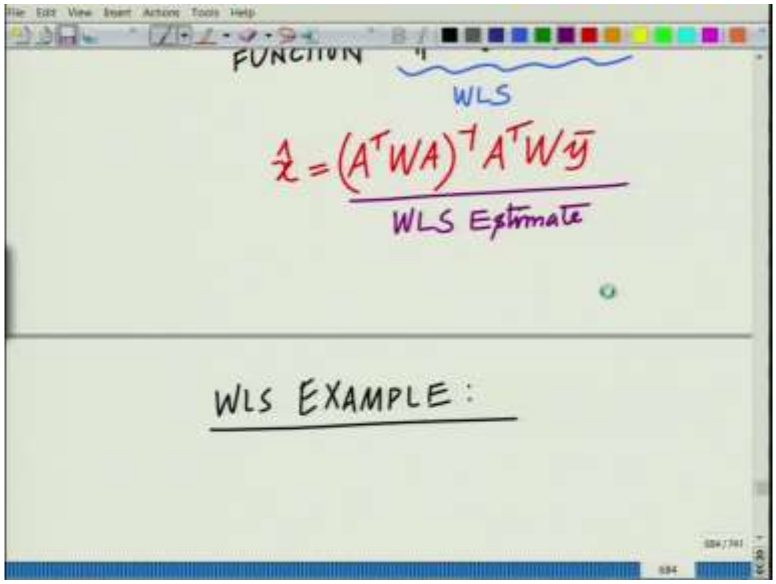


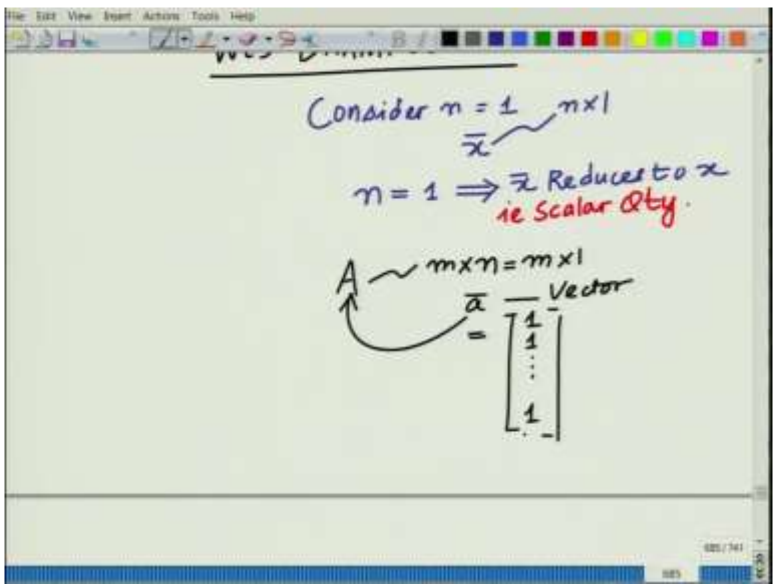


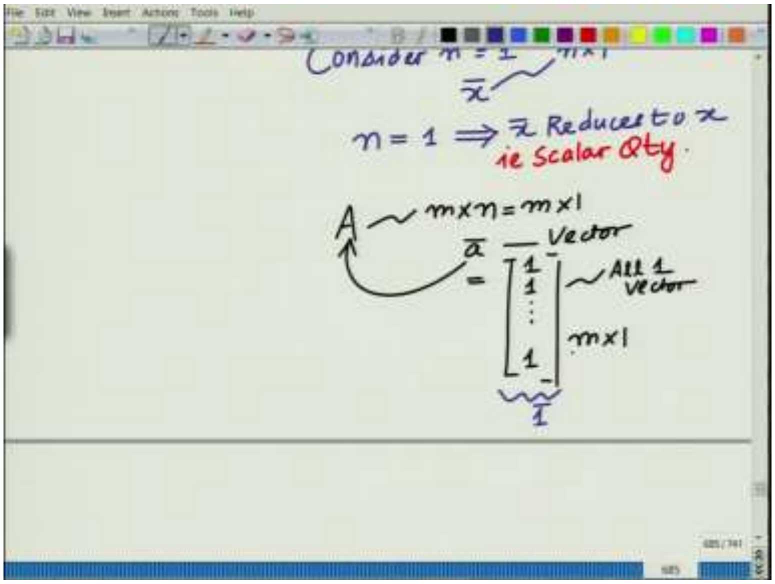
And, in fact, every positive semi definite matrix we have seen this can be decomposed as W transpose or you can decompose it as W half transpose into W half. So, therefore, I can also write this WLS cost function, the WLS cost function that can also be written as a norm of W half y bar minus Ax bar square, x bar whole square. So, I can write this as follows this is your weighted least squares cost function, weighted least squares cost function.

Anyway, we have simplified this further in the previous module we have derived what is the weighted least squares estimate and so on and so forth. And we have seen that the weighted least squares estimate is given as follows. So, I am going to, I am going to write the weighted least squares estimate. So, you have x hat equals A transpose the weighting matrix A inverse A transpose the weighting matrix y bar A transpose times W into y bar. And now, so this is our least square (())(04:41). Let us look at an example for this.

(Refer Slide Time: 04:44)



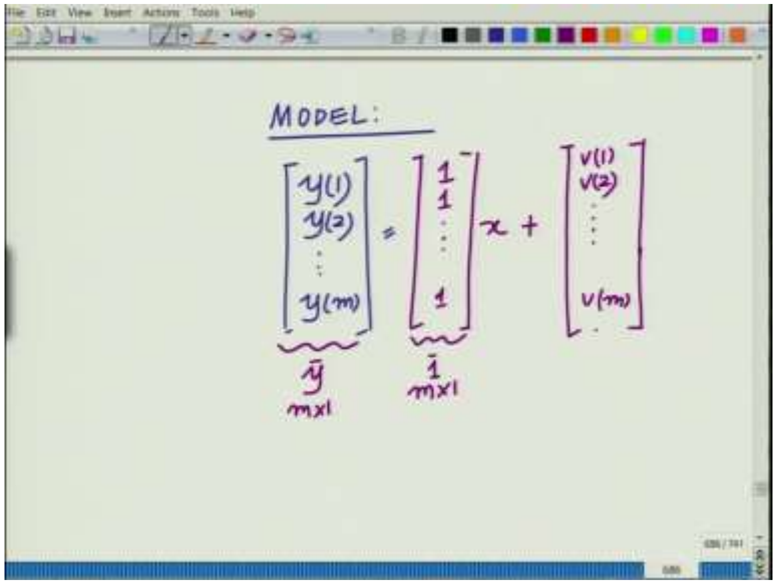


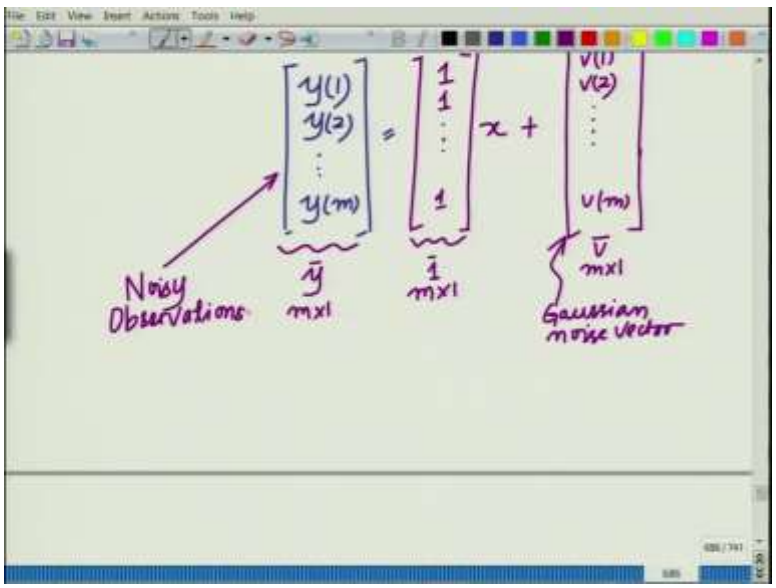


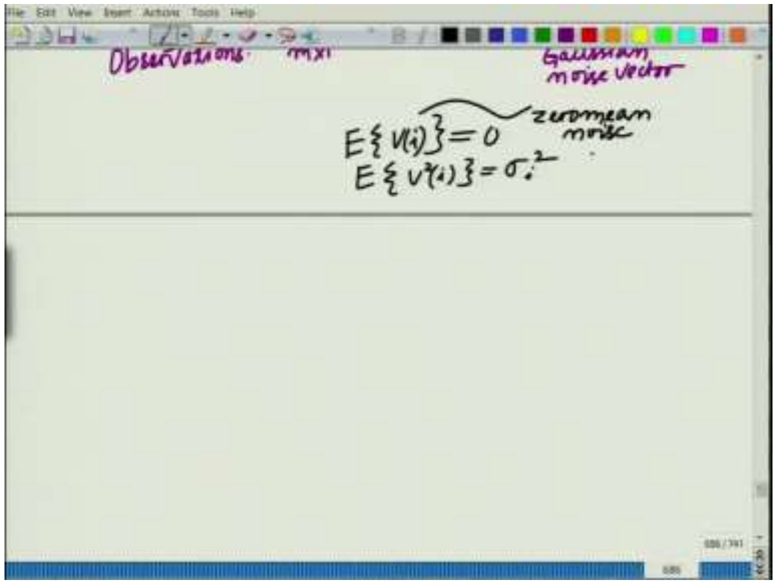
Let us look at an simple example and try to understand this better. So, this is our, this is our WLS estimate, this is our WLS estimate, let us now look at a simple example. Consider n equal to 1. So, to make, consider n equal to 1 small n equal to 1, remember we have our n is basically the size of, x bar is basically of size n cross 1. So, n equal to implies that x bar reduces to simply a scalar quantity reduces to x that is a scalar quantity i.e. reduces to a scalar quantity.

Now therefore, we are considering, as a special case we are considering the estimation of a scalar quantity. And further now, we are going to assume this matrix which is A, which is an m cross n that will be m cross 1, let us assume this to be a vector of all 1. So, now, if you look at A, A is of size m cross n which is equal to m cross 1. So, I can think of this rather as the vector A bar. And we are going to set this equal to the all one vector in this example, we are going to set this equal to the all one vector. So, we are going to set this equal to the all one vector. Size of this is m cross one and I am going to also denote this by one bar.

(Refer Slide Time: 07:16)







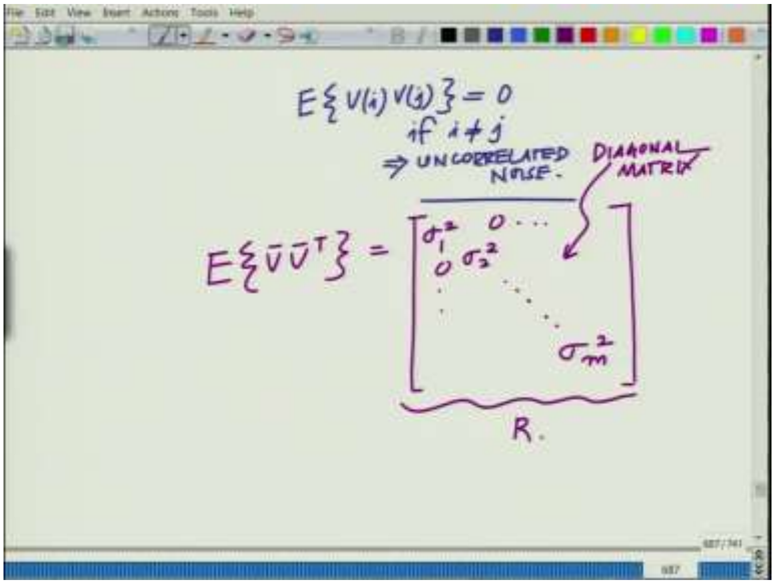
So, we can write our model, as our model as follows that is our least squares model, for least squares estimation as follows. So, we have the model y 1, y 2 if you look at this this is y m this becomes equal to your quantity which is you're a, which is the vector of all one. So, this is your vector, write this is your vector A bar.

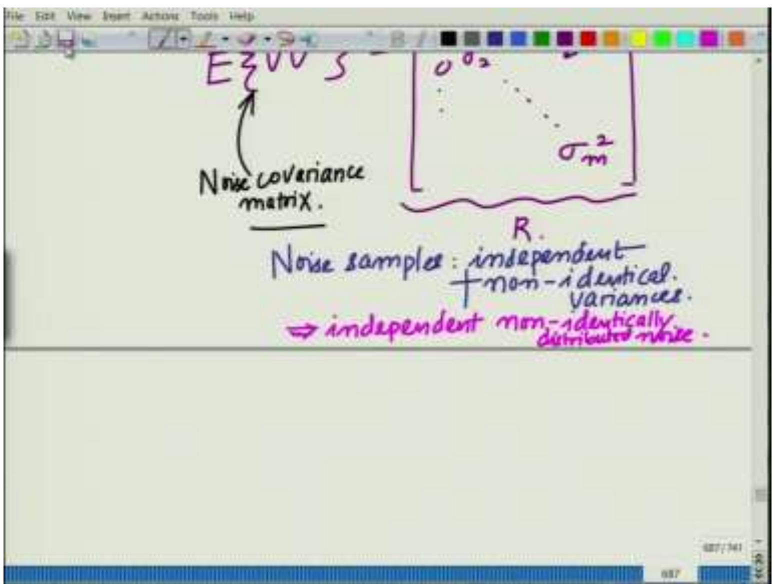
So, this is your y bar which is an m cross 1 vector. This is a vector A bar which is your m cross 1. In fact, this is the vector 1 bar which is a vector of ones times x plus, now you can write this as this is our estimation model you can write this as noise v 1, v 2, v m. So, this is the noise vector v bar which is your m cross m, let us call this as your Gaussian noise vector.

So, this is your noise, so this is a model which is a noisy estimation model. So, you are making several observations y 1, y 2, y m of this parameter x in Gaussian noise which is where v 1, v 2, v m. So, these are your observation. So, these are essentially your, so these are essentially your noisy, these are essentially your noisy observations or simply known as the observations in noise or simply call them as observations.

Now, typically, when we look at the noise we also have to describe the properties of this noise that is this noise properties. What are the properties of this noise, namely what are the mean and the variance. So, typically when we look at this noise we termed the Gaussian noise, the mean is essentially vi square or we can say vi square or the mean is 0, so this is 0 mean noise. So, this is 0 mean noise. And the variance is expected value of v square i, this is equal to sigma A square.

(Refer Slide Time: 10:06)





And now if we look at the correlation between these noise samples, vi to vj this is equal to 0 if i not equal to j implies uncorrelated noise which, because the noise is Gaussian, it is also independent. So, the Gaussian's noise, Gaussian random variables they are uncorrelated that also was that they are independent, that is a property of the Gaussian.

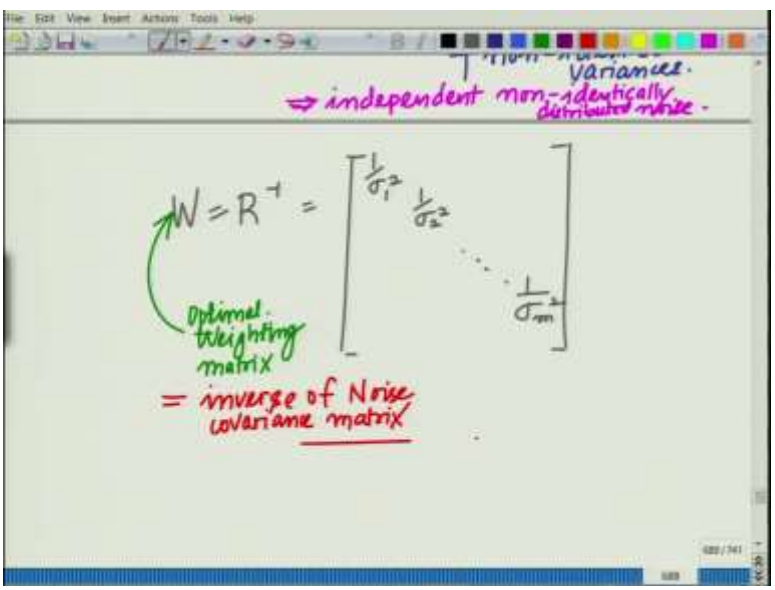
In any case, now, if you look at the covariance matrix of this noise, now if you look at the covariance matrix of this noise, which has the variance which ith element as the variance sigma i square. If I look at the covariance matrix of the noise, this is going to be a diagonal matrix with the elements sigma 1 square, sigma 2 square, so on sigma m square. So, this is basically your diagonal matrix which we are basically going to call as R.

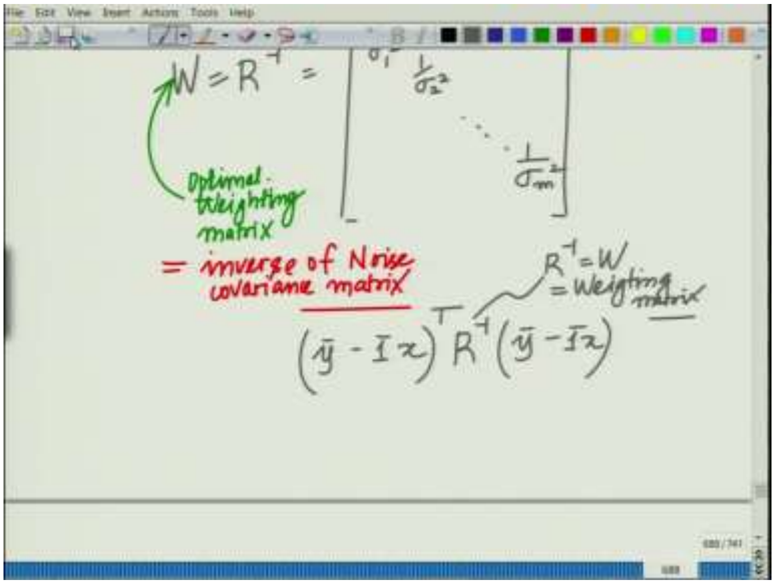
So, this is essentially you can see this is because the noise samples are uncorrelated and but they can have possibly non identical variance as sigma 1 square, sigma 2 square, sigma m square. If the variances are identical, then it become sigma square all the entries are sigma squared this becomes a diagonal matrix sigma square times identity.

But now this is diagonal matrix, but the entries are not identical, the entries are sigma 1 square, sigma 2 square, sigma m square. So, this is your covariance matrix, the first thing is this is your noise covariance matrix, this is your noise covariance, this is the noise covariance matrix and because the noise variances are, the noise samples are independent, but variances are not identical.

So, noise samples these are independent plus non identical variances. So, the such a noise is known as independent non identically distributed noise. So, this is essentially independent non identically distributed, this is independent non identically distributed noise. What it means is the noise samples are independent, but they are non identically distributed the covariance matrix is diagonal, with the entry sigma 1 square, sigma 2 sigma square, sigma m square.

(Refer Slide Time: 13:42)



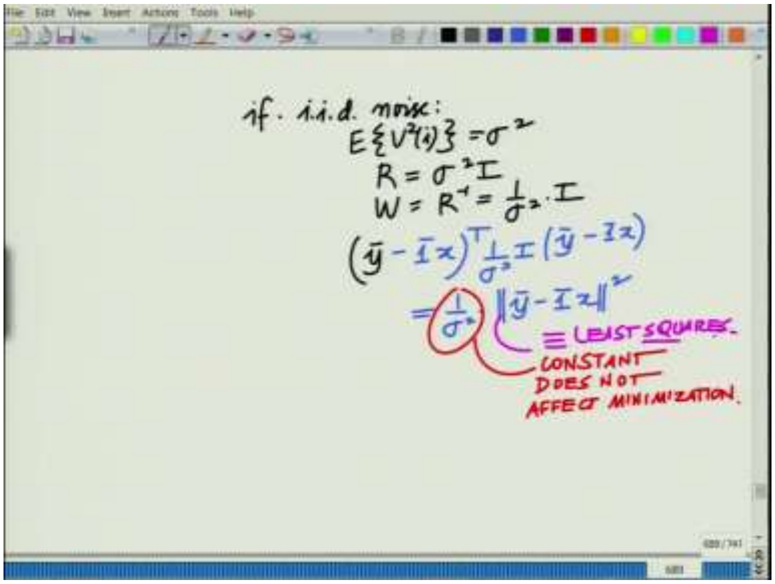


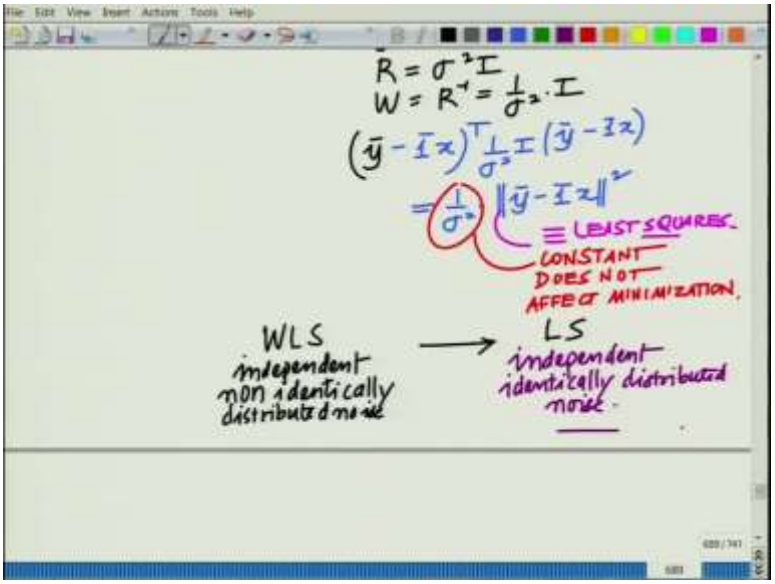
And now, if you look at R inverse, now if you look at R inverse that will also be a diagonal matrix with all the diagonal entries inverted. So, that will be 1 over sigma on square, 1 over sigma 2 squared, so on, 1 over sigma m square this is R inverse. And now, for our estimation problem, for our least square estimation turns out that in such a situation where you have the noise samples independent but non identically distributed, turns out that the weighting matrix for estimation is our inverse, that is the point.

So, W equal to R inverse, this can be shown that this is the optimal weighting, and we will justify this, this is the optimal weighting, which essentially is equal to the inverse of the noise covariance matrix that is the interesting point, inverse of, this is the inverse of the noise covariance matrix. That is, we want to consider the estimation problem y bar minus 1 bar x transpose, I believe that I have written that correctly y bar minus 1 bar x transpose into R inverse y bar minus 1 bar x.

So, this R inverse equals W which is essentially your, becomes now your weighting matrix. So, for this least squares problem, the weighting matrix becomes the, weighting matrix is the inverse of the covariance matrix and now, we can once again see if the noise samples are independent, identically distributed.

(Refer Slide Time: 15:59)





Now if, if noise is i.i.d. that is, you have all of them have identical variance sigma 1 square, then covariance becomes sigma square times identity weighting matrix becomes inverse of the covariance which is 1 over sigma square times identity and the problem becomes bar y minus 1 bar x transpose 1 over sigma square times identity y bar minus 1 bar into x, which is now 1 over sigma square, this identity goes so you will have simply y bar minus 1 bar x transpose y bar minus 1 bar x which I can write as y bar minus 1 bar x norm square.

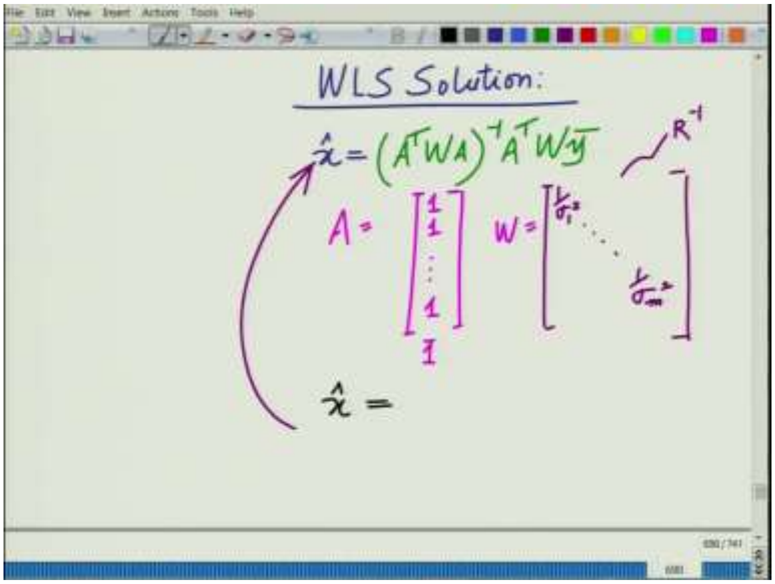
And since this is a constant, this does not affect, this is a constant. So, does not affect the minimization because, whatever x minimizes the norm square you multiply it by constant this is the same x that minimizes this cost function. Therefore, this reduces to the least squares. Now, for this reduces to the least squares.

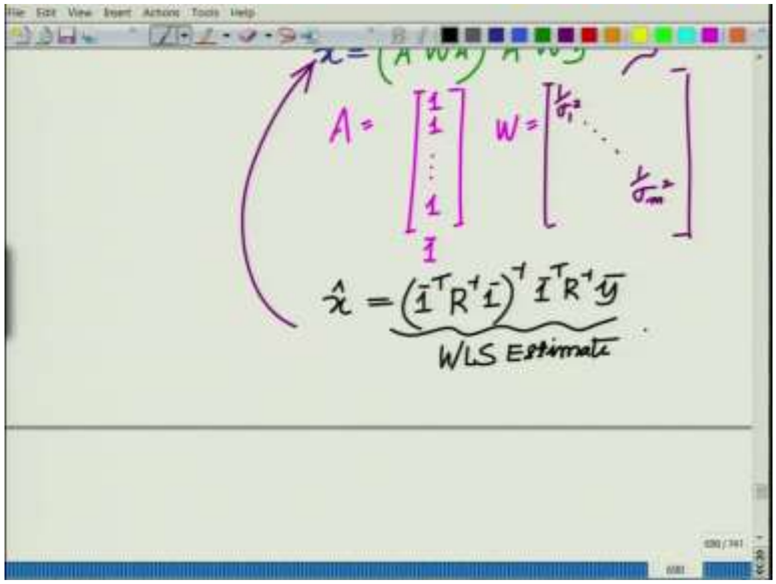
So, what this essentially means is that when the noise is independent identically distributed, the weighted least squares optimization problem reduces to the least squares, that you can, you can say that is essentially your conclusion or when the noise is independent non identically distributed least squares becomes the weighted least squares.

So, the weighted least squares estimation problem or the weighted least squares problem is the optimal cost function to consider, to estimate or to determine this unknown quantity, what we are also calling as the parameter, you can also think of this as your regression coefficient and so on and so forth. So, these are all the essentially these are all the same terminology. And now, what we can think of this is, so this does not affect the minimization.

And now, what is your estimate, (())(18:35) the way to think about this is weighted least squares is for independent non identically distributed. And this becomes your least squares for independent identically distributed noise, this becomes your least squares four independent identically distributed noise. And now, we can write this as the following thing. Therefore, now in this case you can write it as what is the estimate. Now, we know what is the least square. Now, we know what is a weighted least square solution.

(Refer Slide Time: 19:45)

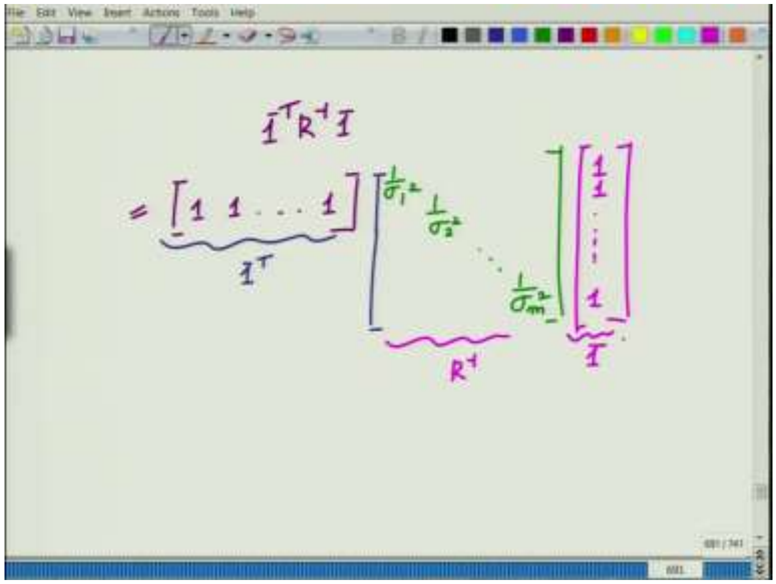


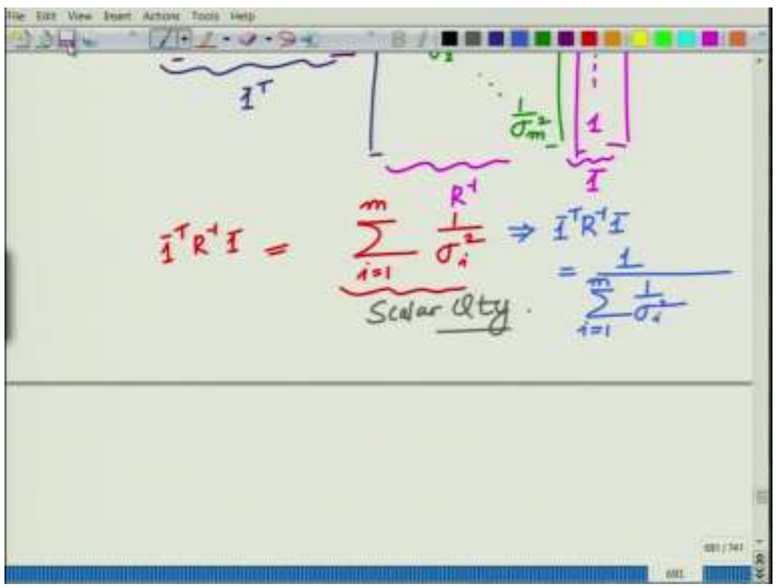


So, the WLS solution, we know what is a weighted least square solution recall this is x hat equals A transpose W A inverse A transpose W y bar. And A equals the vector of all ones this is your 1 bar and W equals your R inverse that is 1 over sigma square 1 over sigma m square. So, this is essentially your weighting matrix which we already said, this is your R inverse. And therefore, now I can write this, substituting this in this formula the weighted least squares estimate this is obtained as x hat, I hope this discussion is clear to all of you.

I have replaced this A by 1 bar the vector of m dimensional vector of all ones the weighting matrix becomes your inverse of your covariance matrix which is essentially the diagonal matrix with the inverse of the, inverses of the variances. So, this becomes 1 bar transpose R inverse 1 bar further inverse of this times 1 bar transpose R inverse y bar, this is your WLS estimate, this is your weighted least squares estimate.

(Refer Slide Time: 21:49)



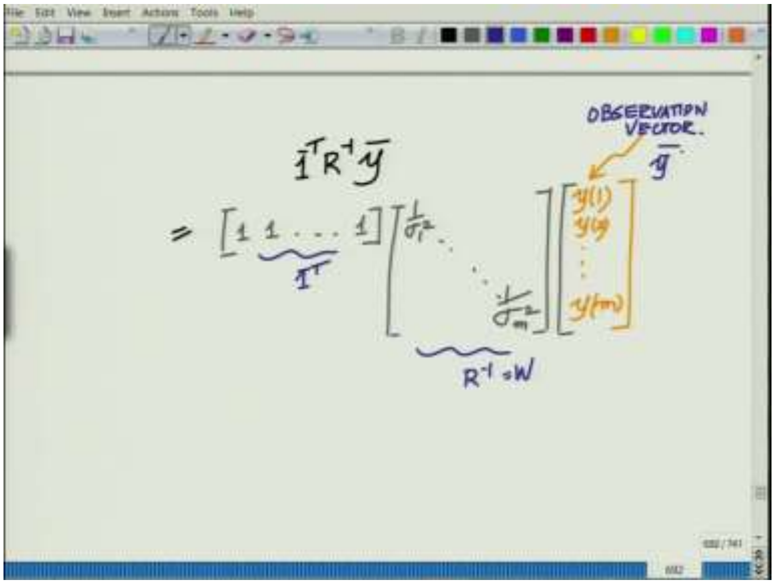


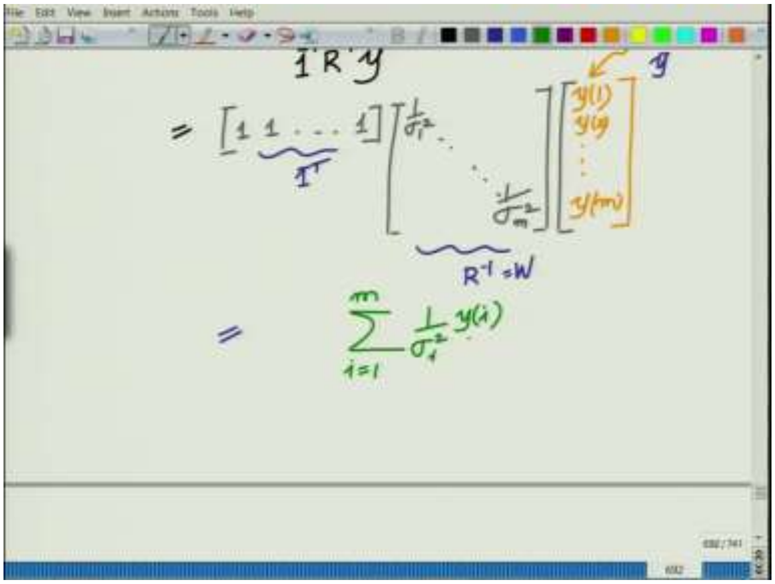
Let us simplify these quantities. Now, 1 bar transpose R inverse 1 bar it is not very difficult to see what this quantity is going to be, 1 bar transpose is your row vector of all ones this is essentially your 1 bar transpose R inverse we already seen, that is your diagonal matrix of 1 over sigma 1 square, 1 over sigma 2 square, 1 over sigma m square. And this is your 1, 1 vector of all ones this is your R inverse, this is your vector of all ones.

And therefore, if you simplify this you can clearly say this is nothing but summation i equal to 1 to m 1 over sigma square. This is essentially your 1 bar transpose R inverse 1 bar and not difficult to see this is your, this is a scalar quantity.

So, this is a scalar quantity, this is a scalar quantity. Therefore, I can write, essentially implies the inverse of this is nothing but 1 bar transpose R inverse 1 bar, this I can write as 1 bar 1 over summation i equal to 1 two m 1 over sigma i square. So, this is a scalar quantity. So, I can simply write it a tickets reciprocal, I cannot normally do that, but in this case because it is a scalar quantity.

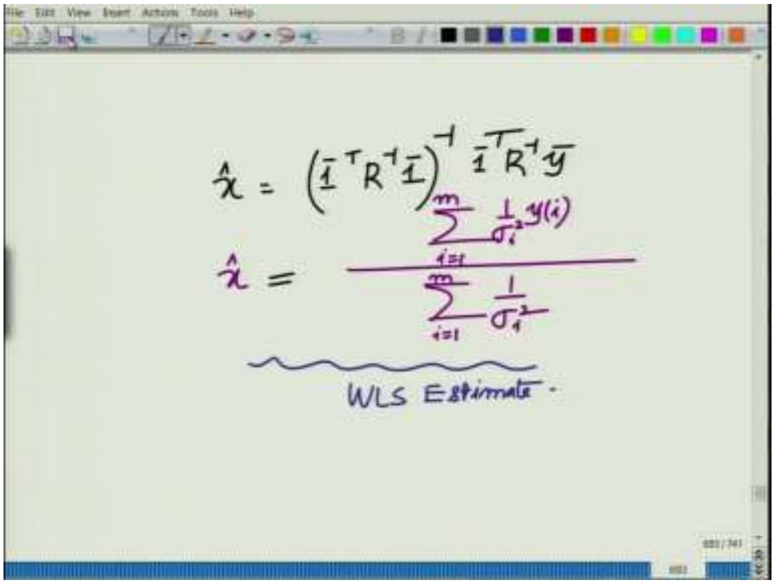
(Refer Slide Time: 23:47)

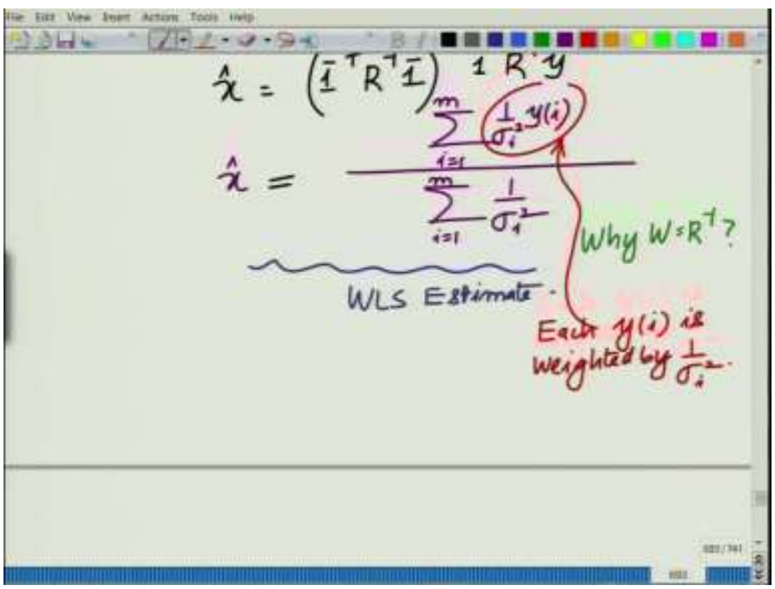




What about the other quantities that is your 1 bar transpose R inverse y bar. Once again if you look at that, that is essentially your row vector, take the row vector of all ones put your matrix R inverse 1 over sigma 1 square so on, 1 over sigma m square times your column vector that is your y 1, y 2, upto remember this is your column vector of observations, this is your observation vector, this is the observation vector. And this is your 1 bar transpose, this is your R inverse that is your weighting matrix, this is y bar, and this you can readily see this reduces to nothing but very simply, this is essentially your I equal to 1 to m 1 over sigma i square yi.

(Refer Slide Time: 25:11)



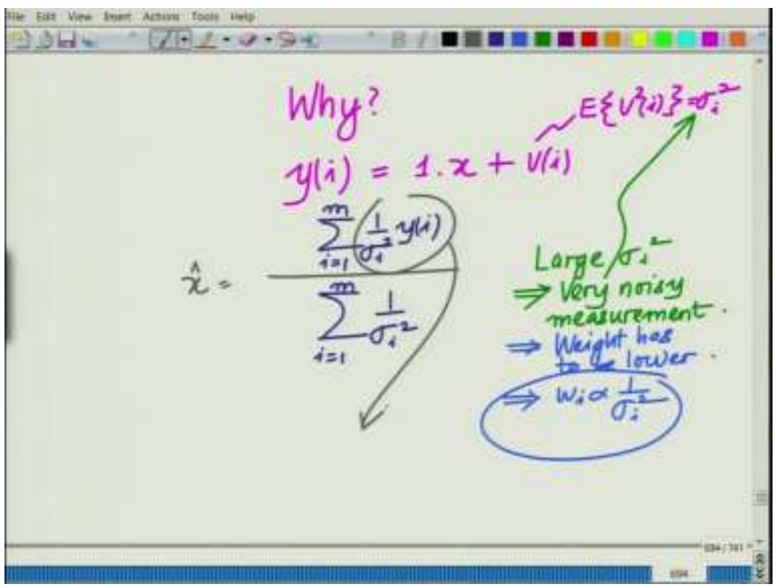


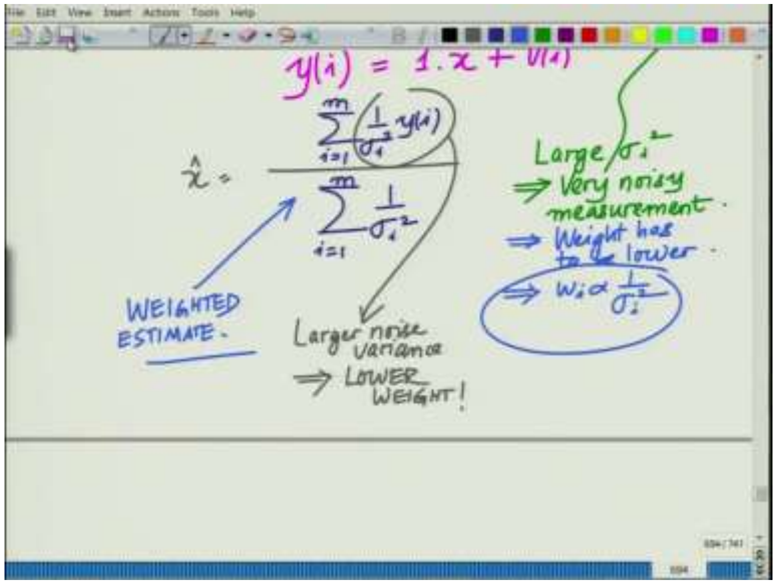
And this is the numerator and therefore, if you look at x hat once again that is 1 bar transpose R inverse 1 bar inverse 1 bar transpose R inverse y bar, which is essentially now, I can write it in the following fashion, this is essentially the denominator is summation i equal to 1 over m 1 over sigma i square numerator is i equal to 1 over m 1 over sigma i square y i.

So, this is your essentially your weighted least squares estimate for this simple problem. So, this is the WLS. For this problem with independent, independent non identically distributed noise. At now I said, remember I said I am going to show you, or I am going to describe to you why this paradigm makes sense, why this W equal to R inverse makes sense why, why W equal to R inverse?

Why not W equal to R? Why do not we set the weighting matrix directly as the covariance matrix rather than as the inverse of the covariance? But of course, there is a method to do that in a mathematically rigorous fashion, but even in this simple formula, you can see it reflected in the fact that if you observe this carefully, you will see each y i is weighted by each y i is weighted by, each y i is weighted by 1 over sigma square.

(Refer Slide Time: 27:19)





Now, why is this significant, now let us reflect on this for a little bit why is this significant? Now, remember our moral is y equals 1 times x plus V i and we have expected value V square i equals sigma i square. Now, sigma i square is high large sigma i square implies very noisy measurement, measurement is very noisy measurement, that is the noise is, the measurement is very noisy.

And when the measurement is very noisy, employ, naturally the weightage has to be lower not larger, so implies that the weight has to be lower. So, implies weight, implies the weight Wi is proportional to 1 over sigma square and this is why it makes us, you cannot have the weight proportional to sigma i square because that would mean if the larger the noise the greater is the weight.

Now, if you look at this estimation, you have a very interesting interpretation for that and that is as follows that is summation i equal to 1 to m 1 over sigma i square y i divided by 1 over i i equal to 1 to m 1 over say y square. So, you are weighting noisy, so as (())(29:18) 1 over sigma square which means that the noisy measurements have lower, the which means larger, worse the noise weight is lower, which means larger noise variance implies lower weight.

And therefore, this is a weighted estimate and that is the interesting aspect of this. And it is also very, so this is a weighted estimate. And what you are doing is, is also logical that is those measurements where the noise variance is greater, you are ascribing a lower weight to them, that is you are weighing them by 1 over sigma i square.

So naturally, the variance is larger, the weight to that measurement is lower. Now, and also the opposite also holds vice versa. That is, if sigma i squared is small noise variance is low, 1 over sigma squared is large. Therefore, the weight is larger. So that, both it holds both ways. And therefore, now this weighted estimate, estimate gives you a much more reliable, much better estimate of the quantity x rather than one way you are ascribing equal weights to all the quantities.

And worse, if your weights are sigma i square, that is, which means that your larger the noise variance, larger is the weight that will completely give you a contradictory kind of estimation principle where the unreliable measurements are being weighted by larger weight. So, the weighting matrix it makes a lot it makes logical, so it is logical to consider the weighting matrix as R inverse rather than considering, a concentrate as W equal to R.

And in fact, if you look at the maximum likelihood estimation principle and so on, this can be derived more rigorously, but even with a simple kind of a logical argument and intuitive argument, we can clearly see why the weighting matrix has to be R inverse and not R, where are remember is the covariance matrix of the noise. So, with this interesting example, let us conclude this module here. And we will look at, continue this discussion, including about this and other such similar paradigms in the subsequent examples. Thank you very much.