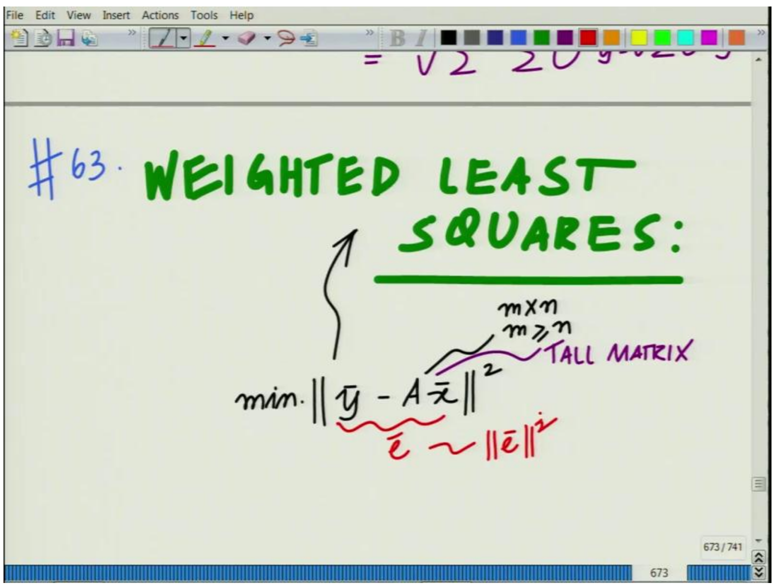
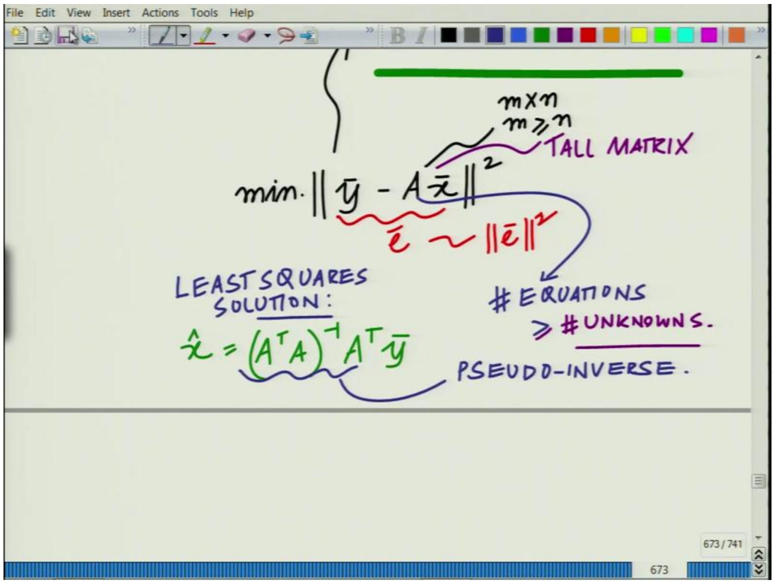
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor Aditya K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Lecture 63
Weighted Least Squares

Hello, welcome to another module in this massive open online course. So, we are looking at least squares and let us look at another extension, another interesting and very useful extension of this least squares paradigm which is termed as Weighted Least Squares, which is you can think of this as a generalization of the least squares principle.

(Refer Slide Time: 00:32)



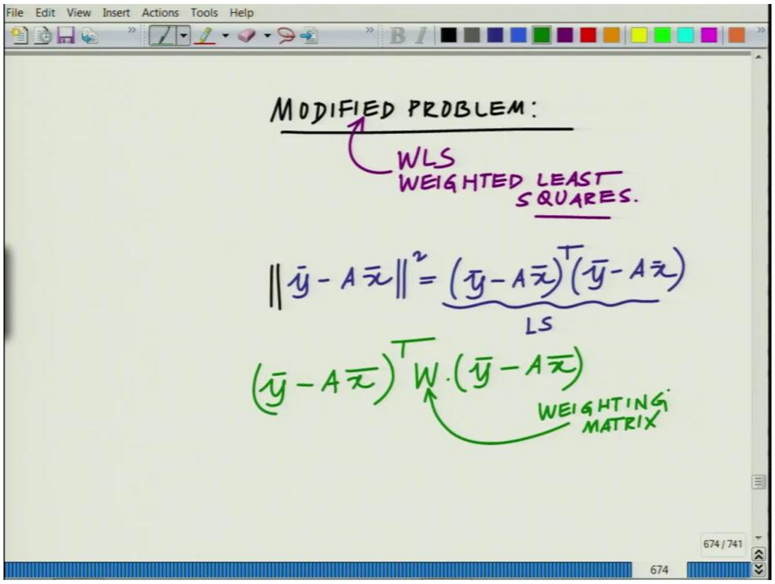


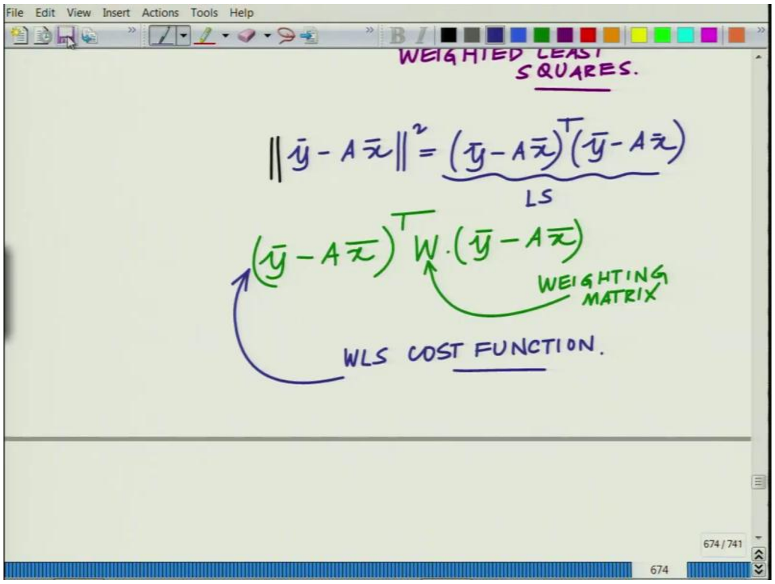
So, we want to look at the concept of weighted, we want to look at this notion of weighted least squares. Now, remember we have the conventional least squares in which you have, you want to minimize the square of the error. So, you have y bar minus Ax bar square where A, A is an m cross n matrix with m greater than or equal to n. So, this is also termed as a tall matrix we have seen that.

So, this is your traditional what we have seen as least squares, this is the norm y minus Ax bar square. You want to find the y x bar which such that Ax bar is the best possible approximation to y bar the observation vector and the error, the norm, the square of the norm of the error, this is the error, if you remember this is essentially your error and we are talking about the norm square of the error which we are minimizing, that is essentially what the least squares problem is achieving.

And the least square solution is, and so, A is a tall matrix this implies that number of equations, another way also to look at this is that the number of equations is greater than or equal to the number of unknowns. So, this is another way to look at this. And the solution of the least squares, the least squares solution this is given us x hat equals A transpose A inverse A transpose times y bar where this matrix A transpose A inverse A transpose this is also termed as the pseudo, this is also termed as the pseudo inverse of the matrix A. So, this is essentially your least squares problem and the least squares solution.

(Refer Slide Time: 03:32)



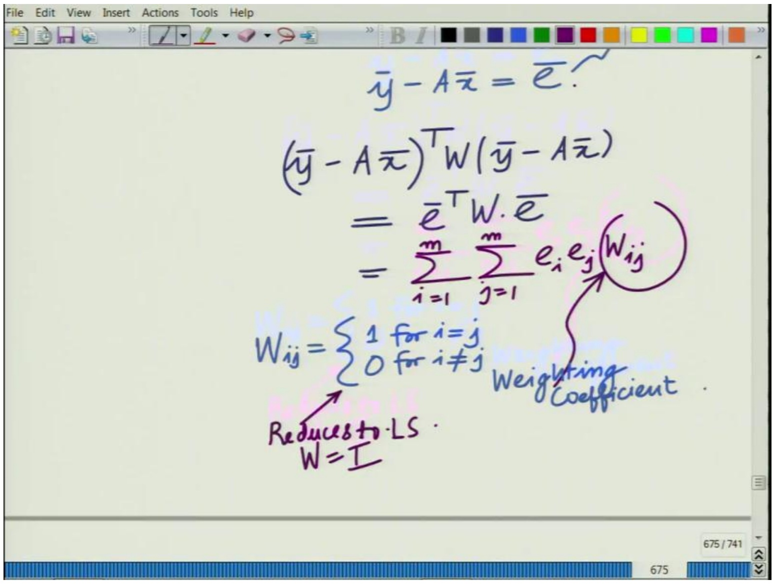


Now, let us look at a slight extension to this problem this is known as the Weighted Least Squares or the WLS problem. So here the LS, which is the least squares and then you have the you have the modified problem which is what we call as the WLS, the WLS or the weighted least squares. And what this problem is that is you have instead of so, if you look at the least squares problem that is norm y bar minus Ax bar square.

So, we can write this as y bar minus Ax bar transpose y bar minus Ax bar. Now, this is your least squares problem. Now, now, introduce a weighting matrix, so we introduced a weighting matrix. So, this is your least squares, this is your LS.

Now, instead of this you have y bar minus Ax bar transpose times W times y bar minus Ax bar where W is the weighting matrix, W is a weighting matrix and this is essentially your weighted least squares cost function, those this is the what we term as the WLS. So, what we are essentially doing over here is we are introducing this weighting function that is essentially what we are trying to do and essentially that takes the different errors.

(Refer Slide Time: 05:56)



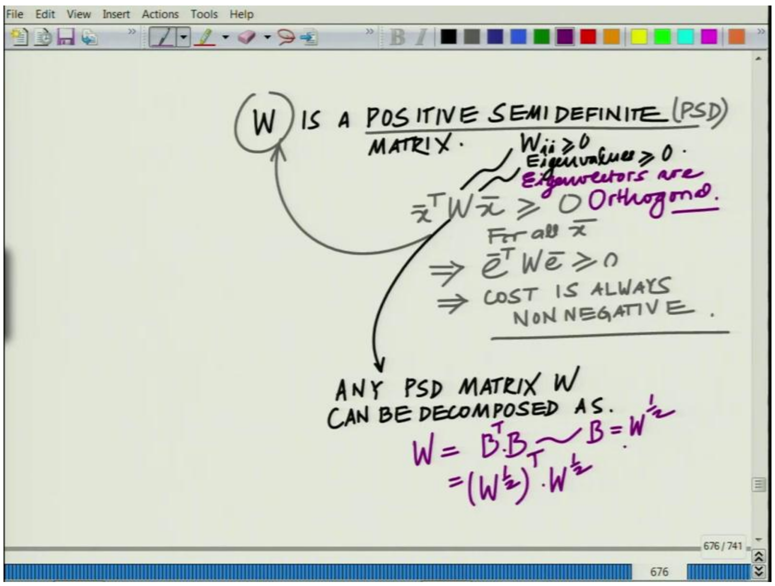
If you are calling this as your W y bar minus, so if you look at this y bar minus Ax bar if you call it as this vector e bar, which you can see this is essentially going to be an m cross 1 vector. So, this y bar minus Ax bar transpose. So, this cost function y bar minus Ax bar transpose W into y bar minus Ax bar that is nothing but e bar transpose W into e bar which is the error vector. And you can write this as, now if you can see, this essentially is sigma equal to 1 to m j equal to 1 to m ei that is ith element of this times ej times Wi times Wij, where this Wij thus is essentially now nothing but your weighting coefficient.

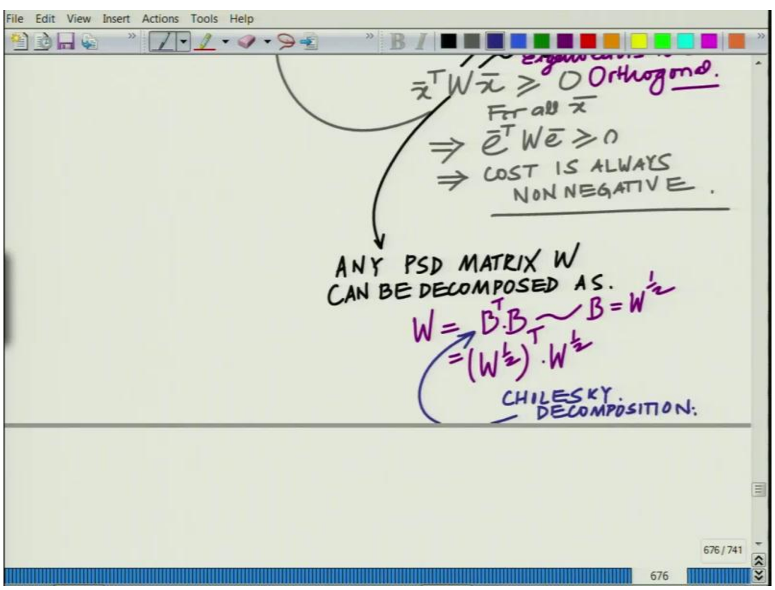
This is your weighting coefficient. And therefore, you are weighting the errors that is essentially what you are doing you are not considering cost, you are not considering all the weights all the errors to be equal, but you are weighing some errors more and some errors, that is you are giving more importance to some errors and less weightage to some errors, that is why it comes a weighted least squares.

Now, if Wij equal to 1 for, if you set the weighting matrix is the Wij equal to 1 for i equal to j is 0 for i not equal to j then this becomes reduces to your normal this squares. Then, then for this kind of scenario reduces to your conventional, you can see. Because this only survives when i call to j, i equal to j and for that it is 1 when i equal to j. So, this becomes summation magnitude (())(08:16) square which is nothing but norm e bar square.

So, this reduces the LS when the, or when you can say W is the identity matrix that is a weighting matrix is the diagonal matrix with all ones, then it reduces to your least, your conventional least squares.

(Refer Slide Time: 08:46)





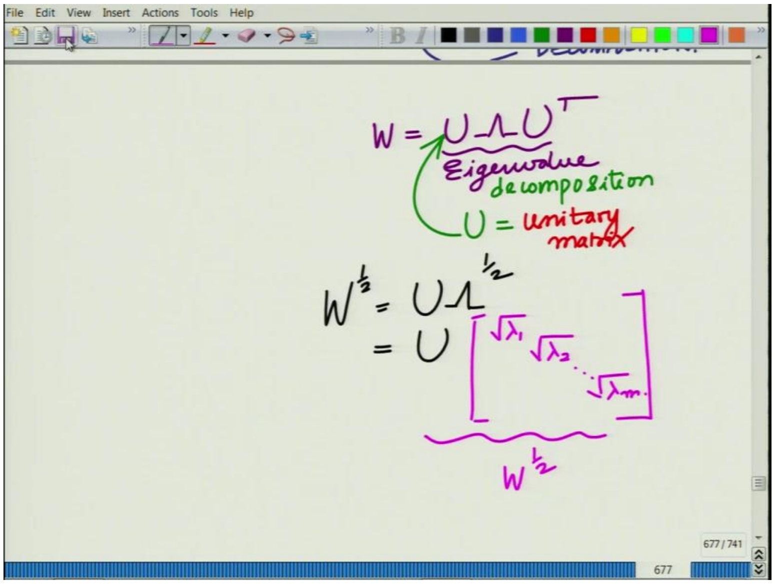
Now, now, another important criterion for W is you cannot take any W because e transpose W e bar might not be positive. So, we choose W which is a, W is a positive semi definite. So, W is a, is restricted to be is a positive, W is a positive semi definite matrix. We also abbreviate this as PSD, this implies, what is the meaning of Positive Semi Definite matrix, this implies x bar transpose Wx bar greater than or equal to 0 for all values of x bar, this also implies therefore, this also implies that when you look at the error e bar transpose W e bar is always greater than equal to 0 implies the cost is always non negative.

The cost is always, just like the least squares non, the cost is always non negative, because this is important because we are looking at weighing the errors of the squares of the errors the cost has to be non negative. And therefore, we choose this to be a positive semi definite matrix such that the cost of this weighted least squares is always non negative.

Now, the of course, positive semi definite matrices satisfy many properties, the diagonal elements Wii have to be greater than or equal to 0, Eigen values have to be greater than or equal to 0, Eigen vectors corresponding to distinct Eigen values are orthogonal, the Eigen vectors are orthogonal and so on. Now, also any positive semi definite matrix if you look at this satisfies, must satisfy an important property that is if you look at any positive semi definite matrix, any PSD matrix W can be decomposed as, any PSD matrix W this can be decomposed as W equals W half.

So, W equals, you can write it as W equals, let us see, you can write, always write it as W equals B, some matrix B times or some matrix B transpose times B and B is in fact known as square root of W. So, you can denote B as W half. So, you can write it as W half square root of W transpose times W half, this is known as the Cholesky decomposition, this is known as, this possible for any PSD matrix. So, this is known as the Cholesky decomposition or simply the matrix square root.

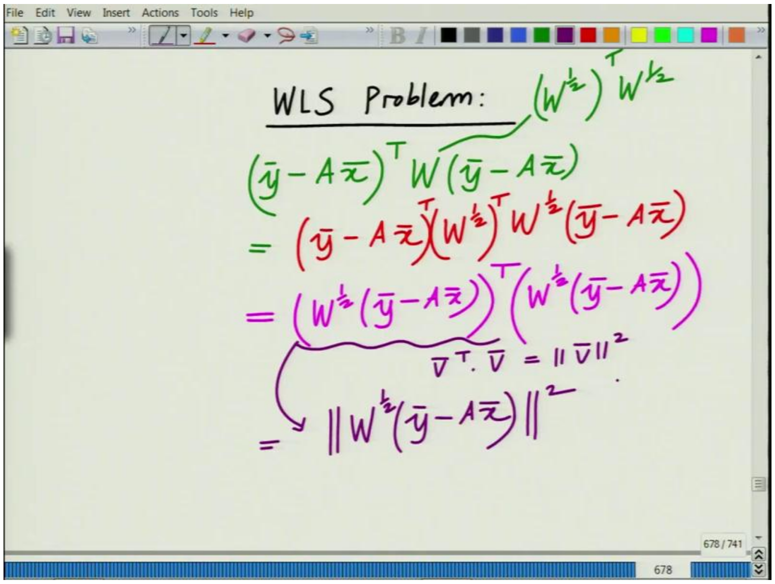
(Refer Slide Time: 12:48)

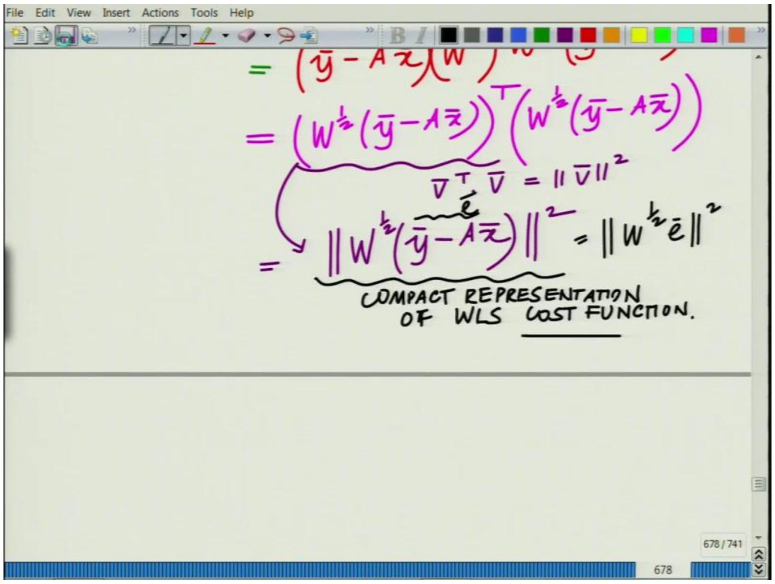


And this can also be obtained as follows, that is if we know that the for a positive semi definite matrix the Eigen value decomposition is given as U lambda U transpose this is the Eigen value decomposition. This is your Eigen value, where U is a unitary matrix, U is a unitary matrix, where U is a unitary matrix. This implies I can write W half as U lambda to the power of half that is U (take) times you take the lambda which is the diagonal matrix of Eigen values and simply take.

So, so this is U square root of lambda 1 square root of lambda 2 square root of W is an m cross n matrix, so this will be square root of lambda m. So, this is essentially your W, this is essentially your W half. So, this is what you have. And so, we use the property that this such a weighting matrix which is a positive semi definite matrix can be decomposed as follows that is W half transpose times W half, this is known as the, the Cholesky decomposition of the positive semi definite matrix.

(Refer Slide Time: 14:37)





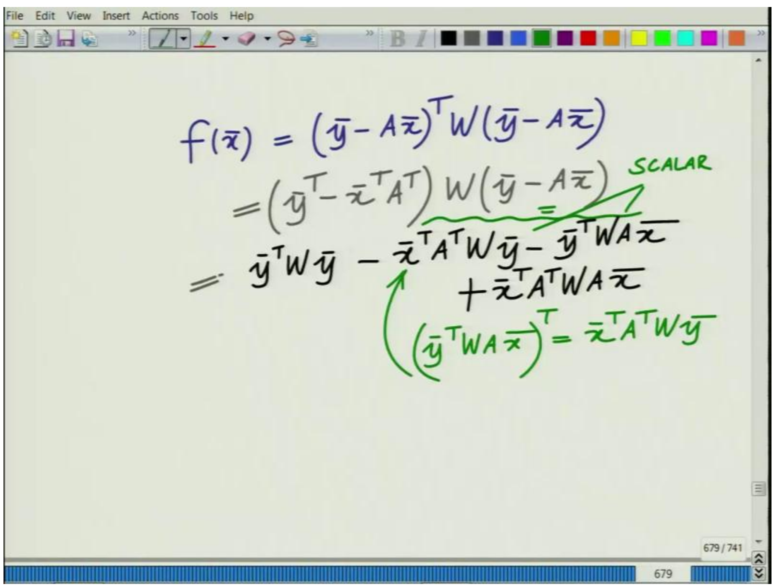
Now, we rewrite the weighted least squares. Now, let us back, get back to our weighted least squares problem and that can be written as follows. The weighted least squares problem is you have your W bar minus Ax bar transpose W into W y bar minus Ax bar. I can now use the decomposition this is W half transpose times W half.

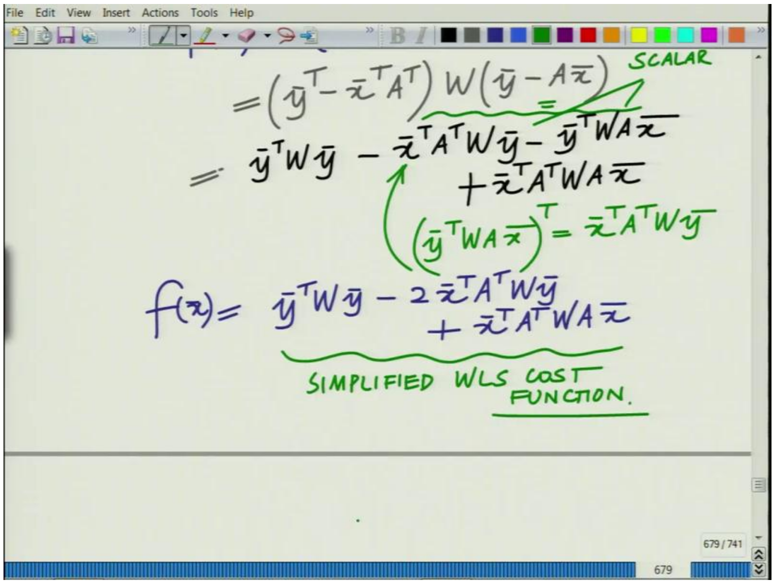
And therefore, this is equal to y bar minus Ax bar times W half transpose times W half y bar minus Ax bar which is equal to, now I can write this as W half y bar minus Ax bar transpose times W half into y bar minus Ax bar which is now you can see this is essentially, essentially vector v bar transpose v bar which is norm v bar square.

So, I can write this as, this is equal to essentially norm of the weighting matrix W half times y bar minus Ax bar whole square. So now you have the, now this is essentially your weighted least square, so this the compact representation of your WLS cost function, this is a compact representation of your weighted least squares cost function.

And that is norm of W half y bar minus Ax bar whole square, you are representing it as the norm square of the error, but this is now you can think of this as the weighted error. So it is exactly, so this is essentially what was your previous, what was your norm e bar square. So, you can instead of looking directly at norm of e bar square what you are looking at is norm W half into e bar square, that is essentially what you are looking at.

(Refer Slide Time: 17:41)





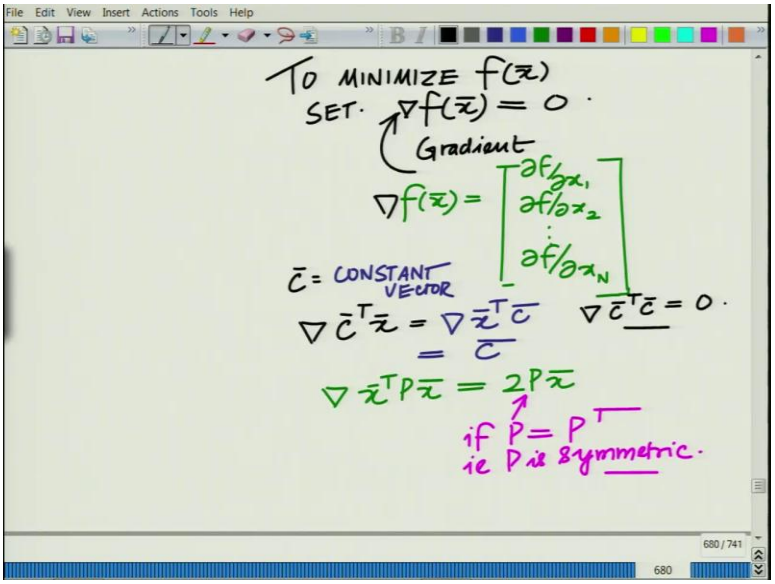
Let us now expand this cost function, that is f of x bar I can write it as y bar minus Ax bar transpose W into y bar minus Ax bar. And this I can simplify it as follows, I can simplify this as y bar transpose minus x bar transpose A transpose into W into y bar minus Ax bar. And now, if you write out the terms, this is going to be y bar transpose W y bar minus x bar transpose A transpose W y bar minus y bar transpose W Ax bar plus x bar transpose A transpose W Ax bar.

So, this is essentially what you can see. And now, you can see these two terms, these two terms are equal because each term is the transpose of the other that is what you see in your typical least squares expansion that is, this term x bar transpose A transpose W bar is the transpose of the other term you can quickly look at it, you can have y bar transpose W Ax bar and these are scalar quantities.

So, this transpose is equal to x bar transpose A transpose W y bar which is equal to the other quantity. So, these two quantities are equal. And moreover, these are scalar quantities, so these are, note that these are scalar quantities implies that they are number. So, when you have number and you take the transpose of its, of the number, it is equal to the transpose because it is a one dimensional quantity, scalar quantity.

And therefore, now and we simplify it. And now therefore, one can write it in this fashion at y bar transpose W y bar minus twice x bar transpose A transpose W y bar plus x bar transpose A transpose W A x x bar transpose A transpose W Ax bar. So, this is your simplified weighted least squares, this is your simplified. So, this is the simplified WLS cost function, this is the simplified weighted least squares cost function. And now, in order to minimize this, we take, we have to find the x which minimizes x bar which minimizes this, take the gradient and set it equal to 0.

(Refer Slide Time: 21:33)

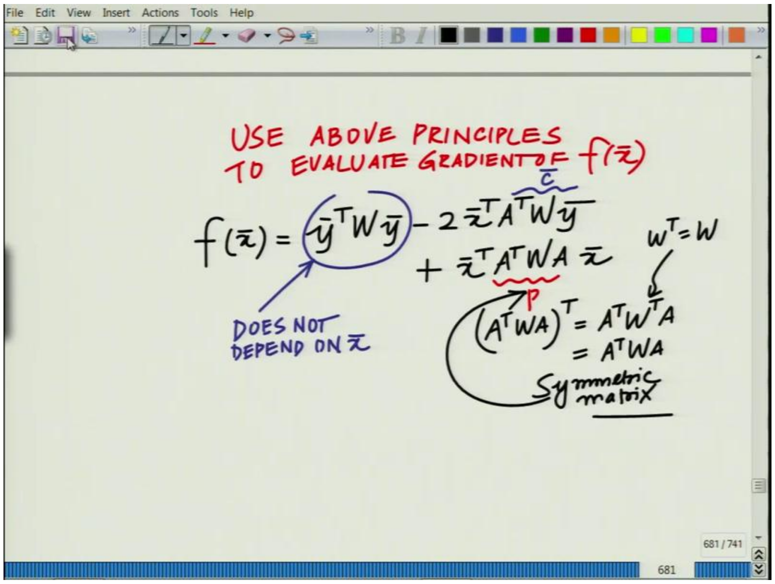


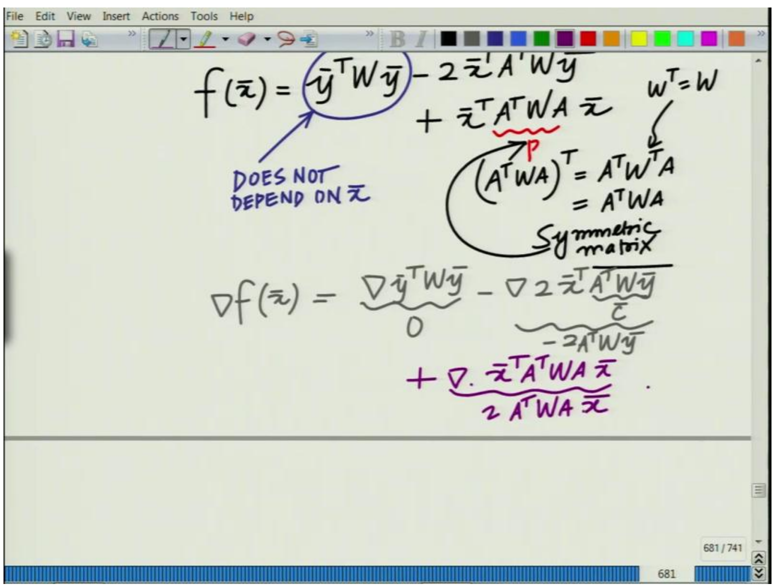
To minimize the cost, to minimize f of x to minimize set the gradient f of x bar equal to 0 remember this is the gradient which is defined as, essentially if you have the gradient n dimensional vector x you take the partial with respect to each component of x. This is something that we have seen, but you might do well to recall that this is the definition of the, this is our definition, this is the definition of the gradient.

So, you take the gradient of this f of x bar and set it equal to 0, we have already seen some principles of how to evaluate the gradient. For instance, let us say we have a constant vector delta c bar transpose x bar. So, c bar equals a constant vector that does not depend on x delta c bar are gradient of c bar transpose x bar equals gradient of x bar transpose c bar, since both these quantities are equal this is equal to c bar.

Further if you have a symmetric matrix gradient of x bar transpose P x bar equal to twice P x bar if, but remember this holds only if P equal to P transpose that is, P is a symmetric matrix. And of course, it goes without saying that if you simply have a constant vector that is a gradient of c bar transpose c bar this will be equal to 0. So, if you have constant vector that does not depend on x bar if you take the derivative, partial derivative with respect to each component is 0, so the gradient is 0.

(Refer Slide Time: 23:56)





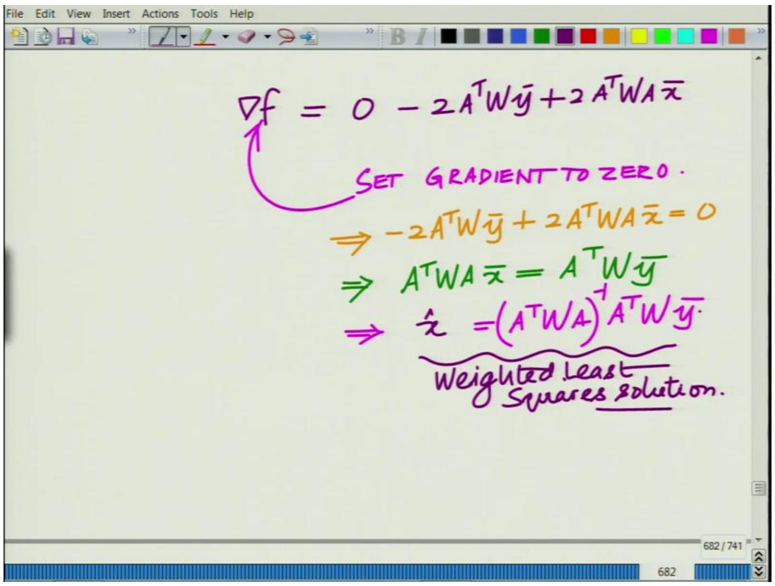
Now, we use these principles, now use this principles, now use these principles to evaluate gradient of the WLS cost. So, use above principles of the x bar which is the, weighted least cost function. Remember we have f of x bar equals y bar transpose W y bar minus twice x bar transpose A transpose W y bar plus x bar transpose A transpose W A into y bar. Now, you look at this, this is essentially your constant, does not depend on x bar implies this will go to 0.

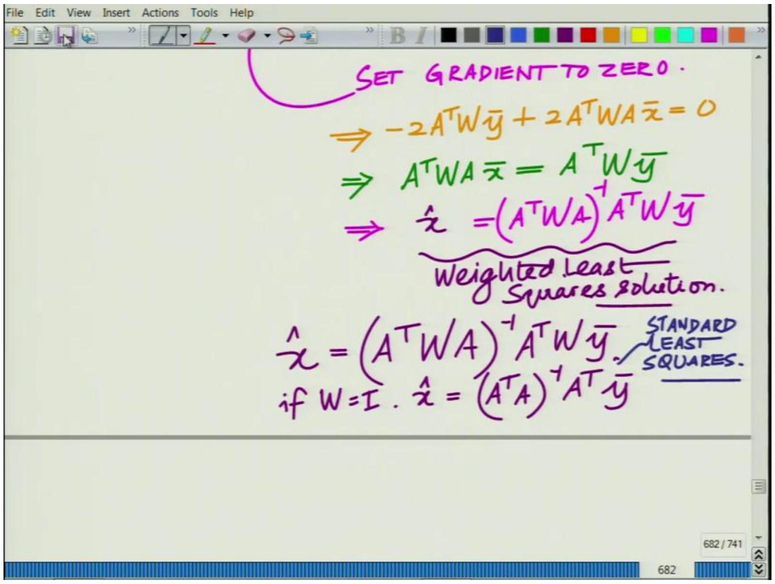
And this is of the form x bar transpose c bar and this is of the form x bar transpose P times, I am sorry, x bar transpose P times x bar and this is a symmetric matrix because A transpose W A transpose is nothing but A transpose W transpose A but W transpose is equal to W which is equal to A transpose W into A. So, W transpose equal to W because this is a symmetric positive semi definite matrix.

Therefore, this implies that this is symmetric. So, this implies that this P this is a symmetric matrix, P is a symmetric matrix. And therefore, now if you compute the gradient, evaluate the gradient. So, I take the gradient of f, so I take the gradient of this f of x bar that is equal to the gradient of y bar transpose W y bar which is 0 minus twice the gradient of minus the gradient of twice x bar transpose A transpose W into y bar.

This is, we already seen this is your c bar. So, this will be essentially minus twice c bar that is minus twice A bar transpose A transpose W into y bar. And we have plus the gradient of x bar transpose A transpose W A x bar A transpose W a symmetric matrix. So, this thing is going to be twice A transpose W A into x bar.

(Refer Slide Time: 27:44)





So, the gradient reduces to 0 minus twice A transpose W y bar plus twice A transpose W Ax bar. So, we set gradient to 0, so set gradient to 0. So, we set the gradient to 0, this implies that minus 2 A transpose W y bar plus 2 A transpose W Ax bar equal to 0 which implies A transpose W Ax bar equals A transpose W y bar which essentially implies that now the least square solution is given as x bar equal to A transpose W A inverse A transpose W y bar.

So, this is the weighted least square solution. So, this is your, this is essentially your and we can call this as x hat because, this is the x hat which minimizes the least square solution this is the weighted least squares, this is your weighted least. This is your WLS solution, which is essentially if you have to write it, let me write it a little prominently.

This is your x hat equals A transpose W A inverse A transpose W y bar. So, this is essentially the inverse of the least squares solution. So, this is essentially the solution of the least square solution where W is your weighting matrix. And now you can see if W equal to identity that is the conventional least squares, now if W equal to, then you again get back the previous solution which is A transpose A inverse A transpose y bar this is your conventional, you can call this as your standard least squares, SLS, this is your standard least squares.

So, you they had the W which is essentially the weighting matrix and that is essentially what makes this solution different and you have the option of weighting the different errors differently rather than the straightforward, the standard least squares where all the errors, error terms are weighted equally. So, this gives the option the flexibility of weighing or attributing different levels of importance to different errors which can occur for various reasons as we will go into, as we will see in an example that we will study next. So, let us end this, stop this module here and continue our discussion in the subsequent modules. Thank you very much.