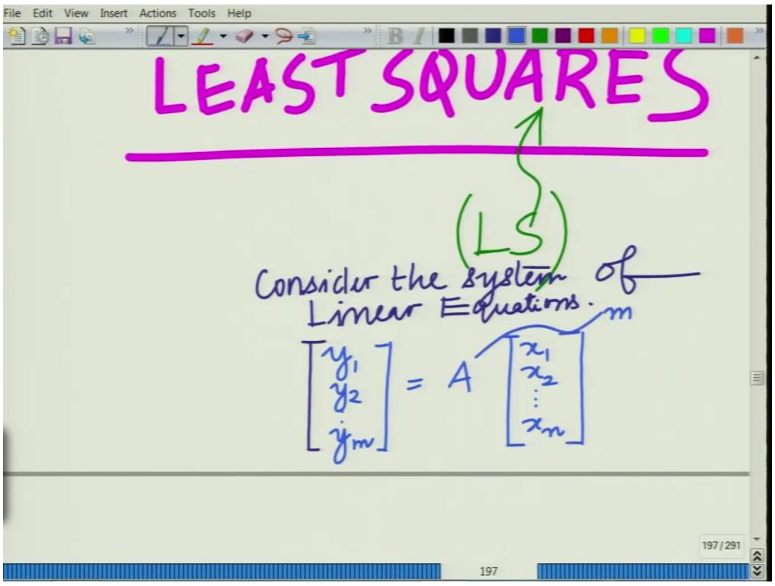
**Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning**
**Professor Aditya K Jagannatham**
**Department of Electrical Engineering**
**Indian Institute of Technology Kanpur**
**Lecture 21**
**Least Squares (LS) solution, pseudo-inverse concept**

Hello. Welcome to another module in this massive open online course. In this module, let us start looking at another very important topic in linear algebra and that is of the least squares solution.

(Refer Slide Time: 0:30)

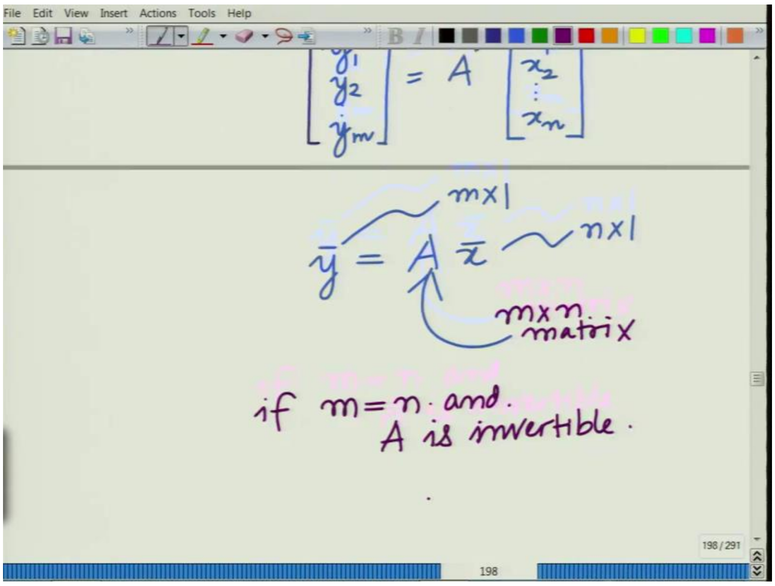



Alright? So, this is another very, very important concept that has many significant implications. This is what is called as the least square. It is often abbreviated simply as the LS, LS for least square. This is known as the least square solution. Now, to understand this, consider the system of linear equations. Let us say we have this system of linear equations, so

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \mathbf{A} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

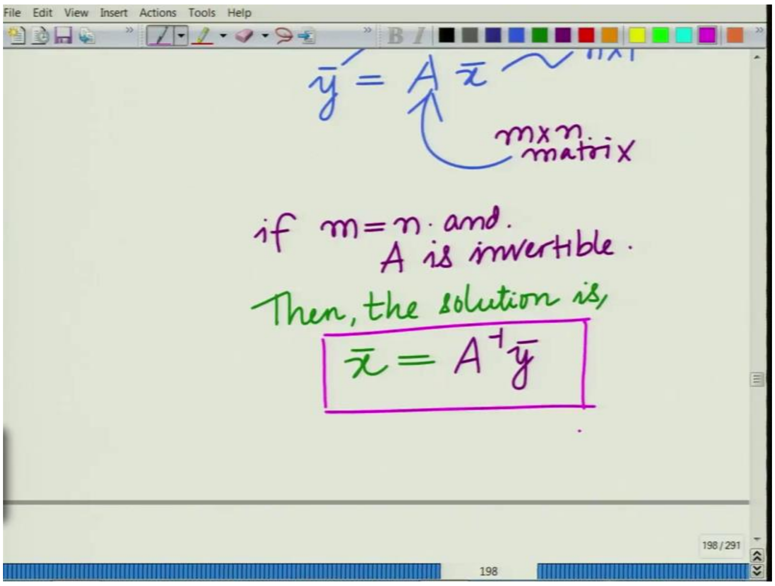
This is our system of linear equations where $\mathbf{A}$ is naturally an $m \times n$ matrix.

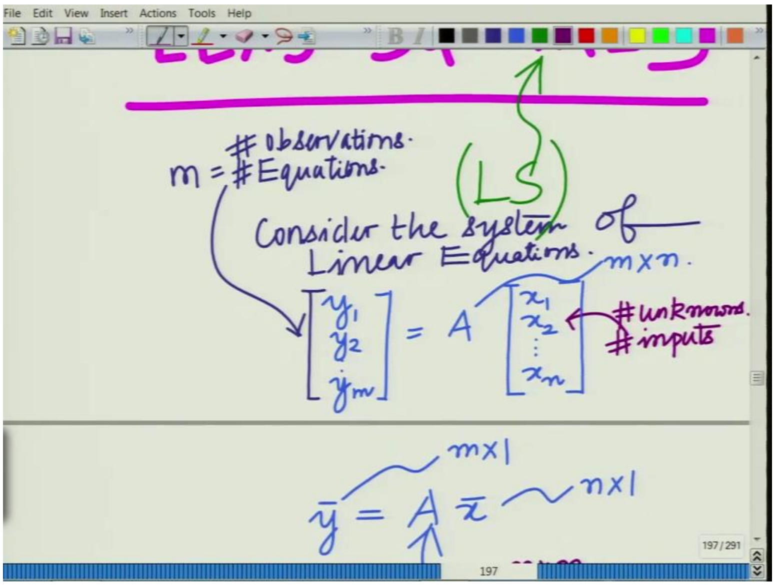
(Refer Slide Time: 2:00)




So, you can write this as $\bar{\mathbf{y}} = \mathbf{A}\bar{\mathbf{x}}$, where $\bar{\mathbf{y}}$ is $m \times 1$ vector, $\bar{\mathbf{x}}$ is $n \times 1$ and $\mathbf{A}$, therefore naturally, which maps the $n \times 1$ vector $\bar{\mathbf{x}}$ to the $m \times 1$ vector $\bar{\mathbf{y}}$, and is an $m \times n$ matrix. Now point here is, we have seen if $m = n$ and $\mathbf{A}$ is a square matrix. $\mathbf{A}$ is a square matrix only then you can talk about the inverse but even then, it is not guaranteed. Any square matrix need not be invertible. Only if $\mathbf{A}$ is not singular then $\mathbf{A}$ is invertible, the inverse exists.

(Refer Slide Time: 3:01)



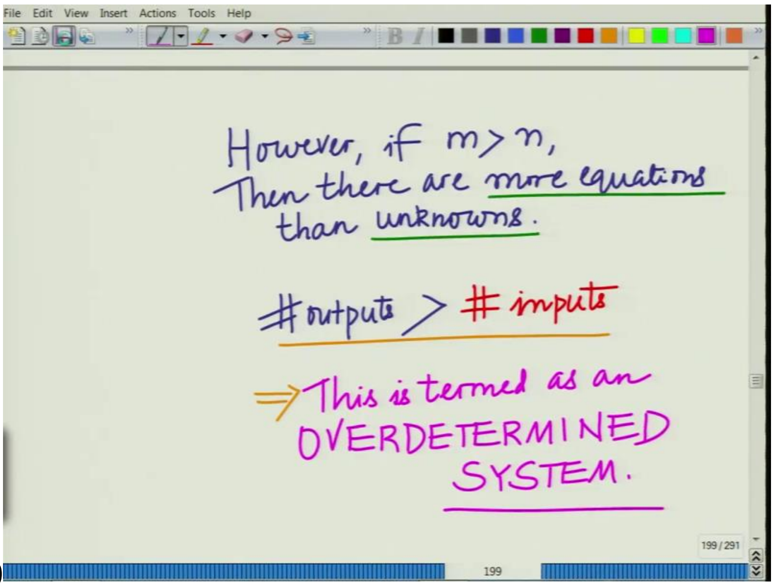



Then the solution is given as $\bar{\mathbf{x}} = \mathbf{A}^{-1}\bar{\mathbf{y}}$.

This is the solution, if **A** is invertible, only if **A** is invertible. But however, now you see in this system, if you look at this, what is $m$, $m$ is equal to the number of equations or number of observations. So, $m$ equal to number of equations. You can also think of this in a practical system as the number of outputs.

In your wireless communication system, $\bar{\mathbf{y}}$ are the number of outputs or observations. And what is $\bar{\mathbf{x}}$, these are the number of unknowns, or the number of transmit symbols or the number of inputs. In your practical multiple-input multiple-output wireless communication system that we have seen before, $x_1, x_2, .., x_n$,these are the transmitted symbols. Now if $m$ is greater than $n$ then there are more equations.

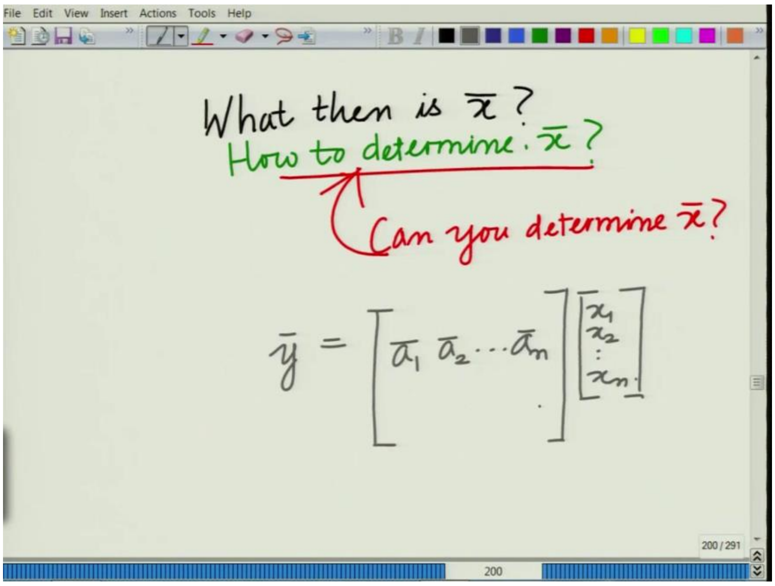
(Refer Slide Time: 4:23)

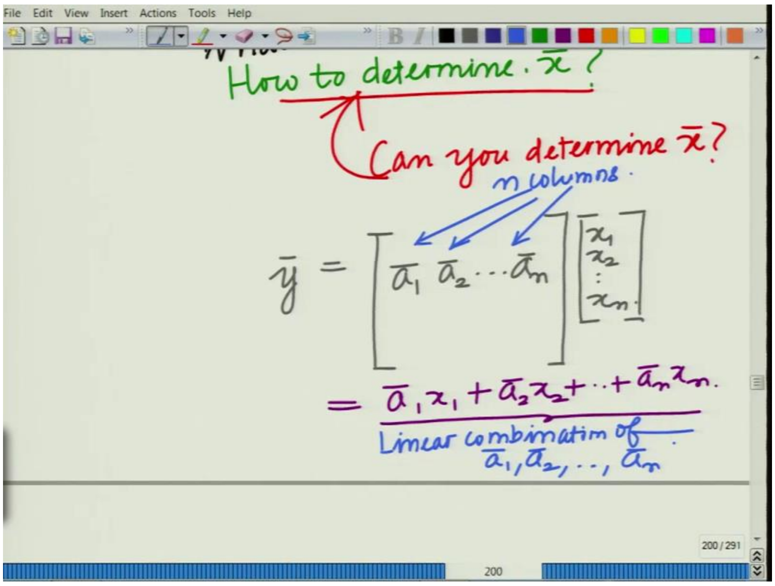


However, if $m$ is strictly greater than $n$, then there are more equations than unknowns. That is number of observations greater than number of inputs, speaking in terms of a linear system as an input-output system, number of outputs is greater than the number of inputs. In such a situation, we call this, this is known as an overdetermined system. This is termed as an overdetermined system.

Such a system of linear equations where the number of equations $m$ that is the dimension of $\bar{\mathbf{y}}$ is greater than the number of unknowns $n$, that is the dimension of $\bar{\mathbf{x}}$. This is known as an overdetermined system. Now what can we do now. No, naturally $m$ is greater than $n$. **A** is not square. This implies that of course invertibility of **A** is out of the question because, we talk about invertibility only if **A** is a square matrix and that too if **A** is not singular. So now how do you solve this system? So, what is the solution?

(Refer Slide Time: 6:43)



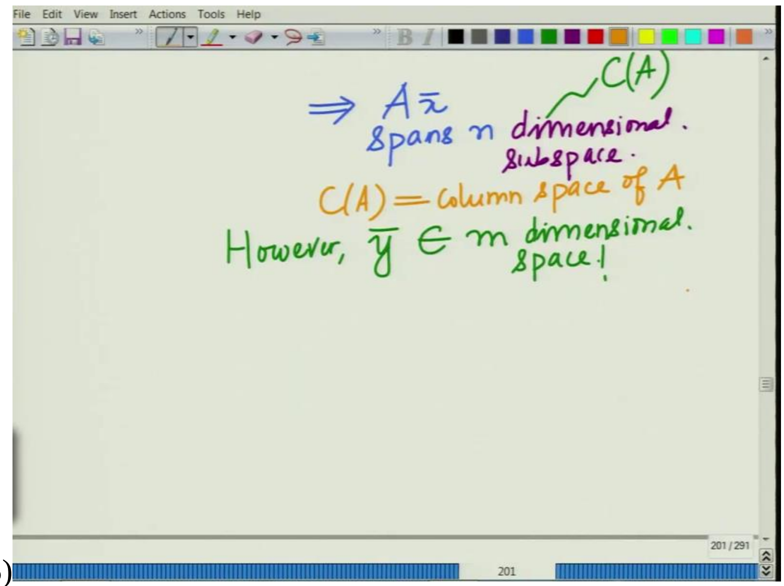



So, what then is $\bar{\mathbf{x}}$ or how do you determine $\bar{\mathbf{x}}$. First of all, can you determine $\bar{\mathbf{x}}$, I think that is the more important question. The first thing, even before how to determine $\bar{\mathbf{x}}$? I think the more relevant question is can you determine $\bar{\mathbf{x}}$? Does not $\bar{\mathbf{x}}$ exist to satisfy this? That is an important question. And you will see if you look at that system again consider this $\bar{\mathbf{y}}$ equal to, you write this in terms of this column,

$$\bar{\mathbf{y}} = [\bar{\boldsymbol{a}}_1 \quad \bar{\boldsymbol{a}}_2 \quad \dots \quad \bar{\boldsymbol{a}}_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \bar{\boldsymbol{a}}_1 x_1 + \bar{\boldsymbol{a}}_2 x_2 + \cdots + \bar{\boldsymbol{a}}_n x_n.$$

This span, if you look at this, this is a linear combination. So, if you look at the matrix **A**, this has $n$ columns, and because this $\bar{\mathbf{y}}$ is a linear combination of $n$ columns of the matrix **A**, implies this spans an $n$ dimensional subspace.

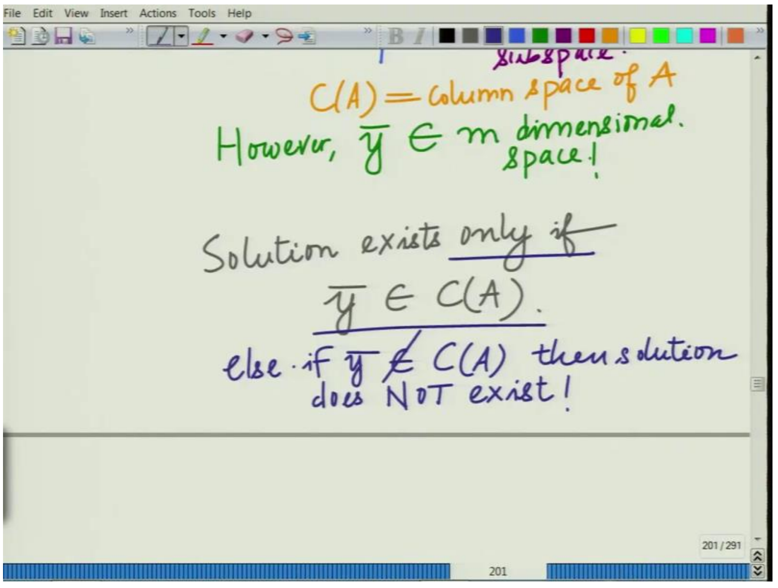
(Refer Slide Time: 9:05)



Implies $\mathbf{A\bar{x}}$ spans an $n$ dimensional subspace. This spans an $n$ dimensional subspace at most, if the columns are linearly independent. But, let us say for simplicity, let us assume that $\mathbf{A}$ is full rank, the maximum rank because $m$ greater than $n$ implies that the maximum rank is $n$. However, $\bar{\mathbf{y}}$ belongs to an $m$ dimensional space. It can be any vector in an $m$ dimensional space, implies so if you look at this $n$ dimensional subspace, if you call this as $C(\mathbf{A})$, that is the column space span by the columns of $\mathbf{A}$.

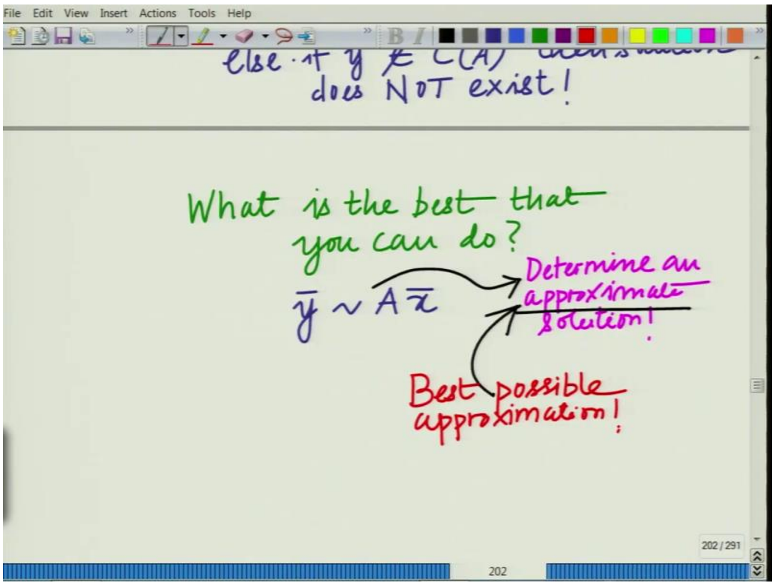
If you call this as $C(\mathbf{A})$, and what is $C(\mathbf{A})$? $C(\mathbf{A})$ equal to column space of $\mathbf{A}$, which is an $n$ dimensional space. Now solution exists. Now it is easy to see when does a solution exists. Solution exists only if this $\bar{\mathbf{y}}$, which can lie general, anywhere in this $m$ dimensional space, the solution exists only if $\bar{\boldsymbol{y}}$ belongs to this $n$ dimensional subspace that is a column space of $\mathbf{A}$. Otherwise, the solution does not exist.

(Refer Slide Time: 10:51)




Otherwise, solution does not exist, else this has no solution. That is, you cannot determine an $\bar{\mathbf{x}}$ which satisfies the condition $\bar{\boldsymbol{y}} = \mathbf{A\bar{x}}$ because any $\mathbf{A\bar{x}}$ has to lie in the span of the columns of $\mathbf{A}$ and $\bar{\boldsymbol{y}}$ lies outside this $n$ dimensional subspace.

(Refer Slide Time: 12:18)

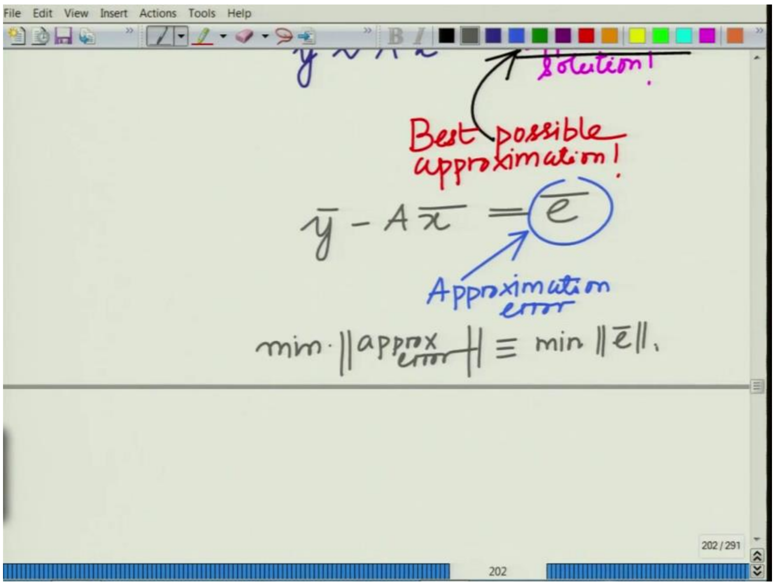



Now therefore, if you cannot determine a $\bar{\boldsymbol{y}}$ then what is the best thing you can do? Now you can ask the question as an engineer, in particular, we do not give up when we cannot find an exact solution, because we do not give up, what is the best. Many times, in life we cannot find the exact solution, then what is the best that we can do?

Now you would like to ask, fine, you cannot find the exact solution $\bar{\boldsymbol{y}} = \mathbf{A}\bar{\mathbf{x}}$. That does not mean that you have to give up. What is the best that you can do and the best that you can do is find an $\bar{\mathbf{x}}$ such that $\bar{\boldsymbol{y}}$ is approximately equal to $\mathbf{A}\bar{\mathbf{x}}$. So now you change from exact to $\bar{\boldsymbol{y}}$ is approximately equal to $\mathbf{A}\bar{\mathbf{x}}$. So, it is a novel concept, if you cannot determine an exact solution, determine an approximate solution. And we want to find the best approximation.

Remember, naturally as engineers and smart people in general, you would like to determine the approximate, I mean any solution is an approximate solution. I mean you can give $\bar{\mathbf{x}}$ equal to $0, 0,$ the all $0$ vector. So, we just randomly generate a vector but you would like to find the best possible approximation. Therein lies the ingenuity. So therefore, now we would like to find the approximate solution but what do we mean by approximate solution, we would like to qualify this and we would like to find the best possible approximation. Not any approximation, but the best possible approximation.

(Refer Slide Time: 14:19)

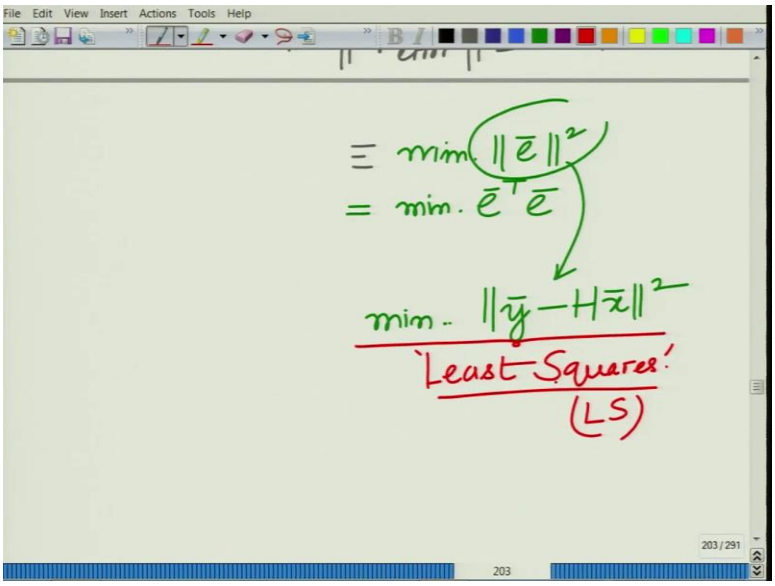



What do we mean by the best possible approximation, that is if you look at $\bar{\boldsymbol{y}} = \mathbf{A}\bar{\mathbf{x}}$ and call that as $\bar{\boldsymbol{e}}$, that is

$$\bar{\mathbf{e}} = \bar{\mathbf{y}} - \mathbf{A}\bar{\mathbf{x}}.$$

this is the approximation error. Then we would like to minimize the approximation error but remember this $\bar{\mathbf{e}}$ is a vector so I cannot minimize the approximation error. It does not make sense to say we are going to minimize a vector. We can minimize the norm of this vector, the length of this vector or minimize the norm of the approximation error which is essentially minimize norm of $\bar{\mathbf{e}}$.

(Refer Slide Time: 15:26)

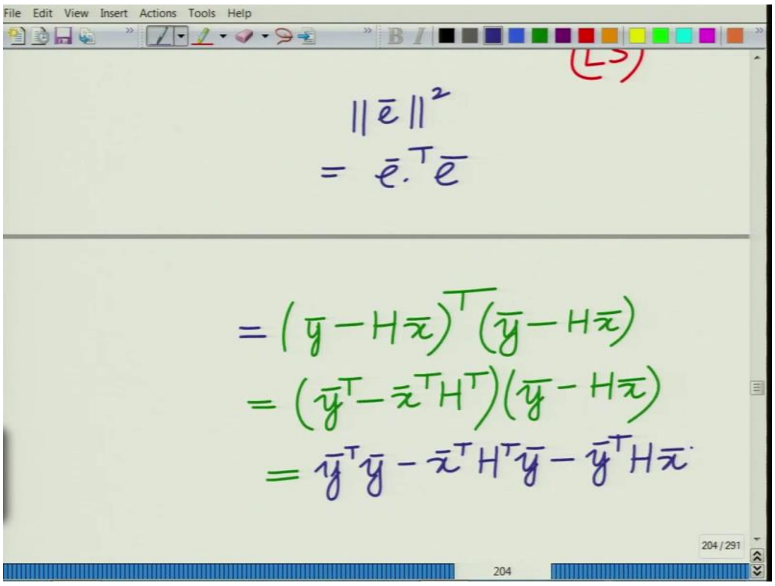


That is the problem that we want to formulate, which is essentially also similar to saying minimize the $||\bar{\boldsymbol{e}}||^2$, which is equal to minimize $\bar{\boldsymbol{e}}^T\bar{\boldsymbol{e}}$. So, we would like to minimize the $||\bar{\boldsymbol{e}}||^2$. That is if you look at this quantity that is nothing but

$$\min\left|\left|\bar{\mathbf{y}} - \mathbf{A}\bar{\mathbf{x}}\right|\right|^2$$

and this is known as the least squares problem. Find the vector $\bar{\mathbf{x}}$ such that the error $\bar{\mathbf{y}} - \mathbf{A}\bar{\mathbf{x}}$, if you look at the norm of the error and square of the norm of the error, that is the least. Therefore, this is known as the least square problem. So, this is a very popular problem in entire signal processing, machine learning, data analysis. I mean, whatever you look at it and this is basically known as a, this is a very fundamental problem this is essentially the least square problem which is essentially the title of our module. This is in case you are wondering what is the least squares problem. This is the least squares problem (LS) and now we want to find the solution to this problem that is the least squares solution.

(Refer Slide Time: 17:05)




Naturally, this is an optimization problem so let us first simplify the objective. So

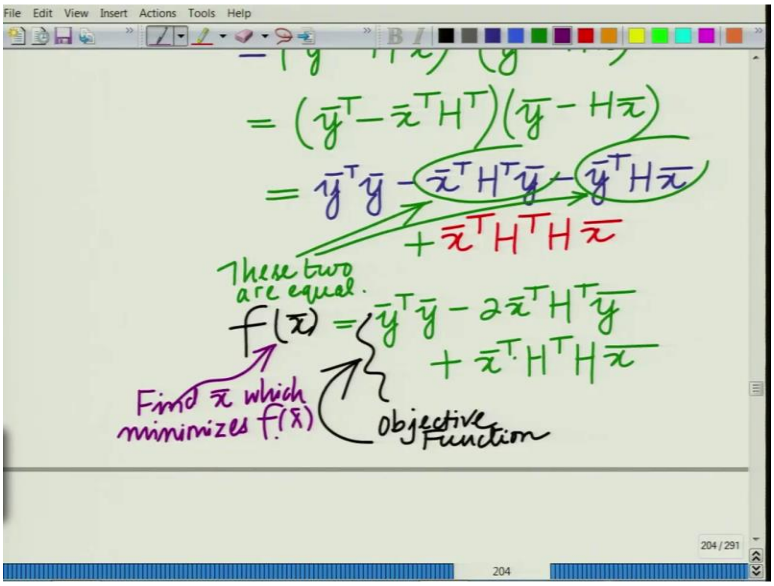
$$||\bar{\boldsymbol{e}}||^2 = \bar{\boldsymbol{e}}^T\bar{\boldsymbol{e}}$$

$$= (\bar{\mathbf{y}} - \mathbf{A}\bar{\mathbf{x}})^T(\bar{\mathbf{y}} - \mathbf{A}\bar{\mathbf{x}}) = (\bar{\boldsymbol{y}}^T - \bar{\mathbf{x}}^T\mathbf{A}^{\mathrm{T}})(\bar{\mathbf{y}} - \mathbf{A}\bar{\mathbf{x}})$$

$$= \bar{\mathbf{y}}^{\mathrm{T}}\bar{\mathbf{y}} - \bar{\mathbf{x}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\bar{\boldsymbol{y}} - \bar{\mathbf{y}}^{\mathrm{T}}\mathbf{A}\bar{\mathbf{x}} + \bar{\mathbf{x}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\bar{\mathbf{x}}.$$

The terms $\bar{\mathbf{x}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\bar{\boldsymbol{y}}$ and $\bar{\mathbf{y}}^{\mathrm{T}}\mathbf{A}\bar{\mathbf{x}}$ are equal, note that these two quantities are transpose of each other. These two are equal because they are the transpose of each other. So, you can write this as

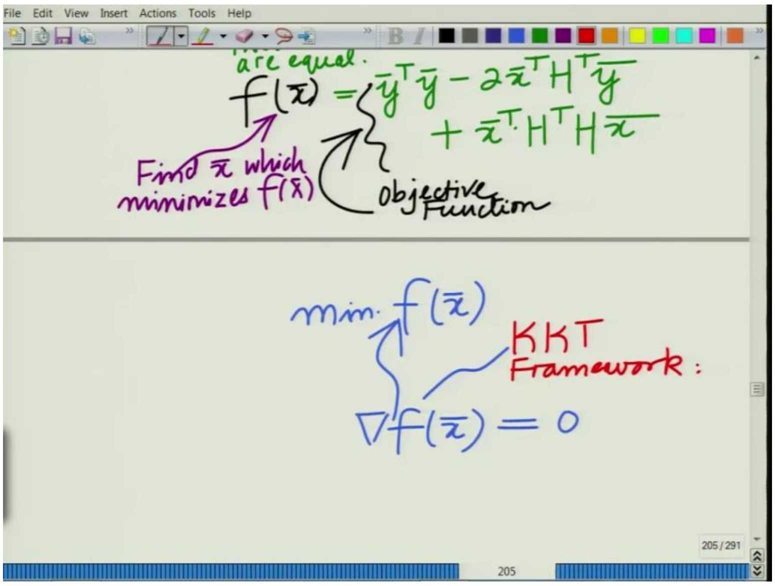
$$f(\bar{\boldsymbol{x}}) = \bar{\mathbf{y}}^{\mathrm{T}}\bar{\mathbf{y}} - 2\bar{\mathbf{x}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\bar{\mathbf{y}} + \bar{\mathbf{x}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\bar{\mathbf{x}}.$$




.

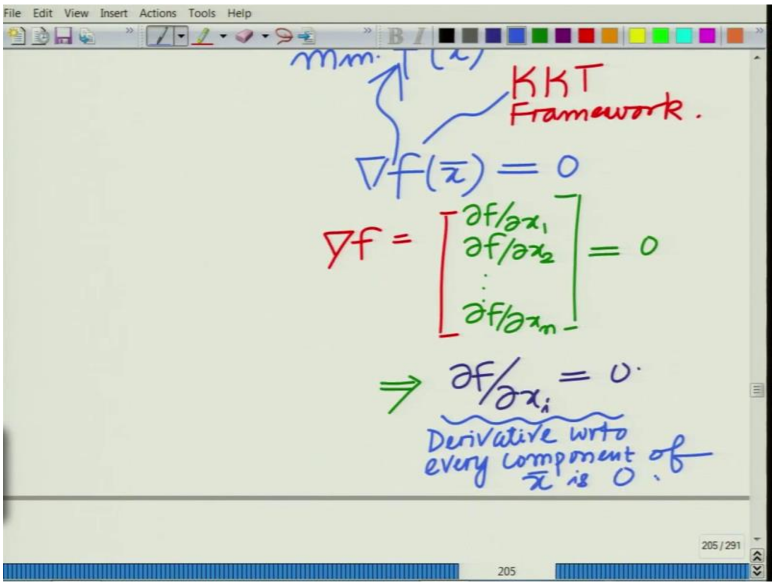
So, this is essentially going to be the objective function or this is the simplified version of the objective function. Now we have to find the minimum. Now we have to find the $\bar{\mathbf{x}}$. which minimizes this objective function, alright? The task is very simple. Find $\bar{\mathbf{x}}$ which minimizes $f(\bar{\mathbf{x}})$ and the way to do is a very straightforward procedure which is known as the Karush–Kuhn–Tucker (KKT) framework.

(Refer Slide Time: 20:18)




So, to minimize, this is like an optimization problem which you are all familiar with $f(\bar{\mathbf{x}})$ compute the gradient with respect to $\bar{\mathbf{x}}$ and set it equal to 0 to find what is known as the stationary point or in this case it will of course be a minima. Because this is unbounded. You can clearly take this as close to infinity as possible. So, this is essentially the concept when you find a stationary point, this is going to be a minima. So, you set the gradient equal to 0, this is known as a KKT Framework, Karush–Kuhn–Tucker Framework, a very standard framework in optimization and it is very intuitive. Essentially for a scalar function, what you are doing is? you are computing the derivative and setting it equal to 0, except now that $\bar{\mathbf{x}}$ is a vector so you have to compute the partial derivative with respect to each component of $\bar{\mathbf{x}}$**,** i.e., $x_1, x_2, \dots, x_n$, and set it equal to 0. So that is what this is.

(Refer Slide Time: 21:20)

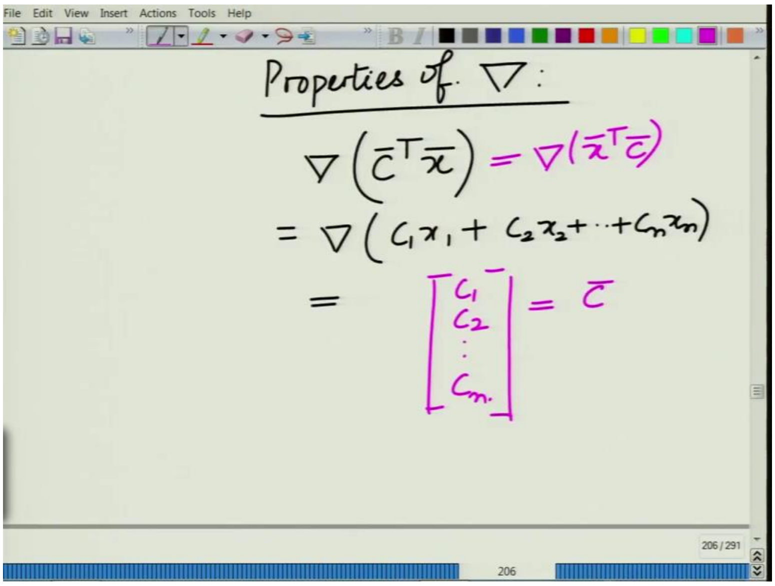



If you look at this, this implies that $\nabla f(\bar{\mathbf{x}})$ is nothing but the vector of partial derivatives, that is

$$\nabla f(\bar{\mathbf{x}}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \mathbf{0} \Rightarrow \frac{\partial f}{\partial x_i} = 0, \forall\, i.$$

Which basically implies that you are setting the partial derivative with respect to every component of $\bar{\mathbf{x}}$ equal to 0. Derivative at the optimum point, derivative with respect to every component of $\bar{\mathbf{x}}$ is equal to 0.

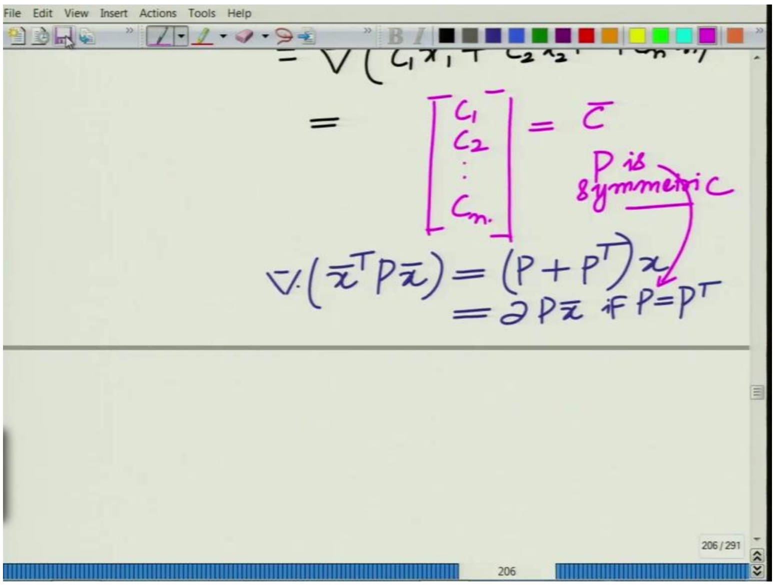
(Refer Slide Time: 22:31)

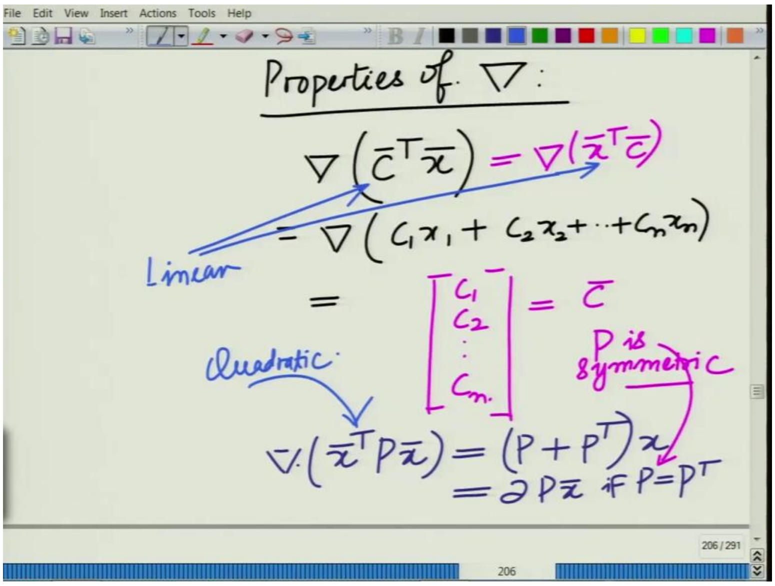



Let us see how to do this. Consider, let us look at some properties of the gradient operator. If you have a constant vector $\bar{\mathbf{c}}$ then

$$\nabla(\bar{\mathbf{c}}^T\bar{\mathbf{x}}) = \nabla(\bar{\mathbf{x}}^{\mathbf{T}}\bar{\mathbf{c}}) = \nabla(c_1x_1 + c_2x_2 + \cdots + c_nx_n) = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \bar{\mathbf{c}}.$$

(Refer Slide Time: 23:38)





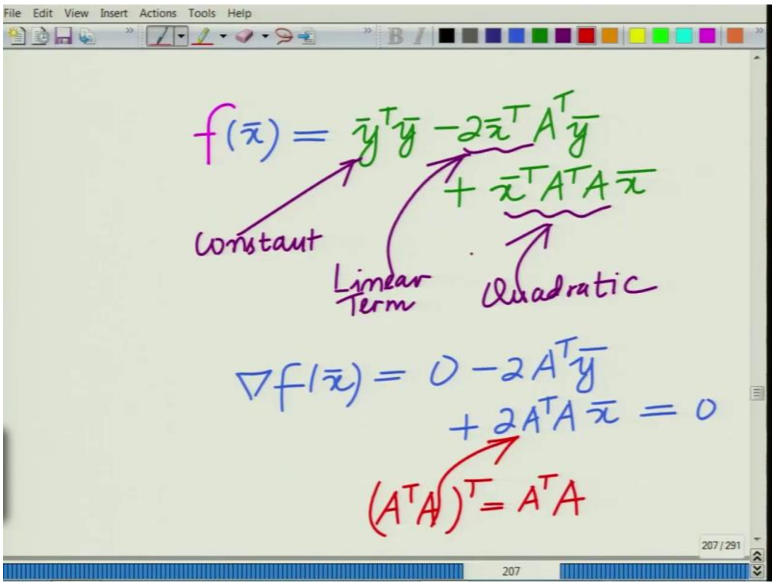

And now if you look at the quadratic function $\bar{\mathbf{x}}^{\mathbf{T}}\mathbf{P}\bar{\mathbf{x}}$, the gradient with respect to $\bar{\mathbf{x}}$, for instance,

$$\nabla(\bar{\mathbf{x}}^{\mathbf{T}}\mathbf{P}\bar{\mathbf{x}}) = (\mathbf{P} + \mathbf{P}^{\mathrm{T}})\bar{\mathbf{x}} = 2\mathbf{P}\bar{\mathbf{x}}, if\ \mathbf{P} = \mathbf{P}^{\mathrm{T}},$$

which implies basically that **P** is a symmetric matrix.

Let us come back now to our objective function. Now let us use these properties. So, $\bar{\mathbf{x}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\bar{\mathbf{y}}$ is what we are calling as a linear. Now, $\bar{\mathbf{x}}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{A}\bar{\mathbf{x}}$ this is the quadratic, second order term and $\bar{\mathbf{y}}^{\mathbf{T}}\bar{\mathbf{y}}$ is a constant term.

(Refer Slide Time: 24:51)



$$+\ \bar{x}^T A^T A \bar{x}$$

Constant

Linear Term

Quadratic

$$\nabla f(\bar{x}) = 0 - 2A^T\bar{y} + 2A^TA\bar{x} = 0$$

$$(A^TA)^T = A^TA$$

$A^TA$ is symmetric.





$$\min \|\text{approx error}\| \equiv \min \|e\|.$$

$$\equiv \min \|\bar{e}\|^2$$

$$= \min \bar{e}^T\bar{e}$$

$$\min \|\bar{y} - A\bar{x}\|^2$$

'Least Squares' (LS)

$$\|\bar{e}\|^2$$

$$= \bar{e}^T\bar{e}$$





$$= (\bar{y} - A\bar{x})^T(\bar{y} - A\bar{x})$$

$$= (\bar{y}^T - \bar{x}^TA^T)(\bar{y} - A\bar{x})$$

$$= \bar{y}^T\bar{y} - \bar{x}^TA^T\bar{y} - \bar{y}^TA\bar{x} + \bar{x}^TA^TA\bar{x}$$

These two are equal.

$$f(\bar{x}) = \bar{y}^T\bar{y} - 2\bar{x}^TA^T\bar{y} + \bar{x}^TH^TH\bar{x}$$

Find $\bar{x}$ which minimizes $f(\bar{x})$

Objective Function

$$f(\bar{x})$$

And in fact, if you look at our objective function, $f(\bar{\mathbf{x}})$, you will notice something interesting. This is essentially equal to

$$f(\bar{\boldsymbol{x}}) = \bar{\mathbf{y}}^{\mathrm{T}}\bar{\mathbf{y}} - 2\bar{\mathbf{x}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\bar{\mathbf{y}} + \bar{\mathbf{x}}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{A}\bar{\mathbf{x}}.$$

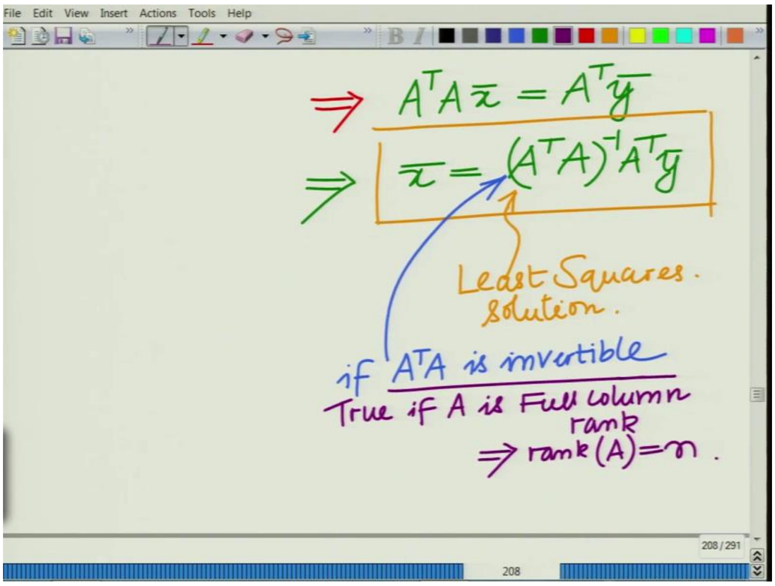
So overall this is a quadratic expression. And therefore now if you find the gradient of this that is the gradient of $f(\bar{\boldsymbol{x}})$ that is equal to

$$\nabla f(\bar{\boldsymbol{x}}) = \nabla(\bar{\mathbf{y}}^{\mathrm{T}}\bar{\mathbf{y}}) - 2\nabla(\bar{\mathbf{x}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\bar{\mathbf{y}}) + \nabla(\bar{\mathbf{x}}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{A}\bar{\mathbf{x}})$$

$$= \mathbf{0} - 2\mathbf{A}^{\mathrm{T}}\bar{\mathbf{y}} + \mathbf{A}^{\mathbf{T}}\mathbf{A}\bar{\mathbf{x}} = \mathbf{0} \Rightarrow \bar{\mathbf{x}} = (\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\bar{\mathbf{y}}.$$

It is basically, you can easily see that $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ is symmetric.

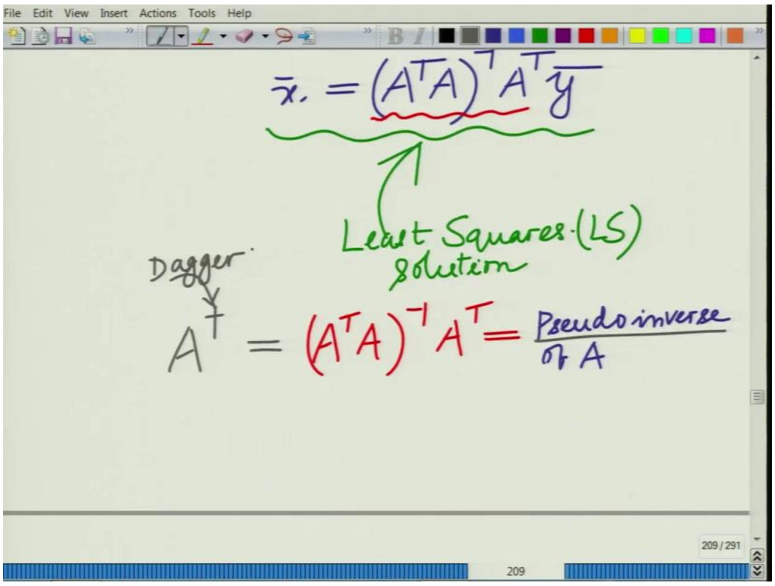
(Refer Slide Time: 28:29)




This is basically your least square solution. So, this is a very interesting solution.

And of course, here we are assuming that $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ is invertible. If $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ is invertible which is not very difficult to show that this exists. This is true if $\mathbf{A}$, we will not prove this explicitly, is full column rank which is that is this implies that $rank(\mathbf{A})$ is equal to $n$. $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ is invertible, whenever the matrix $\mathbf{A}$ is a full column rank. It has a rank equal to the number of columns that is $rank(\mathbf{A})$ is equal to $n$, i.e., the number of columns.

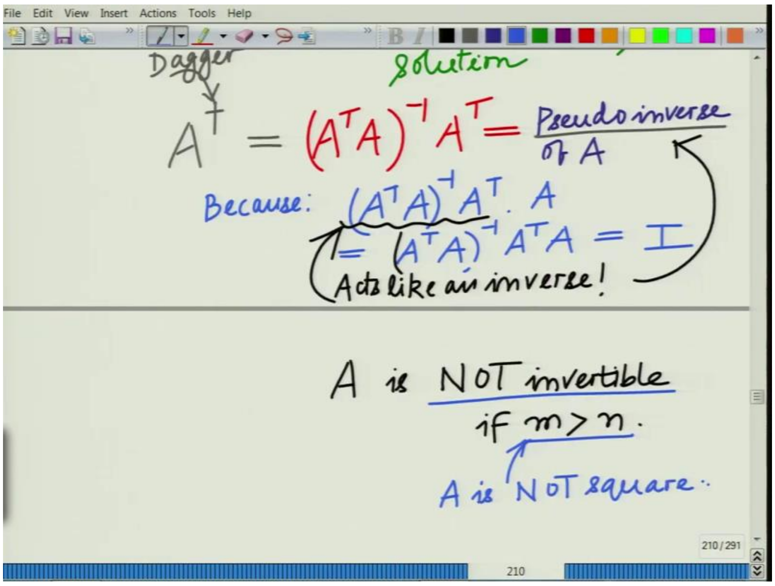
(Refer Slide Time: 30:31)




Now, this is a very interesting solution. Let us look at the solution. This is a very interesting solution. We have the least square solution. The problem is the least square problem, and $(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\bar{\mathbf{y}}$ this is essentially termed the least square solution. This is essentially termed the least square solution or basically as you also know we call the least square abbreviated as LS so this is basically termed as the LS solution, alright?

And $(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}$, this matrix, this is an interesting structure so if you look at this matrix which is this $(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}$. This has a very important meaning. This is termed as the pseudo-inverse. This is the pseudo-inverse of $\mathbf{A}$. This $(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}$ is often denoted by $\mathbf{A}^{\dagger}$, the symbol †, this is what is termed as the dagger.

(Refer Slide Time: 32:01)



This $(\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}$, and this is termed as pseudo-inverse because this has a very interesting property, because, now remember if $m > n$, then **A** is not invertible. But if you look at matrix $(\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}$, and multiply it with **A** on the left

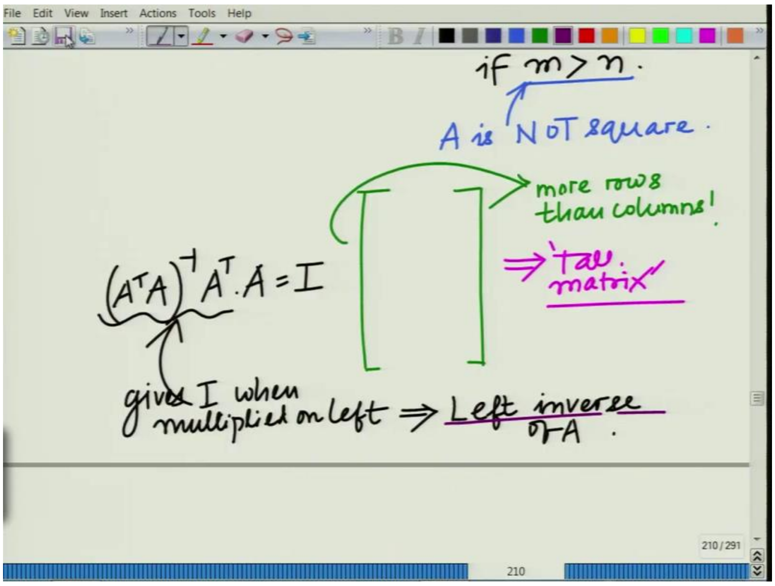
$$(\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\mathbf{A} = \mathbf{I}.$$

So, this acts as an inverse, more specifically it acts as a left inverse of **A**, hence it is termed as pseudo-inverse. So, if $m > n$, strictly speaking, **A** is not invertible but this matrix is acting as an inverse. It is appearing as if it is an inverse of **A** but it is not really an inverse of **A.** You can clearly see because it does not satisfy any properties of the inverse.

For instance, if you multiply it on the right, any inverse, if **AB** is a square matrix which is invertible, and if **AB** is identity then **BA** also has to be identity. But, in this case, if you multiply $(\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}$ on the right of **A**, you will not get the identity. So, in fact that is a very interesting. So, this is only a pseudo-inverse because it appears to be like an inverse, acts like an inverse.

So, this quantity essentially acts like an inverse. That is why this is termed as a pseudo-inverse. It is not an inverse. It acts like an inverse. Because note that **A** is not invertible if $m > n$. This is always important to remember because if $m > n$, **A** is not a square matrix

(Refer Slide Time: 34:20)




In fact, if you look at **A**, this is an interesting structure, if $m > n$, it has more height than width, more rows than columns. This is often termed as tall matrix; it looks very cylinder and tall. This is often termed colloquially in linear algebra; it is often termed as a tall matrix. In engineering and linear algebra this is often known as a tall matrix because it looks like a tall matrix, more rows than columns. This is some, take it with a pinch of humor. This is not something that you use in a formal paper but it is something that is termed as a tall matrix and also because it only gives the identity when it is multiplied on the left. You note that $(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{A} = \mathbf{I}$, that is, it is multiplied on the left, gives and that is why this is termed as a left inverse because on the right it does not give identity.

And in case you are more interested, another interesting fact that you will also realize that this left inverse of $\mathbf{A}$ is not unique. If $m > n$, then left inverse is not unique. The pseudo-inverse is one of the left inverses. So, there might be some confusion especially if you are seeing this for the first time because there is an inverse and there is a pseudo-inverse, but it is very simple. The pseudo-inverse as the name implies is not an inverse. If $m > n$ then this is only a pseudo-inverse.

Now the interesting point is if $m = n$ and $\mathbf{A}$ is invertible, then pseudo-inverse reduces to the inverse. For square matrices again you might be confused which one to use? should I use the pseudo-inverse? or should I use the inverse? Does not matter. For a square matrix, if the inverse exists, pseudo-inverse equals the inverse and that you can also see in a straightforward fashion. So, it is a very interesting concept, a very powerful, very important concept, arises everywhere, signal processing, machine learning, engineering, communication and all fields of engineering.

In fact, there is hardly any field of science engineering, probably business, management, even probably human, economics, wherever you have data, wherever you analyze data, wherever you are fitting data, anywhere, probably even in politics. Well. wherever you have data and you are trying to fit data, this is the problem that always arises and we are going to see several applications. One of the most important profound and fundamental applications, I would say, in entire linear algebra. So, let us stop here. Please go through this once again to understand this thoroughly. Thank you very much.