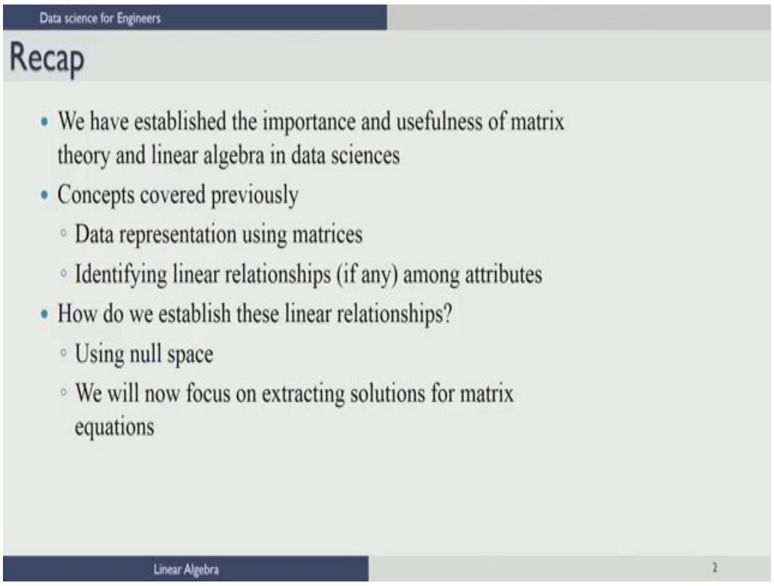
Data Science for Engineers
Prof. Raghunathan Rangaswamy
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 13
Solving Linear Equations

In this lecture, we will discuss solutions to linear equation. As we saw in the previous lecture we had established the importance of matrix theory and linear algebra in data science.

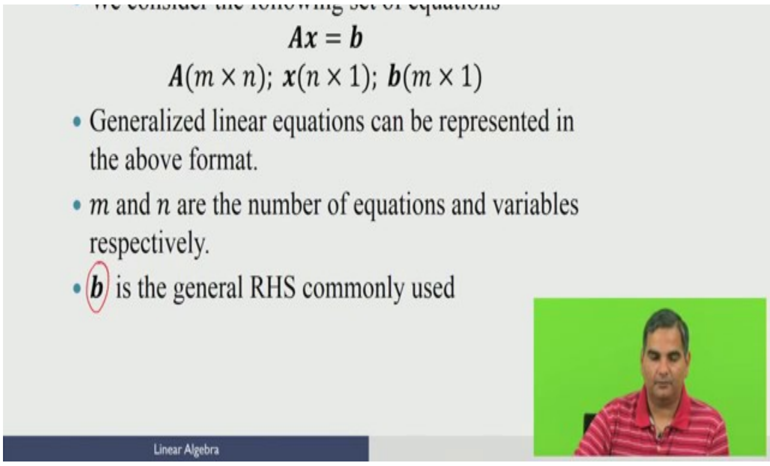
(Refer Slide Time: 00:26)



The concepts that we covered previously are data presentation using matrices and from matrix of data, we saw how to identify linear relationships if any, among the variables or attributes. We saw that we could establish these linear relationships using the concept of null space.

In this lecture we will focus on the next important topic of extracting solutions for matrix equations, which is the most fundamental aspect of data science, where we might have several equations and we might have to find solutions to those equations. So, we will look at solving matrix equations.

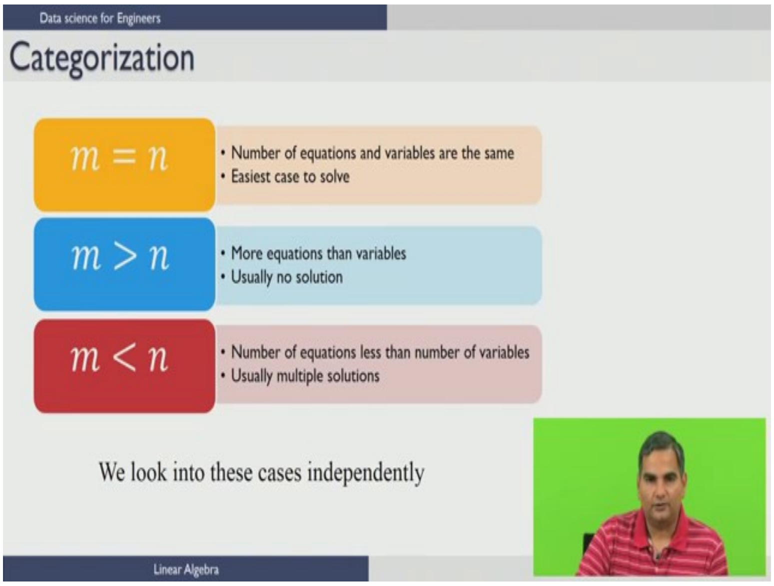
(Refer Slide Time: 01:11)



In general when we have a set of linear equations, when we write it in the matrix form, we write this in the form Ax = b where A is generally a matrix of size m by n, and as we saw in the last class m would represent the number of rows and n would represent the number of columns, and for matrix multiplication to work, x has to be of size n by 1 and b has to be of size m by 1.

Now if you take each row of this equation you will have a left hand side and the right hand side. And the left-hand side will have terms corresponding to multiplying the first row of A with x and the right-hand side will have the term corresponding to b. If you take the first row, it will be the first equation and so on. So, from that viewpoint, m represents the number of equations in the system of equations and n represents the number of variables, and in general b is the constant matrix that is used on the right hand side. So, when we write Ax = b, this represents a set of m equations in n variables.

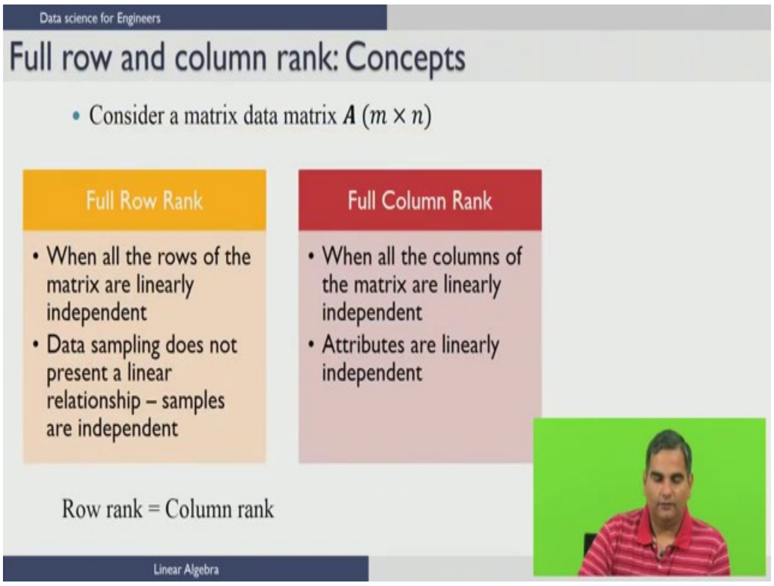
(Refer Slide Time: 02:39)



Now, clearly there are three cases that one needs to address when m = n; that means, the number of equations and variables are the same. So, this turns out to be the easiest case to solve. When m is greater than n; that means, we have more equations than variables. So, we might not have enough variables to satisfy all the equations. In the usual case this will lead to no solution when m is less than n.

The number of equations are less than the number of variables. What this basically means, is that we have lot more variables than necessary to solve the given set of equations. So, in a general case this will usually lead to multiple solutions. So, the first case is the easiest to solve the second case does not have solution usually, and the third case has multiple solutions. What we are going to do, is we are going to look into these cases independently, and then combine all of them using the concept of pseudo inverse.

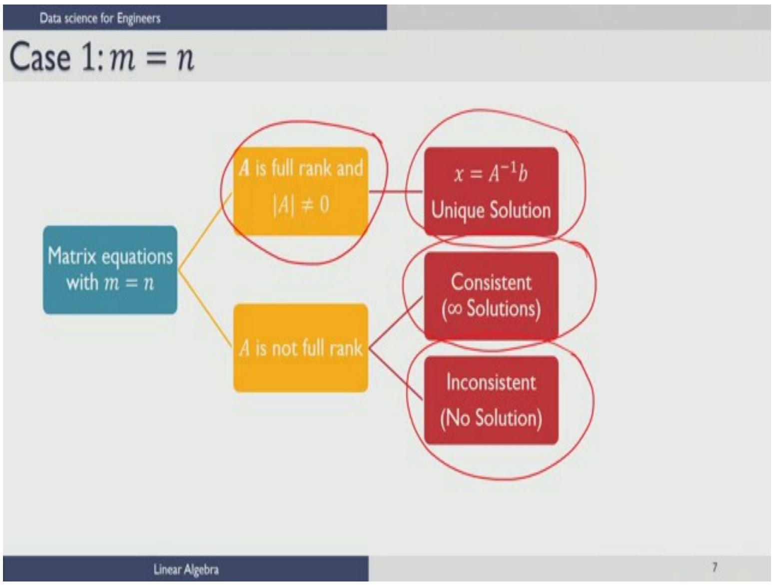
(Refer Slide Time: 03:43)



We had already discussed the concept of rank in the previous lecture, but let us talk about it again, because this is going to be useful, as we talk about solving equations. If you can consider a matrix A m by n. Now if all the rows of the matrix are linearly independent, then remember we said the rows represent data. So, this, this basically means that the data sampling does not present a linear relationship; that is the samples are independent of each other.

Now when all the columns of the matrix are linearly independent that basically means the variables are linearly independent, because columns represent variable. In a general case if I have a matrix m by n, if m is smaller than n, the maximum rank of the matrix can only be m. So, the maximum rank can be the less of the two numbers. So, in cases where I have A matrix m by n, where n is smaller than m, then the maximum rank that is possible is n. In general whatever be the size of the matrix, it has been established that row rank = column rank. So, you cannot have a different row rank and column rank. What it basically means is, whatever be the size of the matrix, if you have a certain number of independent rows, you will have only those many numbers of independent columns and so on. So, this is something important to bear in mind.

(Refer Slide Time: 05:18)

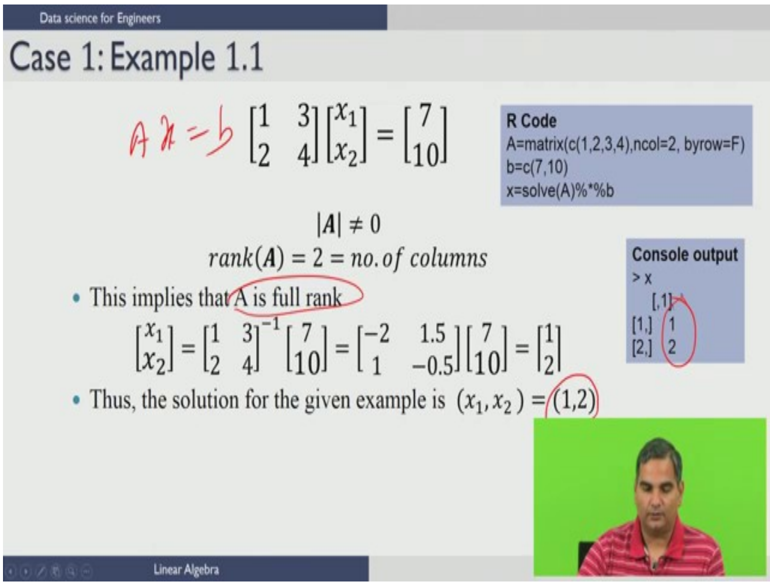


Now, let us look at the case of m = n; that is we have the same number of equations and variables. If A is full rank, what does full rank mean? Now we have the same number of equations and variables m = n. So, the rank of the maximum rank of the matrix can be m or n, because both are the same. Now if the rank of the matrix turns out to be m, then it is what is called the full rank matrix, what this basically means, that all of these equations on the left hand side are independent of each other. In other words you can never get any equation on the left hand side as a linear combinations of other equations on the left hand side. In this case there is a unique solution to the Ax = b problem, and that unique solution is x = A inverse b.

Now from your high school and so on, you would have learned that if the determinant is not 0, A^{-1} is possible to compute. So, one could simply compute x = A^{-1} p as a solution to this problem, the difficulty arises only when A is not fulltime; that means, the rank of the matrix is less than n. In this case what it means, is if I take the left-hand side of the equation Ax = b, and then make some linear combinations of some rows of A, at least one of the rows of A is going to be a linear combination of the other rows of A; that is the reason why the rank of the matrix became less than m. In this case, depending on what the values are on the right hand side, you could have two situations; one situation is what we are going to call as a consistent situation.

I will explain this through an example in later slides when you have a consistent situation, then you will have infinite number of solutions. There could be many solutions for A x = b. And in the case where the system of equations become inconsistent, there will be no solution to this problem. So, to summarize when m equal n this is what is called a square matrix, and if the matrix is full rank determinant A not = 0, then there is a unique solution x = A^{-1} b, and when A is not full rank there are two situations that are possible! One is what we call as a consistent scenario, where we could have infinite solutions, and the other one is what is called the inconsistent scenario where we might have no solution.

(Refer Slide Time: 08:07)

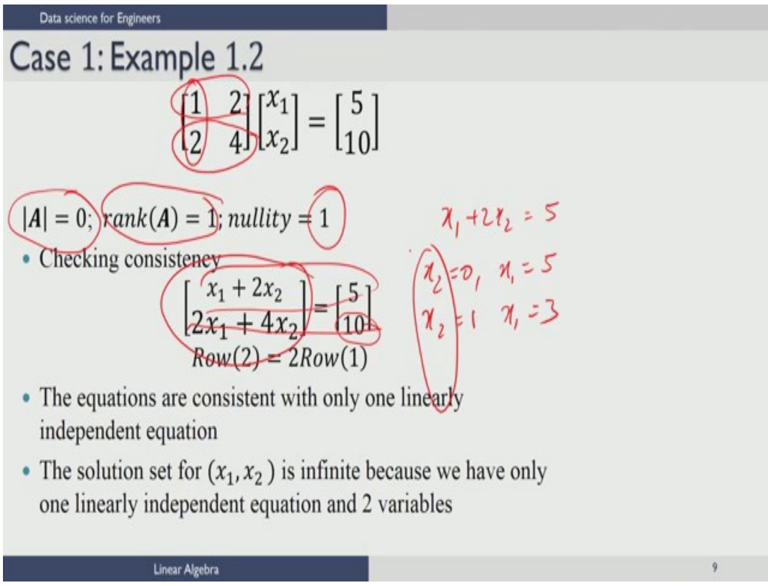


Let us take a simple example where I have on the top of the screen, the matrix in the form A x = b. In this case matrix A is 1 3 2 4 x is x_1 x_2 7 10. So, you can notice that there are two equations in two variables x_1 and x 2. The first equation is basically $x_1 + 3x_2 = 7$ and the second equation is $2x_1 + 4x_2 = 10$. We can see that this is full rank, because whatever multiple of the first column you take, you can never represent the second column and similarly whatever multiple of the first row you take you can never represent the second row, and this can also be seen from the fact that the determinant of A is not 0.

From your high school, you know the determinant is going to be in this case simply 4 times 1, - 2 times 3. So, this is not 0. So, rank of A is 2. So, we have maximum rank that matrix size is 2 by 2, and the rank is 2. So, this implies that the matrix is full rank. Now we said in the previous slide that x_1 x_2 can be written as A inverse b. So, this is A inverse matrix 1 3 2 4, this inverse b. You can do an inverse computation and then calculate the solution as 1 2 and if we put the solution back into the equation, you will see 1 times 2 + 1 times 1 + 3 times 2 is 7 and 2 times 1 + 4 times 2 is 10. So, the solution 1 2 satisfies this equation.

Now if you want to write an r code for this. It is very simple, you write a matrix put the numbers in c and then define the number of columns, you de ne what b is and then simply use the command solve for x to get the solution. And as you notice here the solution is 1 2. So, this is a case of full rank where I get an unique solution. The important thing to note here is, no other solution will be able to satisfy these two equations.

(Refer Slide Time: 10:28)



Now, let us take another example to illustrate what happens when the rows or columns become linearly dependent. So, here is another set of equations and we have to read these equations as $x_1 + 2x_2$ is $5 + 2x_1 + 4x_2 = 10$.

Now if you notice these equations, through the matrix A you will see that, if I multiply the first column by 2 I get the second column. So, the second column is linearly dependent on the first column or the first column is linearly dependent on the second column, whichever way you want to say it. Similarly if you divide for example, the second row by 2 you will get the first row or if you multiply the first row by 2 you get the second row. So, the rows are also linearly dependent and, and as I said before, there is only one independent column and that necessarily means that there will be only one linearly independent row.

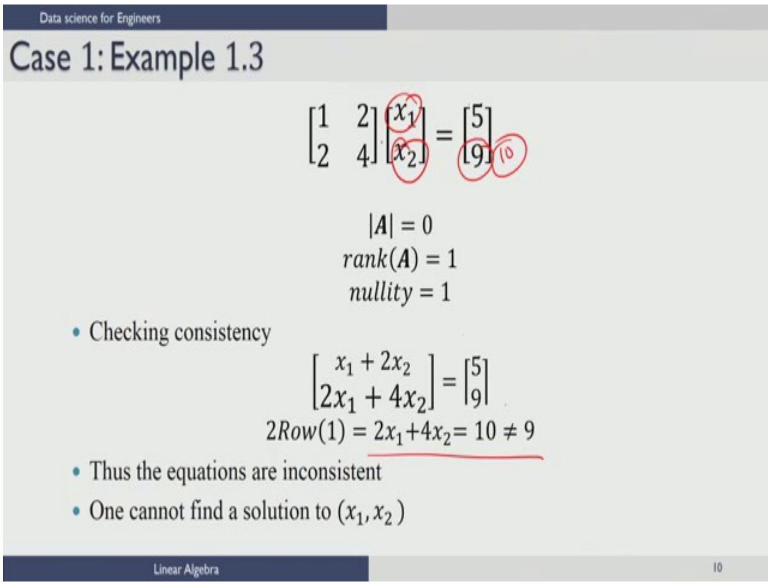
Now another way to check this linear dependence is to calculate the determinant of this matrix, and when we calculate the determinant of this matrix, you will get 1 times 4 - 2 times 2, which is 4 - 4 which is 0. So, determinant also says there is linear dependence here. And from the previous lecture we know that when the rank is 1 and the number of columns are 2 the nullity is 1. So, there is one vector in the null space. Now let us look at the equations when you write these equations, as I said before the first equation $x_1 + 2x_2 = 5$, and the second equation is $2x_1 + 4 = 4x_2 = 10$.

When we talk about the linear dependence of the rows of A, we are only talking about the left hand side, we never talked about the right hand side. Now whenever the left hand side becomes linearly dependent, if the same linear dependence is maintained on the right hand side also then we have the situation of consistent equations. So, if you take a look at this example, we know the left hand side, if I take the first equation and multiply it by 2 I get the second equation on the left hand side. So, $x_1 + 2x_2$ multiplied by 2 gives me $2x_1 + 4x_2$. Now if the same linear dependence is maintained on the right hand side; that is if I take the first number 5 and multiply it by 2 I should get this number.

In this case, we have constructed this example in such a way that we get this number. Now not only is the left hand side linearly dependent, the same linear dependence is also maintained on the right hand side. So, as a whole the equations become consistent, but linearly dependent on each other. So, in this case you notice that if I solve this equation. I do not have to solve this equation, because I multiply this by 2 I get this equation. So, whatever 1 and x_2 will solve this equation will also solve the second equation. So, I can drop one of these two equations. Let us assume I drop this equation out. So, I am just left with $x_1 + 2x_2 = 5$, but now notice that I have one equation in two variables, that basically tells me I have one free variable. So, for example, if I take this equation $x_1 + 2x_2 = 5$.

If I set $x_2 = 0$ I get $x_1 = 5$. If I said $x_2 = 1$ I get $x_1 = 3$ and so on. So, all of these are solutions to these equations. I am just pointing out two. Now you can notice that I can take any value for x_2 and then calculate an x_1, which will satisfy this equation. Since I can take any value for x_2. There are infinite choices for x_2 corresponding to each one of these infinite choices, I will get a value of x_1. So, that pair would be a solution to this set of equations. So, when this A becomes rank less than full rank, if the equations are consistent then they will get infinitely number of many infinitely many solutions.

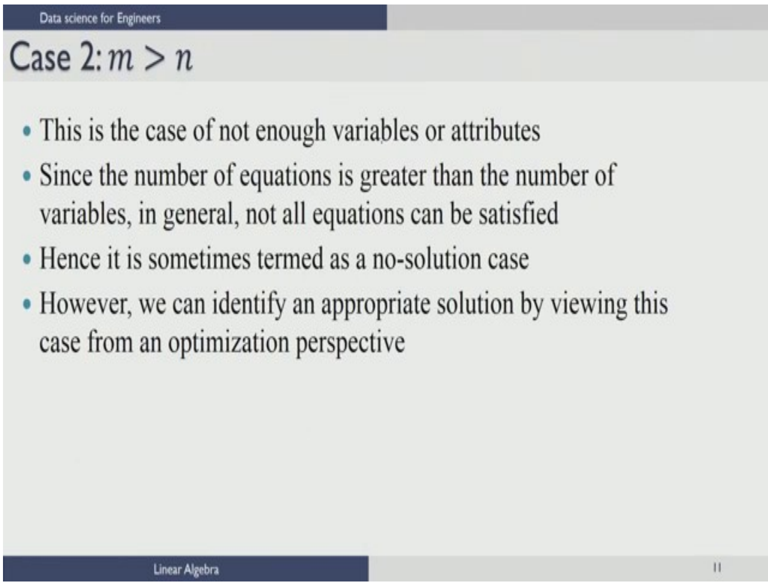
(Refer Slide Time: 14:50)



Let us take another example to explain the idea of inconsistent equations. Let us look at this example here, this is the same example as the previous case except for one change. This number has been changed from 10 to 9. In the previous case we saw that the left hand side, if I multiply the first equation by 2, I get the second equation. And since in the previous example this number was 10, when I multiplied 5 by 2 I get 10. So, both the equations became consistent and I could drop one equation ; however, if this number was anything other than 10 then what it basically means is, that while the left hand side will be linearly dependent where the second equation will be twice the first equation the right hand side will always be inconsistent.

If this number is not there that basically means whatever x_1 and x_2 values that you use for solving the first equation, If you plug that value into the second equation, your right hand side will always be 10, but you want it to be 9; so which is never possible. So, this is a case where you will not have any solution. So, in other words if this number is anything other than 10, you will have no solution to these equations, this will become a inconsistent case. Now again inconsistent case is possible, only when the rank of the matrix is less than full rank. That is there are some rows which are linearly dependent on the left hand side, which you can again verify by the same determinant 1 times 4 - 2 times 2 is 0. And since the second row is inconsistent, these equations are inconsistent and one cannot find a solution to this problem of A x = b.

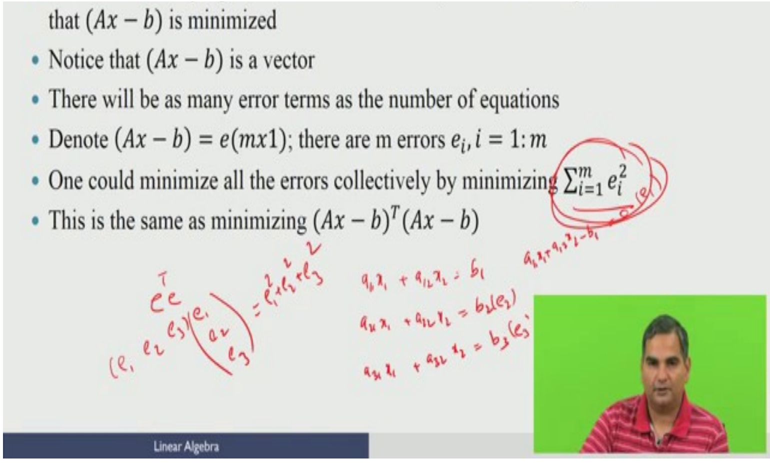
(Refer Slide Time: 16:44)



So, that finishes the case where the number of equations and variables are the same. So, till now we saw the case, where m = n. Now let us take the second case, where m is greater than n. Since m is greater than n this basically means that I have more equations than variables. So, this is the case of not enough variables or attributes to solve all the equations. Since the number of equations is greater than the number of variables in general, we will not be able to satisfy all the equations; hence we termed this as a no solution case.

However we still want to identify some solution which makes sense and which we can generalize for all the three cases. So, what we are going to do is, we are going to identify an appropriate solution by viewing this case from an optimization perspective and I explain what that is presently.

(Refer Slide Time: 17:47)



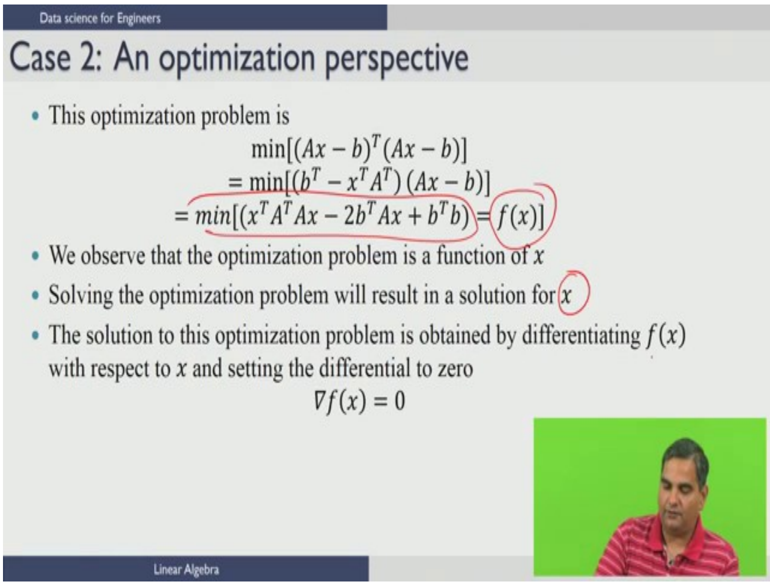
Let us look at a solution to Ax = b when the number of equations are more than the number of variables. As we mentioned before we are going to take an optimization perspective here. When we try to solve Ax = b, we can write that equation as A x - b, and if there is a perfect solution to the set of equations then Ax - b will be = 0. However, since we know that the number of equations are a lot more than the number of variables,there might not be a perfect solution.

So, what we want to do is, we want to identify a solution in such a way that Ax - b is minimized. Why do we want to do this? Notice that Ax - b is a vector and if you take each term in that vector you can think of each of those terms as an error in an equation. So, to give you an example, if I have $a_{11}x_1 + a_{12}x_2 = b_1$ $a_{21}x_1 + a_{22}x_2 = b_2$ $a_{31}x_1 + a_{32}x_2 = b_3$. Now if I had a perfect solution x_1 x_2 which will satisfy all the three equations.

Then when I write this as $a_{11}x_1 + a_{12}x_2$ - b1, this will be = 0; however, when I cannot find a perfect solution then let me call this as an error. In this equation correspondingly I allow an error in this equation, I have error in this equation. So, you notice that there are three errors - as many errors as there are equations. So, how do we collectively minimize all of these errors? One thing to immediately think of both is to minimize $e_1 + e_2 + e_3$, but that would not be a good idea simply, because I could have a_1 as a very large error in the positive direction e_2 as a very large error in the negative direction and e_3 as 0 that will still give me an answer 0. So, that is not a good answer at all, one way to do this is to collectively minimize all of them by minimizing what we call as the sum of squares errors. So, you take this example instead of minimizing $e_1 + e_2 + e_3$. You are going to minimize $e_1^2 + e_2^2 + e_3^2$. In this case notice irrespective of whether e_1 e_2 e_3 are positive or negative as long as they are away from 0. The contribution to the error term will be high. So, it will automatically ensure that you do not go very far away from zero.

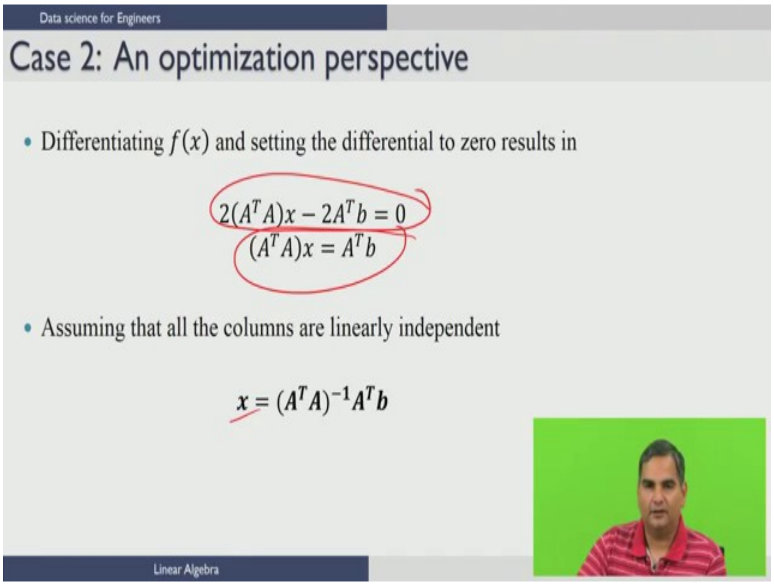
Now, this is the least squares solution you could also minimize instead of e_1^2, you can minimize modulus e_1 + mod e_2 + mod e_3, because mod is always positive; that is also possible, but in general we are going to talk about least square solution where we minimize this sum of squares of errors. Now this is the same as minimizing Ax - b transpose times Ax - b simply, because Ax - b is e. So, $(Ax - b)^T$ is e^T e. So, if I have numbers e_1 e_2 e_3 ^{T} and multiply by e_1 e_2 e_3, this will lead to $e_1^2 + e_2^2 + e_3^2$ which is the same as this right here. So, minimizing this, is the same as minimizing $(Ax – b)^T (Ax – b)$.

(Refer Slide Time: 21:38)



So, the optimization problem is minimize $(Ax – b)^T (Ax – b)$. After some algebraic manipulation we can write this objective as a function of the solution f(x). We observe that this optimization problem becomes a problem, where the objective is a function of x. Solving this optimization problem will result in a solution for x, which is what we are going for. So, the way to get the solution to deck this optimization problem, is to take f(x) differentiate it with respect to x and set it to 0.

(Refer Slide Time: 22:20)



So, differentiating f(x) and setting the differential to 0 results in the following equation, and this can be simplified to this equation, where A^T a times x is A^T b. Now to solve this equation we will assume that all the columns are linearly independent which allows us to take the inverse of this matrix, and then we can come up with a solution x equal $(A^T A)^{-1} A^T$ b. While this solution x might not satisfy all the equations, this solution will ensure that the errors in the equations are collectively minimized. So, this is an optimization view for case 2, where the number of equations are more than the number of variables or m is greater than n.

We will conclude this lecture at this point. What I will do in the next lecture is, take an example to illustrate what happens in case 2 in terms of how you get a solution and whether some of these equations are satisfied or not satisfied and so on, and after that I will move on to $case_3$ and show an optimization perspective for solving those types of equations, where the number of variables become greater than the number of equations, and then in the next lecture I will also show how all of this can be combined into one elegant solution through the concept of pseudo inverse.

Thank you.