Hello! Welcome to Linear Algebra on demand brought to you by Professor Jen-Mei Chang at California State University, Long Beach. In this video lesson today, we'll look at some background information on the approximation to the linear system $Ax = b$. Keep in mind that the goal is to eventually develop a direct method that approximates the solution to the linear system $Ax = b$. In the linear system $Ax = b$, there are typically three different scenarios that can come up. The first case is when the size of the matrix $n$ is equal to $p$. It is to say that the number of equations is exactly the same as the number of unknowns. Then the ideal situation is that the system $Ax = b$ has an exact solution whenever the matrix $A$ is invertible.

If that's the case, then the solution to the system $x$ is precisely obtained by doing a inverse times $b$. And keep in mind that this is just a expression to find the $x$, the solution $x$, but we typically don't solve $x$ this way due to the numerical instability. In general we solve for $x$ using the Gaussian elimination process. Now this is the best case scenario, where we have the number of equations equals to number of unknowns. In most of the situations, and that's often what happens in reality, is that we have number of equations greater than the number of unknowns. The case number two is you have a matrix $A$ size of $n$ by $p$ where $n$ is greater than $p$. This is to say that the number of equations is greater than the number of unknowns. In this situation, we call the system $Ax = b$, an overdetermined system. The reason why it's called the overdetermined system is because in general if we have $p$ unknowns, then we only need $p$ equations to solve or resolve that ambiguity.

But in this case, you actually have more equations than the number of unknowns, so you actually have redundant information, or too much information to resolve ambiguities. That's why it's overdetermined. Now think of this in the situation where we're looking for the regression. If you have a data sets such as the following where the $x_1$ and $x_2$ corners were given, so think of the $x_1$ corners as a time and think of the $x_2$, for example, could be some sort of measurement over time. For example, if something is measured at time equals 1 and gives the value of a 2 and when it's measured at times 2 you get a measurement of 3 and if we're looking for a linear equation that goes through those two points, then those two points determine that line uniquely. Therefore, two equations and two unknowns. What if I give you one more data point? Say that at time equals 3, we got a measurement of 3.5? Now ask yourself the questions.

How could I then find a straight line that goes through all three points simultaneously? The answer shouldn't be so obvious anymore. It doesn't seem like that it would be a single straight line that's going to go through all three points. In such a case, the system is called the overdetermined system because those three points then gives rise to three equations but with two unknowns only. Suppose the line it goes through has the form of $y = Ax + b$. Suppose that this regression line goes through or approximately goes through at these three points which has the coordinate of 1, 2; 2, 3; and 3, 3.5. That means these three points need to approximately lie on this line which means that they should satisfy this equation simultaneously. The first point satisfies the equation we get 2 equals 1a plus $b$. The second point satisfies the equation gives us 3 equals 2a plus $b$. And the third point needs to go through this line which then gives us 3.5 equals 3a plus $b$. 
But if you take these three equations and try to write it as a matrix equation, then this column right here corresponds to the b vector but this matrix here has to be a coefficient of 1, 2, 3 and 1, 1, 1 and to verify that, if you do the row column multiplication, you get 1a plus 1b equals 2 and that's the first equation; 2a plus 1b is equal to 3 and that's the second equation and so on. So essentially, solving the equation of the regression line corresponds to solving the solution to the linear equation Ax equals b. And we know that an exact solution shouldn't happen because if an exact solution actually does exist, that means this line needs to go through all three points simultaneously. So I can tell you right away this equation is not going to have a unique solution.

Or the best we can do is come up with an approximation solution to the system so the result of multiplying a with x gives rise to a vector b that comes as close to the actual vector b as possible. [inaudible] every time we have an overdetermined system, we're looking for an x hat such that a times x hat is approximately b. So in this case, we're looking for an x hat such that the difference between a times x hat and b is as small as possible. But if you actually do [inaudible] subtraction, this gives rise to a vector. How do we actually quantify the vector being small? We can't do it unless we have a numerical value associated with it. Therefore, instead of studying this quantity on its own, we're going to study the [inaudible] or the length of this vector. We wanted the length of this vector to be as small as possible. Hence, what we're going to need to do is develop some notions of length in high dimensional spaces and that will be discussed in a separate video.