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 a well-known result from [inaudible] Algebra called "the Approximation Theorem," which then motivates the study for distance between a point and a subspace. Geometrically if we realize W is a subspace in [inaudible] space of RN, and Y is any point in RN. Then we wanted to find the distance between the point Y to the subspace at W to be the shortest path that we can take Y to W. Intuitively this then corresponds to doing a Orthogonal Projection of the point Y onto the space W, and then measure the length here. If you have studied the Orthogonal Projection Theory before, then you would know that when you project a point Y onto the space W, it gives us a vector somewhere in W.

Imagine this is the origin for the space W, with a 0 vector, then this is the Orthogonal Projection, and then the difference between the Y and its Orthogonal Projection gives the residual. And then the distance between the point Y and the subspace W is precisely defined by the length of this residual. But I have to argue that it seems we define distance to be the shortest path between a point and a space I have to make sure that this path here is indeed the shortest. Therefore, I'm going to talk about the Approximation Theorem next. Now, suppose W is subset of RN, is a subspace of RN, and Y is any point in RN, we also define Y hats [phonetic] to be the projection of Y on to W. Then the Approximation Theorem says Y hats is the closest point in W to Y in the sense that its residual, defined by the difference between Y and Y hat, is the smallest when compared to any other residual produced by taking a difference between Y and any other vector in W. So here vector V is any other vector in W, not equal to Y hats.

Pictorially the vector V is any other vector not equal to Y hat. The theorem says that if you now were to produce the residual between the Y and V, taking the difference between Y and V, which then will give you this vector here, then the length of this vector is always going to be greater than the length of the residual produced by taking the projection of Y onto W. Essentially, we're arguing that this length here is always going to be greater than the length over there. I can prove this really quickly using the Pythagorean Theorem. Since this vector here is the residual between Y and the Orthogonal Projection, this must form a right angle with W. And the other vector done this way would not produce a right angle, therefore the hypotenuse of this right triangle. And we know from the Pythagorean Theorem that the hypotenuse is always going to be greater than any leg of the right triangle, hence the results. So here the Approximation Theorem tells us that the residual between Y and its Orthogonal Projection is indeed the shortest path from Y to W.

We can then confident of using this definition now. Let's look in the example calculation here. In this example we're given a vector Y, and 2 directions. You want [inaudible] and define the subspace W as a span of U1 and U2. We wanted to find the distance between Y2 to subspace W. We have to caution ourselves that the definition of the distance is defined when we have Orthogonal Projection. Therefore, we need to actually check to see if these 2 vectors that currently serve as the spinning set is indeed an Orthogonal basis for the space W. And quick inter-product tells us a 5 minus 4 minus 1 is indeed 0. So you want a U2 [inaudible] Orthogonal basis for W. To figure out the distance between Y into W you have to first figure out its projection.
Well, that amounts to doing $Y$ transposing $U_1$ over $U_1$ transposed $U_1$ times $U_1$ plus $Y$ transposed $U_2$ over $U_2$ transposed $U_2$ times $U_2$. I'll spare you with all the calculation, we get the following, 1/2 times 5 negative 2 1, minus 7/2 times 1 2, negative 1, which then gives us the vector negative 1, negative 8, and 4. Then the distance between $Y$ and $W$ is the norm of the residual, $Y$ minus $Y$ hats, which is norm of the vector negative 1, negative 5 10 minus negative 1, negative 8 and 4, which then is the norm of the vector 0 3 6. The norm of that vector is then square root of 0 squared plus 3 squared plus 6 squared, which is square root of 45. This distance here being a [inaudible] tells us that our original vector $Y$ does now live in the subspace $W$. Next I want to look at the matrix notation for projection, especially when we have an orthonormal set of basis elements for the subspace $W$.

Suppose I have a collection of orthonormal basis elements, $U_1$ through $U_P$ for subspace $W$ of $R^N$ if we let the matrix $U$ to be the matrix with $U_1$ as its column vectors. And this matrix is of size $M$ by $P$. Then if we actually define the matrix $P$ as $U$ times $U$ transposed -- and this matrix $P$ here is called the "Orthogonal Projection Matrix," such that -- again then apply this matrix to my vector $Y$ to obtain the Orthogonal Projection of vector $Y$ onto $W$, I.E., when I do $P$ times $Y$, I will get $UU$ transposed -- I should times $Y$, which then through its definition will give us the calculation of $Y$ transposed $U_1$ times $U_1$ plus $Y$ transposed $U_2$ times $U_2$, which is then precisely the definition of projection of $Y$ onto $W$. In the following proof I will show indeed when I do $P$ times $Y$, I get this linear combination. Well, $P$ times $Y$ is equal to $UU$ transposed $Y$. Here I'm going to realize $U$ as the column vectors.

I'm going to leave $U$ as it is. But I'm going to realize the $U$ transposed as the following, because this is taking each column vector and then transposed make it into a row vector, so I can write it as $U_1$ transposed, which is the row vector this way, $U_2$ transposed all the way to the $U_P$ transposed. The matrix location can be done in either order in the sense that I can perform the multiplication with these 2 first, and then multiply the result with the $U$ on the left. My definition of a row call in multiplication I can write this as $U_1$ transposed $Y$ times $U_2$ transposed $Y$, all the way to the last row of $U_P$ transposed $Y$. But notice that each time I do an inner product or [inaudible] product, because this is a real number I can do the inner product the other way and still get the same results. In the next step I'm actually going to realize my vector -- my matrix $U$ with its columns, and then switching the order on how I do this inner product.

Then using the definition of the matrix multiplication I can realize these as the weights for each columns in the matrix $U$. Then write this as $Y$ transposed $U_1$, which is just the weight times the first column, plus the second weight to the second column, all the way to the last weight to the last column. And now if we compare, this is exactly what was given in the theorem. Since the basis elements are orthonormal, that means when I divide the $Y_1$ transposed $Y_1$ here, it just becomes 1 here, and the same thing for everything else. So this then is equal to the projection of $Y$ onto $W$, which is defined as the $Y$ hats. But furthermore if the matrix $U$ is orthogonal -- now remember, orthogonal matrix means that well the matrix $U$ is actually squared, that means $N$ is actually equal to $P$ here. Well, if $N$ is equal to $P$, that means $U$ transposed $U$ is equal to the $UU$ transposed, which are both equal to the identity matrix. Then the projection $P$ with $Y$, which is defined as a $UU$ transpose, now becomes the identity matrix times $Y$. 
But that's just equal to \( Y \). So then this is saying that if \( U \) is actually orthogonal, which consists of \( N \) orthonormal basis elements, then the projection of the vector \( Y \) onto the subspace \( W \), which is the \( N \) dimensional subspace of \( R^N \), should be equal to itself. Now, of course this definition makes sense, because if the matrix is orthogonal, that means \( W \) is of \( N \) dimensional [phonetic], which is precisely the same dimension as the ambient space \( R^N \). When you're projecting things into its own space, of course you just get the vector back itself. And this concludes the discussion on the Approximation Theorem and the distance between a point to the subspace, as well as a quick discussion on the matrix notation for projection.