GILBERT STRANG: So this lecture is, in a way, a whole linear algebra course created by highlighting five different ways that a matrix gets factored. By factored, I mean a factorization would be-- the first example will be a matrix A. Every matrix A factors into a matrix C and a matrix R, C times R. And I'll describe those. And later ones involve eigenvalues. They involve singular values.

It's linear algebra in a nutshell of these factorizations. So this is coming from the latest and the final edition of my linear algebra book. I'm grateful to say there's a sixth edition this year, 2023.

For the start, let me start right out with some of the key ideas before you see the first factorization. So one key idea is whether a set of vectors is linearly independent or dependent. And that simply means-- dependent simply means that some combination of those vectors gives the zero vector. And of course, I'm not allowing the 0 combination, which takes 0 of every vector, but some other combination like 2 of the first vector plus 3 of the second vector minus 11 of the third vector gives the zero vector. That would mean those vectors were dependent.

And combinations is what you can do to vectors. You take combinations, multiply them by numbers-- x1, x2, x3. They multiply the vectors-- A1, A2, A3. And then you add those together.

So every time I teach linear algebra, this is in the first lecture. How do you multiply or-- and how do you think of a matrix A times a vector x? And I think of it not as a lot of little, tiny computations, dot products, but I think of it vector-wise. I think of A times x as a combination of the columns of the matrix. And we'll come back to that. That's important.

And now that third line is speaking about C times R. Those are matrices. Now we're multiplying a matrix by a matrix. Before, it was a matrix times a vector. So I'll have to show you important ideas because there are several ways to multiply a matrix by a matrix. And they're all important. It's a fantastic subject.

And one example would be every matrix will get factored. This will be factorization number 1. Every matrix A will get factored into a matrix C that just comes from the columns and times a matrix R that comes from the rows, which you will see soon.

Oh, I even wrote it at the top, number 1, A equals CR. But I'm going to leave the details, the all-important details, for the coming slides. I thought here I would tell you factorization number 2 because it's probably the most used factorization in numerical computing. Probably billions of dollars a year are spent in taking-- this is usually a square matrix A. And we're trying to solve n equations and n unknowns, where n could be 10,000.

And the way to do it is a factorization. And it's known as LU. Everybody connects to that-- those words, LU. L is for lower, and U is for upper. And those are used because the two matrices-- L is a lower triangular matrix. Above the diagonal of the matrix is all 0. All the nonzero parts of the matrix L are below, whereas U is the opposite. It's 0 below the diagonal. And its nonzeros are all on the diagonal or above.

So virtually, all matrices factor into a lower triangular times an upper triangular. Sometimes, you need a matrix P, a permutation. You'll see permutations show up here once in a while. That's just a reordering of the rows. Sometimes, if the rows start with-- if the first row starts with a 0, you're not happy with that, and you have to have to switch rows.

And why is this L times U so great? Well, because you can solve Ax equals B, let's say, solve a set of equations, by first solving the problem with the matrix U and then-- let me-- it's-- so we're going to solve it in two steps, as the slide says.

The first step is to solve-- find the solution to LC equal B. So L is lower triangular. B is the right-hand side we want. And C is the answer to that. You could say it's half the problem. It's the lower triangular half of the problem. So that gives us an answer C. That's not the x that we're finally aiming for. But it's halfway.

And then put that C on the right side of the second equation, Ux equals C. Now we're solving for x with the upper triangular part. Do you see the two triangular matrices? And triangular systems are fast to solve. And that's why it's just fantastic to split the matrix into two fast cases.

And then when you put them together, A times x is-- A is LU. So we have LU times x. And the Ux is C. So we have LC. And that was B. So we got Ax equal b. Every course is going to teach that. I don't have to-- I won't say more about that. But of course, it's fundamental.

So I'm going to go back to this C times R, which is not so classical. But I think it's a great way to start linear algebra. An example's always the best. So you see a matrix A. You see it has three columns-- 1, 2, 1, 1, 3, 4, 3, 7, 6-- three columns. And it has three rows. So it happens to be a three-by-three square matrix. And it's small enough that we can find the C and the R.

So listen up because this is what-- this is the key point. So what's the matrix C? C is for columns. So you see that I took the first column and the second column of A and put them into C-- 1, 2, 1 and 1, 3, 4. I did not take the third column in C. And why not? Because it's dependent. That column-- 3, 7, 6-- is a combination of the others.

So what do I mean by a combination? I mean that 2 of the first column would be 2, 4, 2. And then if I add on 1 of the second column-- 1, 3, 4-- I think that 2, 4, 2 and 1, 3, 4 add up to 3, 7, 6. So in a way, 3, 7, 6 is already there in the first two. And that's reflected in the R, in the R matrix, the one that's more horizontal that has row, emphasizes the rows.

So do you see-- here we're seeing how matrices multiply. So you see the matrix C with two columns. And I'm multiplying by the matrix R that has two rows. And let's look at the third column of R.

What do we see in the third column of R? We see the 2 and the 1. That's exactly what we wanted to get the third column of A. So I'll just say again, columns 1 and 2 of A were independent. They go in different directions. So they go straight into the matrix C. Independent columns go in C.

But R is aimed to catch up with the columns like the third column of A that are combinations. And again, you see why it's 2 and 1, because 2 of the first column plus 1 of the second column-- that combination of independent columns gives the dependent column.

And if I jump to the bottom line, that matrix R, that row matrix, starts with the identity. Does everybody know about the identity matrix? It's the equivalent of 1 for matrices. If you multiply by the identity, you get whatever back again. So the identity is that 1, 0, 0, 1 part. And then the F part is the part that reflects the dependent guy, the third column-- 2, 1.

So my point is going to be that every matrix-- I can pick out the independent columns, put them in C, and then R will tell me what combinations of those independent columns gives all the columns.

So here is the mathematical truth of this thing. I'm going to use the letter little r for the count of how many independent columns. So r was 2 before. The n minus r are the other columns, the dependent ones.

The factorization is that every matrix A factors into C times R, where C contains the independent columns. And R has these two or three pieces. So R has the identity. Do you see R? R is the IF with a P.

So three matrices are involved with R-- the identity matrix because when C multiplies the identity matrix, it just gives those independent columns in C. The identity doesn't do anything, the I. C times F-- that's the combinations that give the dependent columns. And this matrix P I'm-- tempted to say this damn matrix P, that permutation-- just if the columns came in a crazy order or a different order instead of-- it's super nice when the independent ones all come first, and then the dependent ones all come last. That's what we saw in the example.

But there's an example at the bottom where they-- can you pick out the independent columns in the matrix on the 1, 2, 3, 4, 1, 2, 4, 5 matrix? So the first column is independent. It's fine. But the second column is just 2 times the first. So that second column is dependent. And then we get an independent column-- 3, 4. That's a new direction. And then we get 4, 5 which is dependent, again.

So this had two independent columns-- the 1, 1 and the 3, 4. And you see them sitting in C. And then you have the R matrix that tells you how much of each independent column do you need to get every column of A. If you want the identity matrix to come first, that's what a permutation. Does it switches the order of the columns. And so it puts that identity matrix, the 1, 0, 0, 1 matrix, first.

Every matrix factors into C times R. And C is the independent columns. And R has this special form. It's famous as the row-reduced echelon form. Whatever.

So onwards. Now here are the facts. I'm using little r for the count of independent columns. It turns out the matrix R has that many rows. And then A is equal to C times R, C times R. Matrix multiplication is central here.

And the C has-- oh, now I've introduced a new word, the column space. So we have a matrix. What's the column space? Let me go back to a matrix and talk about the column spaces.

How about the column space of that matrix A that we now understand pretty well? The column space is all combinations of the columns. So I would say the column space of that matrix is all combinations of the first two columns because the third is already a combination. So I don't have to include it in the basis-- the good word is the basis-- for the column space. If I'm wanting all the independent columns-- if I'm wanting all the columns, I really just need the independent ones because the others are combinations.

So the column space C is a plane here. It's the combination of just two vectors. And the row space R is also a plane. It's the combination of two rows, combinations of two rows. So we have a column space. Space is a big word. And it means take all the combinations of your basis vectors.

And so let me go forward to a great first theorem of linear algebra. So this gave me pleasure. I have to admit that. It's at that last line. The column space of a matrix-- remember, our matrix had two independent columns-- and the row space-- so that matrix has two independent rows. Mathematics is telling me that from A equals C times R, all rows of A are combinations of rows of R.

You'll have to think about these ideas and the first-- and work some problems finding C and R and seeing this wonderful fact. If you had a matrix that was 50 rows and 100 columns, then it would not be clear for-- without a big computer how many rows are independent and how many columns are independent. But the truth will be that those two numbers are the same. If it has 11 independent rows, it's got 11 independent columns.

For me, that's just a wonderful fact. So that's the first great fact. And now here's the big picture of column space. So we are really moving along here. We are moving.

So these boxes, these four boxes, are filled with vectors. So we call them vector spaces. And I'll describe what-- that word, space, is important. So what is the row space? That's all the rows and all combinations of the rows. That makes the row space.

You get a space by starting with some vectors, taking all their combinations. It fills out the plane. It fills out three-- some three-dimensional space. It fills out some R-dimensional space.

Let me just ask you, just-- let's just take it. Suppose I have two vectors in three dimensions, two independent vectors in three dimensions, like as in my example, two independent vectors in three dimensions. So I'm in 3D. I'm in three-dimensional space. And I take all combinations of those two independent vectors. And what do I get? I get a plane that goes through 0, 0 because one possible combination is 0, 0. And it's a nice, flat plane in 3D.

You can visualize that, that-- a base-- it's-- the dimension of the plane is 2 because it took two vectors to give you everything you needed to know about the plane, take all combinations of those two vectors. So we had a two-dimensional plane inside three-dimensional space.

What about the columns? Same idea-- you start with the columns. You take all their combinations. So that's the column space. Again, the space means take all combinations.

So it's a whole-- I've drawn it as a square or rectangle or a box. It's filled in by taking all combinations. It gives me a plane. It gives me a three-dimensional subspace.

Well, if I have a bigger matrix, like a 10-by-10 matrix, then the rows-- then I might have six independent columns. The column space of this matrix would be all combinations of those six independent columns. They would each-- all those were sitting inside 10-dimensional space. Just visualize 10-dimensional space, or don't visualize it. I can't. But anyway, if you can, do it.

And in that space, the 10-dimensional space, is some thin, very thin, part of it, which has only six independent vectors-- so a six-dimensional column space inside the big 10-dimensional space.

Our first great theorem was that the number of independent rows equals the number of independent columns for every matrix. That's just wonderful. So that means that the size of the row space and the size of the column space are the same.

And then the other two guys, the null spaces, are the solutions to Ax equals 0. So Ax equals 0 says some combination of columns or of rows is giving 0. And so that-- this is accounting for the rest. So we had six-dimensional row space inside 10-dimensional space. Then the null space, which I'll come back to-- but the point is it will be four-dimensional so that 6 and 4 make 10.

The whole 10-dimensional space is being split into the independent rows and the null space on one side and the independent columns and the null space of-- oh, you see this symbol, A transpose. A transpose just means reverse the rows and-- with the columns. If I have a matrix A with 10 rows and eight columns, then A transpose will be the other way around, eight rows and 10 columns.

Well, this is a whole linear algebra, of course. But we're moving. I have to tell you a little bit how you would find which columns are independent and which are not. It's an algorithm. How would a computer find the independent rows? We have to be able to answer that question because for big matrices, we need the computer.

You remember, we're-- it's interesting that the computer would look first for R, the row matrix, where I introduced, first of all, C, the columns, the independent columns. Well, it finds them both at the same time, this algorithm. It just progresses column by column across the matrix. Column by column, we go across the matrix. And we ask, is this new column-- say, column K plus 1-- is it a combination of those previous K columns or is it independent? That's the key question.

If it's a combination of the other ones, then it'll go in with the F part. If it's independent of the first ones, it'll be a new independent one. And it'll go-- it'll-- the I, the identity matrix, will grow by 1.

So I'm not going to push you to follow the steps of this elimination. Well, actually, they-- elimination was invented in China 3,000 or 4,000 years ago. So it's fundamental idea. It's just subtract a multiple of one row from another row to get a 0 in a place where you want a 0.

We can understand what's going on when we have-- when our matrix has enough well placed zeros. So elimination is just subtracting rows from other rows to produce zeros in the right position so we can see what's going on.

Here, we're actually going to use elimination and see and-- the idea of null space and the idea of equations. This we can do. This we can do.

So over at the top left, you see three equations. And you see four unknowns-- x1, x2, x3, x4. So we have a three-by-four matrix. And it's a full matrix-- hasn't got any zeros. And the idea of elimination is to get some zeros into that matrix. In fact, we can see where those arrows are showing where we want to go.

Starting with the matrix A, three-by-four matrix with these numbers-- 1, 3, 4, 2, 7, 9, 11, 37, 48-- are those columns independent or not? Elimination will tell us. And we want to know which rows are independent.

So I'm just going to look at that-- those numbers at the top for a little while. Well, I'll go halfway down just to share the nice math notation. So I wrote the equations at the top. But really, I wasted a lot of time writing in all those plus signs and equal signs, and so on.

Really, it's the numbers that count. So you see the numbers 1, 2, 11, 17 in the first row. So they go in the first row of the matrix. I just skip all the plus signs and x1, x2s. I know they're there. It's the numbers that matter. And then we hour our-- we get, by elimination, to this much simpler form, R, with the matrix R. Now, let's see. Maybe we could do some of these in our head.

I think that if I add equation 1 to equation 2, I get equation 3. Do you agree with me? I'm at the top left of the slide. If I add x1 to 3x1, I get 4x1. And if I add 11x3s to 377x3s, I get 48x3s. Of course, I cooked it up this way.

Instead of adding, if I subtract the first equation from the third and I subtract the second equation from the third, then I have nothing left. That equation becomes 0 equals 0. That tells me that really, this matrix, this big matrix A, has only two independent rows that I'm-- they're sitting there in R.

Do you see the-- you see R, the last matrix, the 1, 0, 3, 5? That comes from the first equation after elimination. And the 0, 1, 4, 6-- those are the coefficients of x1, x2, x3, x4 in the second equation. And I didn't even bother to write the 0 equals 0 equation because I don't need it or want it in R. R gets rid of all the 0 equals 0 stuff.

So where are we? We started with a matrix A. Well, we started with four equations. And the numbers in those equations went into A. And I did elimination. I subtracted 3 times row 1 from row 2. And that knocked out the 3x1, and so forth.

So I ended up with the equations on the right, which went into R, the 1, 0, 3, 5 and the 0, 1, 4, 6. You see, as predicted, we start with the identity matrix in R, the 1, 0, 0, 1. And then we have the other stuff, the dependent coming from the dependent stuff.

And we started with the first two columns in C-- 1, 3, 4, 2, 7, 9. Those are our original first two columns. And we didn't have the third column or the fourth column because those are combinations of the first two.

So this is a matrix of rank 2. It has two independent columns, the first two. And it's got two independent rows, which are-- I could choose the first two rows of A or, better, the first two rows of R.

The whole point of this was to be able to solve Rx equals 0. And the-- when I've got-- or Ax equals 0 is the same as Rx equals 0 now. And R is this simple matrix.

So do you see that that vector, minus 3, down where it's talking about the null space-- null means 0. So these two vectors there-- they solve Rx equals 0. And therefore, they solve Ax equals 0. That's what you're looking for in solutions of equations.

If you took that minus 3, minus 4, 1, 0 vector-- those are x1 and x2 and x3 and x4-- they easily solve our simplified equations at the top. So altogether, we have done what professors assign students to do for centuries of linear algebra. We have solved the equations by elimination.

We had some messy equations. We took combinations of those to give some nice equations. Then we solved those nice equations. And we could express the result as these two vectors at the bottom, or-- but we could express the whole elimination as C times R, the two independent columns of A times the two rows of R.

This is my final slide on this first factorization. Then I'm not going to give such detail on the later factorizations because the textbook and your professor know those so well. I just want to say here that matrices-- really, using matrices in matrix algebra is wonderful. And one of the things you can do is to break them into smaller matrices.

So you see that W, H, J, K? Well? That's our big matrix with four submatrices in it. And W is the part of the matrix that's coming from independent columns. So that arrow in the top line says that W moves to the identity.

That was the theme of all those crazy computations we did-- was to make the top left corner into the identity. Anc then it makes the rest of the matrix into what you see, that-- but just to say that if you think of elimination by blocks instead of just single numbers, then you see the big picture, the big picture. But the real picture was done on that slide at the top. That's where we did the elimination. And then we got the nice pieces that we're seeing here.

So that was all factorization 1, A equals CR, because you need some help with that. That's brought up so many of the key ideas and tools of linear algebra-- go into A equals CR. And it's just wonderful.

So now, oh, I'm seeing the word "orthogonal." So this next factorization is-- takes a matrix A. The columns of A are probably not perpendicular to each other. All the examples we had-- the columns of A-- if I looked at the angle between them, it wasn't 90 degrees.

There's a simple test for perpendicular that-- it's called the dot product. And matrices usually-- vectors don't pass that test. They're not perpendicular. But it's wonderful when they do. Perpendicular vectors are super easy to compute with. "Orthogonal" is just another word for "perpendicular." And it's a shorter word and a more used word.

So what's the idea of this factorization? Again, we take any matrix A. And we're going to get it into two factors. But they're not going to be L and U anymore. They're not going to be C and R. They're going to be Q times R. And Q is this new idea, this new idea of perpendicular vectors.

I just want to say that perpendicular vectors are so easy to compute with. If independent vectors are good, perpendicular vectors are super independent because they're going in 90-degree directions.

And look to see the nice equation that you get if you have these perpendicular vectors. So I'll use that letter q, little q. So do you see where I'm multiplying a matrix, Q transpose, times Q in the middle, that matrix multiplication? So Q, the columns of Q, are our orthogonal vectors. And I adjust their length to be 1. That's no problem. So they're perpendicular. And their lengths are 1.

And Q transpose-- well-- so transpose is an operation that we use a lot. It just changes the rows-- the columns into rows. So the columns were q1 to qn in the second factor. And then the first factor-- we change Q1 to Q1 transpose, which is a row.

So we have now n columns perpendicular vectors in Q. And we have n rows that are perpendicular vectors in Q transpose. That T is, for short, for transpose. And the whole point is that when I multiply those two matrices, Q transpose times Q, that's when the perpendicular stuff pays off because Q1 transpose times Q2 is 0. Perpendicular vectors-- they're-- when you multiply them, a row times a column-- if they're perpendicular, you get 0.

So we actually get the identity matrix from Q transpose times Q. So that means Q is a-- well, Q is one-- is for me-- well, I'll say the queens of linear algebra. I'm not making them the kings. But the queens of linear algebra are these matrices whose columns are length 1 and perpendicular to each other.

But this is an important idea-- that if I start with a bunch of independent vectors, they're going in different directions. No combination is 0. Then I can turn-- bend-- turn those vectors by a matrix R so that they are not just going in different directions, they're going in 90-degree perpendicular directions.

It's just terrific to have vectors that are perpendicular. So that's actually factorization 3-- is doing it to the columns of A. And then factorizations 4 and 5 are doing it to-- you'll see. Factorizations 4 and 5 are the great achievements of linear algebra.

So I promised you five factorizations. And we've got three, two to go. And they're actually together here. They'll take time in the course. And so one is about eigenvalues. Every professor wants to explain eigenvalues. Those are nice eigenvalues and eigenvectors.

And the other one is about what are called singular values and singular vectors, not so famous except for the fact that they are the best of all. So I'll try to compare them so you'll see how they're related because eigenvalues are simpler to see. But singular values are important, too.

So let's suppose we have-- it's nice if the matrix is symmetric, S-- so S for a symmetric matrix. That means that if I flip it across the diagonal, it's still the same. So the number that's in the bottom left is also in the top right.

Now, here's the key idea, eigenvalues and eigenvectors of that matrix S. So the big equation is S times X equals the number lambda times X, SX equal lambda X. What does that mean? That means that X is a vector. S is a known matrix, known symmetric matrix. And I'm looking for vectors X which don't change direction when you multiply them by S. They don't change direction.

Previously, we might have been looking for vectors that go to 0, XS equals 0. Not now. The key equation you'll see in that top line is S times X equals a number lambda times X. So it's still in the direction of X. Its length just got changed by that factor, lambda. And that lambda is called the eigenvalue. X is the eigenvector. "Eigen" is a German prefix that is-- everybody uses here.

So XS equal lambda X-- you will study that type of question. It's, how do you find the X's and the lambdas? In a way, the problem is nonlinear because lambda is multiplying X. So it's not like XS equal B, a right-hand side. SX equal lambda X takes new ideas. And you'll see them.

And the beautiful thing is that when S is a symmetric matrix, then these eigenvectors, these X's-- it turns out that they're perpendicular to each other. And so if I stay on the left side of the screen, which is about S, you'll see that SX-- if I use matrix notation, S times-- I can list all the eigenvectors. There are n of them, luckily-- so same size as S. If I multiply S times those eigenvectors, I get those eigenvectors back again times that diagonal matrix of lambdas, which are just numbers.

So SX equals X lambda-- that's the factorization 4. Well, I can put the X on the far right and write it differently as S equals X lambda X inverse. A symmetric matrix, X, is its eigenvector matrix, X, times its diagonal eigenvalue matrix, lambda times the inverse of X. So it's a factorization. It's got three matrices in it now.

You're going to learn about those guys. This is the second half of the course. The first half of the course was elimination, spaces of vectors, Ax equals 0, linear equations. Now we are deeper with eigenvectors and with singular vectors and singular values.

Now, what do those mean? The point is that every matrix-- doesn't have to be square, doesn't have to be symmetric. Every matrix A has a full set of singular values. And what are-- and vectors. Singular vectors are really more important than the value-- than the numbers. The values are those numbers, sigma or lambda, that scale things. It's the vectors that are great. We'll always take the vectors to be length 1. And then the scaling goes into the sigma or lambda.

What's the deal with singular values? I'm now telling you about number 5, factorization 5, the end of a linear algebra course, frequently, or-- and the past courses didn't get as far as singular values. But you got to push to get there because it's so important. It's become more important-- oh, it's hard to say this-- more important than eigenvalues.

Well, it's closely-- so closely related that you-- but you just have to-- singular vectors apply even when the matrix is rectangular. And they apply to every matrix. So they are exceptional.

And what is the main point of singular vectors? Listen up because this is it. Every matrix A has, you can find, a bunch of perpendicular vectors, inputs-- v1, v2, up to vr, the rank, the number of independent vectors. So orthogonal vectors going in, v's, multiply by A. Orthogonal vectors come out.

For eigenvectors, they come out in the same direction. For other matrices, that's too much. You can't expect them the same direction. But the amazing, miraculous thing is that you can-- there are orthogonal vectors going in. If you pick the right orthogonal perpendicular vectors going in, then the-- and multiply by A, then perpendicular vectors come out.

So these are singular vectors that a matrix takes one orthogonal bunch of, you could say-- in orthogonal-- one set of orthogonal vectors in the row space-- multiplies, produces orthogonal vectors U. So that also is a matrix factorization.

And the famous one is the one halfway down the page on the far right, A equal the U matrix times this diagonal sigma matrix times the V transpose, the outputs. Sorry, the v's are the inputs because being on the right, they're going to hit a new vector-- the first.

So if I multiply by x, I hit them with V, or V transpose. That rotates the space. Then I hit them with sigma, this diagonal matrix that just stretches the thing. And then I rotate again.

Isn't that amazing? Every matrix is a product of a rotation, a diagonal stretching matrix, and another rotation. Well, you know what it means to rotate a plane, just rotate a plane through an angle. If you're an airline pilot, you know about rotations in 3D. What are those? Yaw is one of them.

There are three rotations that an airplane can do in the same plane, either horizontal plane or vertical plane, one angle if we have just a-- in a plane, the rotation just takes one angle. In three-dimensional space, it takes three angles. In four-dimensional space, well, that's for the linear algebra course.

So what are we saying? What does this singular value stuff tell me? It says, again, that for every matrix, I can find perpendicular vectors for inputs. And when I multiply the matrix, the outputs are perpendicular, too. That's what it says. That's what it says, perpendicular inputs that give perpendicular outputs.

They're very special. Normally, that won't happen. But there is a set of perpendicular inputs where the outputs are perpendicular. And I'm going to show you that, or try to, in 2D, in 2D

So I'm looking-- I'm thinking I have a two-by-two matrix. And my vectors, V and W, are the usual 1, 0 and 0, 1. So if I multiply them by this matrix-- so they're perpendicular. V is perpendicular to W. No problem.

If I multiply by my matrix, I get A times V and A times W. Well, the odds are a million to one that they're not perpendicular. I can't expect so much luck that the-- maybe if the matrix was a diagonal matrix or some simple guy, then AV and AW would be perpendicular. But generally, they won't be.

So V and W are not the pair that I'm-- the input pair that I'm looking for. Let me make one more try and fail again. I'm going to try to put in W and minus V. So I'm looking at the middle picture now. I'm looking at the middle picture. I'm multiplying W by A.

So that still gives me the AW. And I'm multiplying minus V by A. So it comes out 180 degrees around from the original AV in the first picture is A minus V in the second picture. You need to think about it. This is a little bit of a proof.

So what's the point? Well, this new pair, W and minus V, didn't work, either. The angle was not 90 degrees after I multiplied by A. And that's what I'm shooting for.

But here's the little idea, small idea. The angle in the first case was smaller than 90 degrees when I multiply by A. The AV and the AW had some small angle, theta, between them. But now if I multiply W and minus V by A, I get a big angle, bigger than 90 degrees.

And mathematics steps into the problem here, looking for an idea. And mathematics says that somewhere between the VW that I started with and the minus-- and the W minus V, the second pair, somewhere, there's a little v and a little w that make it right because it was-- AV was too close to AW in the first picture. They were too far apart in the second picture. But as I-- if I go smoothly around, I'll hit a point where they're exactly right. And that's what the third picture is telling me.

So there's a proof that every two-by-two matrix-- this is a two-dimensional proof-- that every two-by-two matrix has a pair little v, little w, which are perpendicular, and then also A times v and A times w are perpendicular. So that's factorization 5. And that's my-- that's-- if you cover that in linear algebra course, you're doing well.

Despite the fact that that slide is number 12 out of 12, there is another slide that doesn't have a number. If you're up for this, this is about deep learning. So if you want a job eventually-- probably, most people would like to have a job-- then a lot of jobs, we know, are coming from deep learning.

So I would tell you a little bit of-- this is the last chapter in the linear algebra book. It's probably not part of the linear algebra course, officially. But why not learn about deep learning, artificial-- AI? What is deep learning?

Well, deep learning starts with some training data, some inputs with known outputs. So I'll call the inputs u1, u2, u3. I maybe have 1,000 known inputs and 1,000 known outputs.

I've got a lot of information from the training data. And then I want to be able to predict for a new input, u, not one of these known one, not part of the training data, but a new guy, what is the output. So I'm going to assume that the relationship, which I'm never going to know-- the relationship is something reasonable, that if I have a-- if my new input is nearer to one of the u's, then the output will be nearer to one of the w's. But it will be affected by all the other u's and w's.

So it's called interpolation or something. I'm looking for a function that fits the data. And then I can use that function for new data, test data, where I don't know the output. That's the thing.

And now I want to tell you the key idea of deep learning. The question is, what kind of a function are we going to choose that we're going to-- it'll have some freedom so that we can make it fit the known training data. We can create-- there are a lot of-- there are a zillion ways I could create a function that was correct on the known data that gave me the right answers, w, from the inputs, u.

What I want is to be good on data that I haven't seen. And that was the important problem, important scientific problem, that has been largely solved and understood much better. But there's more to understand still by deep learning.

So I'm just going to go quickly here. The question is, what kind of function should we have? What kind of function-- should I try one big matrix? Well, matrices are linear. And I don't know any reason why this unknown function should be linear.

So matrices are going to come in there. But also, some nonlinear stuff has to come in. And it turns out-- do you remember the chain rule from calculus? It's a cool rule out of calculus that if a chain of functions-- you take a first-- you take an input, give it input in a function F1. You take that output, put it into F2. Take that output, put it into F3. There's a chain of-- chain of simple functions is a good idea. You quickly can build up quite a array of interesting functions out of very simple ones.

And now comes the very last slide to tell you the particular functions that deep learning uses, that particular functions that deep learning-- so people tried linear, the function Ak times the input plus some vector Bk, like a straight line as Ax plus b. So the first attempts were-- had linear stuff.

Well, how can I teach linear algebra and have to confess here that linear is a limitation, a big limitation? And it was too much. And a good linear interpolation doesn't-- there isn't one. You've got to make something nonlinear.

And here's the crazy thing, the fact that it surprised everybody-- that this particular nonlinear function that everybody calls ReLU, R-E-L-U-- that's a particular function that's 0 for when the input is negative. And it gives back the input y when y is positive.

So the graph of the function is two straight lines, horizontal line below 0, 45-degree line above 0. That's ReLU. So that's a nonlinear function just of a number. The input is a number.

Tell me, so what is ReLU of minus 7? If I input minus 7, the ReLU of minus 7 is 0. What is ReLU of plus 7? 7. So I get 0 or 7 when I input minus 7 or 7.

That's the magic function. Anyway, the idea is to use that nonlinear function in every one of these simple F's. And then you put all these F's in a chain. And then you find these matrices, A and B, that are the substantial part of the function. And it's a giant calculation, which we are not going to do here. But it's a success.

This idea has created functions. It may not be the only way to do it. I would like to say-- I would like to hope and think that there could be-- from this one, we could learn the-- learn what's important and find other ways of creating a-- functions. But this combination of a chain of very simple functions-- each of those simple functions involves linear-- a matrix A and a vector b. Then this ReLU thing that's nonlinear thing that knocks out the negative part-- and then go on to the next function, on to the next function, and choose the good A and b to fit your data. And then you've got a way to predict the output for new inputs. And that's what so many problems are about.

If you know some training inputs and outputs, how do you predict the output from a new and different input? Well, that's what deep learning does. And I hope this last couple of slides, which are-- which represent the final chapter of the sixth edition of *Introduction to Linear Algebra--* and I could-- a lot of people email me to-- about the books or about linear algebra. And that's-- makes my life interesting, too. Thank you.