

GILBERT

OK, so I was speaking about eigenvalues and eigenvectors for a square matrix. And then I said for data for many other applications, the matrices are not square. We need something that replaces eigenvalues and eigenvectors. And what they are-- and it's perfect-- is singular values and singular vectors.

STRANG:

So may I explain singular values and singular vectors? This slide shows a lot of them. The point is that there will be-- now I don't say eigenvectors-- two-- different left singular vectors. They will go into this matrix u . Right singular vectors will go into v . It was the other case that was so special. When the matrix was symmetric, then the left equals left eigenvector. They're the same as the right one. That's sort of sensible.

But a general matrix and certainly a rectangular matrix-- well, we don't call them eigenvectors, because that would be confusing-- we call them singular vectors. And then, inbetween are not eigenvalues, but singular values. Oh, right. Oh, hiding over here is a key. A times the v 's gives σ times the u 's.

So that's the replacement for $Ax = \lambda x$, which had x on both sides. OK, now we've got two. But the beauty is now we've got two of those to work with. We can make all the u 's orthogonal to each other-- all the v 's orthogonal to each other. We can do what only symmetric matrices could do for eigenvectors.

We can do it now for all matrices, not even squares, just this is where life is, OK. And these numbers instead of the λ 's are called singular values. And we use the letter σ for those. And here is a picture of the geometry in 2 by 2 if we had a 2 by 2 matrix.

So you remember, factorization breaks up a matrix into separate small parts, each doing its own thing. So if I multiply by vector x , the first thing that's going to hit it is v^T . v^T is an orthogonal matrix. Remember, I said we can make these singular vectors perpendicular. That's what an orthogonal matrix-- so it's just like a rotation that you see.

So the v^T is just turns the vector to get here to get to the second one. Then I'm multiplying by the λ 's. But they're not λ 's now. They're σ . The matrix, so that's capital Σ . So there is σ_1 and σ_2 . What they do is stretch the circle. It's a diagonal matrix. So it doesn't turn things. But it stretches the circle to an ellipse because it gets the two different singular values in-- σ_1 and σ_2 .

And then the last guy, the u is going to hit last. It takes the ellipse and turns out again. It's again a rotation-- rotation, stretch, rotation. I'll say it again-- rotation, stretch, rotation. That's what singular values and singular vectors do, the singular value decomposition. And it's got the best of all worlds here. It's got the rotations, the orthogonal matrices. And it's got the stretches, the diagonal matrices.

Compared to those two, those are the greatest. Triangular matrices were good when we were young an hour ago. Now, we're seeing the best. OK, now let me just show you where they come from. Oh, so how to find these v 's. Well, the answer is, if I'm looking for orthogonal vectors, the great idea is find a symmetric matrix and with those eigenvectors.

So these v 's that I want for A are actually eigenvectors of this symmetric matrix A transpose times A . That's just nice. So we can find those singular vectors just as fast as we can find eigenvectors for a symmetric matrix. And we know there, because A transpose A is symmetric. We know the eigenvectors are perpendicular to each other orthonormal.

OK, and now what about the other ones because remember, we have two sets. The u 's-- well, we just multiply by A . And we've got the u 's. Well, and divide by sigmas, because these vectors u 's and v 's are unit vectors, length one. So we have to scale them properly. And this was a little key bit of algebra to check that, not only the v 's were orthogonal, but the u 's are orthogonal. Yeah, it just comes out-- comes out.

So this singular value decomposition, which is maybe, well, say 100 years old, maybe a bit more. But it's really in the last 20, 30 years that singular values have become so important. This is the best factorization of them all. And that's not always reflected in linear algebra courses. So part of my goal today is to say get to singular values.

If you've done symmetric matrices and their eigenvalues, then you can do singular values. And I think that's absolutely worth doing, OK, yeah, so and remembering down here that capital Sigma stands for the diagonal matrix of these positive numbers, σ_1 , σ_2 down to σ_r there. The rank, which came way back in the first slides, tells you how many there are.

Good, good. Oh, here's an example. So I took a small matrix because I'm doing this by pencil and paper and actually showing you that the singular value. So there is my matrix, 2 by 2. Here are the u 's. Do you see that those are orthogonal-- 1, 3 against minus 3, 1? Take the dot product, and you get 0. The v 's are orthogonal.

The sigma is diagonal. And then the pieces from that add back to the matrix. So it's really, it's broken my matrix into a couple of pieces-- one for the first singular value in vector, and the other for the second singular value in vector. And that's what data science wants. Data science wants to know what's important in the matrix?

Well, what's important is sigma 1, the big guy. Sigma 2, you see. Well, it was 3 times smaller-- $3/2$ versus $1/2$. So if I had a 100 by 100 matrix or 100 by 1,000, I'd have 100 singular values and maybe the first five I'd keep. If I'm in the financial market, those guys, those first numbers are telling me maybe what bond prices are going to do over time. And it's a mixture of a few features, but not all 1,000 features, right.

So this is singular value decomposition picks out the important part of a data matrix. And you cannot ask for a more than that. Here's what you do with a matrix is just totally enormous-- too big to multiply-- too big to compute. Then you randomly sample it. Yeah, maybe the next slide even mentions that word randomized numerical linear algebra.

So this, I'll go back to this. So the singular value decomposition-- this, what we just talked about with the u's and the v's and the sigmas. Sigma 1 is the biggest. Sigma r is the smallest. So in data science, you very often keep just these first ones, maybe the first k, the k largest ones.

And then you've got the matrix that has rank only k, because you're only working with k vectors. And it turns out that's the closest one to the big matrix A. So this singular value is among other things is picking out, putting in order of importance the little pieces of the matrix. And then you can just pick a few pieces to work with. Yeah, yeah. And the idea of norms is how to measure the size of a matrix.

Yeah, but I'll leave that for the future. And randomized linear algebra I just want to mention. Seems a little crazy that by just randomly sampling a matrix, we could learn anything about it. But typically data is sort of organized. It's not just totally random stuff. So if we want to know like, my friend in the Broad Institute was doing the ancient history of man. So data from thousands of years ago.

So he had a giant matrix-- a lot of data-- too much data. And he said, how can we find the singular value decomposition? Pick out the important thing. So you had to sample the data. Statistics is a beautiful important subject. And it leans on linear algebra. Data science leans on

linear algebra.

You are seeing the tool. Calculus would be functions would be continuous curves. Linear algebra is about vectors. This is just n components. And that's where you compute. And that's where you understand. OK.

Oh, this is maybe the last slide to just help orient you in the courses. So at MIT 18.06 is the Linear Algebra Course. And maybe you know 18.06 and also 18.06 Scholar, SC, on OpenCourseWare. And then this is the new course with the new book, 18.065. So its numbers sort of indicating a second course in linear algebra. That's when I'm actually teaching now, Monday, Wednesday, Friday.

And so that starts with linear algebra, but it's mostly about deep learning-- learning from data. So you need statistics. You need optimization, minimizing. Big functions of calculus comes into it. So that's a lot of fun to teach and to learn. And, of course, it's tremendously important in industry now. And Google and Facebook and ever so many companies need people who understand this.

And, oh, and then I am repeating 18.06 because there is this new book coming, I hope. Did some more this morning. *Linear Algebra for Everyone*. So I have optimistically put 2021. And you're the first people that know about it. So these are the websites for the two that we have.

That's the website for the linear algebra book, math.mit.edu. And this is the website for the *Learning from Data* book. So you see there the table of contents and all and solutions to problems-- lots of things.

Thanks for listening to this is-- what-- maybe four or five pieces in this 2020 vision to update the videos that have been watched so much on OpenCourseWare. Thank you.