

Okay.

This is the lecture on the singular value decomposition.

But everybody calls it the SVD.

So this is the final and best factorization of a matrix.

Let me tell you what's coming.

The factors will be, orthogonal matrix, diagonal matrix, orthogonal matrix.

So it's things that we've seen before, these special good matrices, orthogonal diagonal.

The new point is that we need two orthogonal matrices.

A can be any matrix whatsoever.

Any matrix whatsoever has this singular value decomposition, so a diagonal one in the middle, but I need two different -- probably different orthogonal matrices to be able to do this.

Okay. And this factorization has jumped into importance and is properly, I think, maybe the bringing together of everything in this course.

One thing we'll bring together is the very good family of matrices that we just studied, symmetric, positive, definite.

Do you remember the stories with those guys?

Because they were symmetric, their eigenvectors were orthogonal, so I could produce an orthogonal matrix -- this is my usual one.

My usual one is the eigenvectors and eigenvalues. In the symmetric case, the eigenvectors are orthogonal, so I've got the good -- my ordinary S has become an especially good Q .

And positive definite, my ordinary λ has become a positive λ .

So that's the singular value decomposition in case our matrix is symmetric positive definite -- in that case, I don't need two -- U and a V -- one orthogonal matrix will do for both sides.

So this would be no good in general, because usually the eigenvector matrix isn't orthogonal.

So that's not what I'm after.

I'm looking for orthogonal times diagonal times orthogonal.

And let me show you what that means and where it comes from.

Okay.

What does it mean?

You remember the picture of any linear transformation.

This was, like, the most important figure in And what I looking for now?

A typical vector in the row space -- typical vector, let me call it v_1 , gets taken over to some vector in the column space, say u_1 .

So u_1 is Av_1 .

Okay.

Now, another vector gets taken over here somewhere.

What I looking for?

In this SVD, this singular value decomposition, what I'm looking for is an orthogonal basis here that gets knocked over into an orthogonal basis over there.

See that's pretty special, to have an orthogonal basis in the row space that goes over into an orthogonal basis -- so this is like a right angle and this is a right angle -- into an orthogonal basis in the column space.

So that's our goal, is to find -- do you see how things are coming together?

First of all, can I find an orthogonal basis for this row space?

Of course.

No big deal to find an orthogonal basis.

Graham Schmidt tells me how to do it.

Start with any old basis and grind through Graham Schmidt, out comes an orthogonal basis.

But then, if I just take any old orthogonal basis, then when I multiply by A , there's no reason why it should be orthogonal over here.

So I'm looking for this special set up where A takes these basis vectors into orthogonal vectors over there.

Now, you might have noticed that the null space I didn't include.

Why don't I stick that in?

You remember our usual figure had a little null space and a little null space.

And those are no problems.

Those null spaces are going to show up as zeroes on the diagonal of σ , so that doesn't present any difficulty.

Our difficulty is to find these.

So do you see what this will mean?

This will mean that A times these v -s, v_1, v_2 , up to -- what's the dimension of this row space?

Vr. Sorry, make that V a little smaller -- up to v_r .

So that's -- Av_1 is going to be the first column, so here's what I'm achieving.

Oh, I'm not only going to make these orthogonal, but why not make them orthonormal?

Make them unit vectors.

So maybe the unit vector is here, is the u_1 , and this might be a multiple of it.

So really, what's happening is Av_1 is some multiple of u_1 ,

right? These guys will be unit vectors and they'll go over into multiples of unit vectors and the multiple I'm not going to call λ anymore.

I'm calling it σ .

So that's the number -- the stretching number.

And similarly, Av_2 is $\sigma_2 u_2$.

This is my goal.

And now I want to express that goal in matrix language.

That's the usual step.

Think of what you want and then express it as a matrix multiplication.

So Av_1 is $\sigma_1 u_1$ -- actually, here we go.

Let me pull out these -- u_1, u_2 to u_r and then a matrix with the sigmas.

Everything now is going to be in that little part of the blackboard.

Do you see that this equation says what I'm trying to do with my figure.

A times the first basis vector should be σ_1 times the other basis -- the other first basis vector.

These are the basis vectors in the row space, these are the basis vectors in the column space and these are the multiplying factors.

So Av_2 is σ_2 times u_2 , Av_r is σ_r times u_r .

And then we've got a whole lot of zeroes and maybe some zeroes at the end, but that's the heart of it.

And now if I express that in -- as matrices, because you knew that was coming -- that's what I have.

So, this is my goal, to find an orthogonal basis in the orthonormal, even -- basis in the row space and an orthonormal basis in the column space so that I've sort of diagonalized the matrix.

The matrix A is, like, getting converted to this diagonal matrix σ .

And you notice that usually I have to allow myself two different bases.

My little comment about symmetric positive definite was the one case where it's $A = Q \sigma Q^T$, where V and U are the same Q .

But mostly, you know, I'm going to take a matrix like -- oh, let me take a matrix like four four minus three three.

Okay.

There's a two by two matrix.

It's invertible, so it has rank two.

So I'm going to look for two vectors, v_1 and v_2 in the row space, and U -- so I'm going to look for v_1, v_2 in the row space, which of course is \mathbb{R}^2 .

And I'm going to look for u_1, u_2 in the column space, which of course is also \mathbb{R}^2 , and I'm going to look for numbers σ_1 and σ_2 so that it all comes out right.

So these guys are orthonormal, these guys are orthonormal and these are the scaling factors.

So I'll do that example as soon as I get the matrix picture straight.

Okay.

Do you see that this expresses what I want?

Can I just say two words about null spaces?

If there's some null space, then we want to stick in a basis for those, for that.

So here comes a basis for the null space, $v_{(r+1)}$ down to v_m .

So if we only had an r dimensional row space and the other $n-r$ dimensions were in the null space -- okay, we'll take an orthogonal -- orthonormal basis there.

No problem. And then we'll just get zeroes.

So, actually, w - those zeroes will come out on the diagonal matrix.

So I'll complete that to an orthonormal basis for the whole space, \mathbb{R}^m .

I complete this to an orthonormal basis for the whole space \mathbb{R}^n and I complete that with zeroes.

Null spaces are no problem here.

So really the true problem is in a matrix like that, which isn't symmetric, so I can't use its eigenvectors, they're not orthogonal -- but somehow I have to get these orthogonal -- in fact, orthonormal guys that make it work.

I have to find these orthonormal guys, these orthonormal guys and I want Av_1 to be $\sigma_1 u_1$ and Av_2 to be $\sigma_2 u_2$.

Okay. That's my goal.

Here's the matrices that are going to get me there.

Now these are orthogonal matrices.

I can put that -- if I multiply on both sides by V^{-1} , I have $A = U \Sigma V^{-1}$, and of course you know the other way I can write V^{-1} .

This is one of those square orthogonal matrices, so it's the same as $U \Sigma V^T$.

Okay.

Here's my problem.

I've got two orthogonal matrices here.

And I don't want to find them both at once.

So I want to cook up some expression that will make the U s disappear.

I would like to make the U s disappear and leave me only with the V s.

And here's how to do it.

It's the same combination that keeps showing up whenever we have a general rectangular matrix, then it's $A^T A$, that's the great matrix.

That's the great matrix.

That's the matrix that's symmetric, and in fact positive definite or at least positive semi-definite.

This is the matrix with nice properties, so let's see what

will it be? So if I took the transpose, then, I would have -- $A^T A^T$ will be what?

What do I have?

If I transpose that I have $V \Sigma^T U^T$, that's the A^T .

Now the A -- and what have I got?

Looks like worse, because it's got six things now together, but it's going to collapse into something good.

What does $U^T U$ collapse into?

I , the identity.

So that's the key point.

This is the identity and we don't have U anymore.

And $\Sigma^T \Sigma$, those are diagonal matrixes, so their product is just going to have Σ squared on the diagonal.

So do you see what we've got here?

This is V times this easy matrix $\Sigma_1^2 \Sigma_2^2$ times V^T .

This is the $A^T A$.

This is -- let me copy down -- $A^T A$ is that.

U is out of the picture, now.

I'm only having to choose the V s, and what are these V s?

And what are these Σ s?

Do you know what the V s are?

They're the eigenvectors that -- see, this is a perfect eigenvector, eigenvalue, $Q \lambda Q^T$ for the matrix $A^T A$.

A itself is nothing special.

But $A^T A$ will be special.

It'll be symmetric positive definite, so this will be its eigenvectors and this'll be its eigenvalues.

And the eigenvalues'll be positive because this thing's positive definite.

So this is my method.

This tells me what the V s are.

And how I going to find the U s?

Well, one way would be to look at $A A^T$.

Multiply A by A^T in the opposite order.

That will stick the V s in the middle, knock them out, and leave me with the U s.

So here's the overall picture, then.

The V s are the eigenvectors of $A^T A$.

The U s are the eigenvectors of $A A^T$, which are different.

And the sigmas are the square roots of these and the positive square roots, so we have positive sigmas.

Let me do it for that example.

This is really what you should know and be able to do for the SVD.

Okay.

Let me take that matrix.

So what's my first step?

Compute $A^T A$, because I want its eigenvectors.

Okay. So I have to compute $A^T A$.

So $A^T A$ is 4×4 minus 3×3 , and A is 4×4 minus 3×3 , and I do that multiplication and I get sixteen -- I get twenty five -- I get sixteen minus nine -- is that seven?

And it better come out symmetric.

And -- oh, okay, and then it comes out 25.

Okay. So, I want its eigenvectors and its eigenvalues.

Its eigenvectors will be the V s, its eigenvalues will be the squares of the sigmas.

Okay. What are the eigenvalues and eigenvectors of this guy?

Have you seen that two by two example enough to recognize that the eigenvectors are -- that one one is an

eigenvector?

So this here is $A^T A$.

I'm looking for its eigenvectors.

So its eigenvectors, I think, are $(1, 1)$ and $(1, -1)$, because if I multiply that matrix by $(1, 1)$, what do I get?

If I multiply that matrix by $(1, 1)$, I get $(32, 32)$, which is 32 times $(1, 1)$.

So there's the first eigenvector, and there's the eigenvalue for $A^T A$.

So I'm going to take its square root for σ .

Okay.

What's the eigenvector that goes -- eigenvalue that goes with this one?

If I do that multiplication, what do I get?

I get some multiple of $(1, -1)$, and what is that multiple?

Looks like 18 .

Okay.

So those are the two eigenvectors, but -- oh, just a moment, I didn't normalize them.

To make everything absolutely right, I ought to normalize these eigenvectors, divide by their length, square root of two.

So all these guys should be true unit vectors and, of course, that normalization didn't change the 32 and the 18 .

Okay.

So I'm happy with the V s.

Here are the V s.

So now let me put together the pieces here.

Here's my A .

Here's my A.

Let me write down A again.

If life is right, we should get U, which I don't yet know -- U I don't yet know, sigma I do now know.

What's sigma?

So I'm looking for a U sigma V transpose.

U, the diagonal guy and V transpose.

Okay.

Let's just see that come out right.

So what are the sigmas?

They're the square roots of these things.

So square root of 32 and square root of 18.

Zero zero.

Okay.

What are the Vs?

They're these two.

And I have to transpose -- maybe that just leaves me with ones -- with one over square root of two in that row and the other one is one over square root of two minus one over square root of two.

Now finally, I've got to know the Us.

Well, actually, one way to do -- since I now know all the other pieces, I could put those together and figure out what the Us are.

But let me do it the A A transpose way.

Okay. Find the Us now. u_1 and u_2 .

And what are they?

I look at $A A^T$ -- so A is supposed to be $U \Sigma V^T$, and then when I transpose that I get $V \Sigma^T U^T$.

So I'm just doing it in the opposite order, $A A^T$, and what's the good part here?

That in the middle, $V^T V$ is going to be the identity.

So this is just $U \Sigma \Sigma^T U^T$, that's some diagonal matrix with Σ^2 and U^T .

So what I'm seeing here?

I'm seeing here, again, a symmetric positive definite or at least semi-definite matrix and I'm seeing its eigenvectors and its eigenvalues.

So if I compute $A A^T$, its eigenvectors will be the things that go into U .

Okay, so I need to compute $A A^T$.

I guess I'm going to have to go -- can I move that up just a little?

Maybe a little more and do $A A^T$.

So what's A ?

Four four minus three and three.

And what's A^T ?

Four four minus three and three.

And when I do that multiplication, what do I get?

Sixteen and sixteen, thirty two.

Uh, that one comes out zero.

Oh, so this is a lucky case and that one comes out 18.

So this is an accident that $A A^T$ happens to come out diagonal, so we know easily its eigenvectors and eigenvalues.

So its eigenvectors -- what's the first eigenvector for this $A A^T$ matrix?

It's just one zero, and when I do that multiplication, I get 32 times one zero.

And the other eigenvector is just zero one and when I multiply by that I get 18.

So this is $A A^T$.

Multiplying that gives me the $32 A A^T$.

Multiplying this guy gives me First of all, I got 32 and 18 again.

Am I surprised?

You know, it's clearly not an accident.

The eigenvalues of $A A^T$ were exactly the same as the eigenvalues of -- this one was $A^T A$.

That's no surprise at all.

The eigenvalues of $A B$ are the same as the eigenvalues of $B A$.

That's a very nice fact, that eigenvalues stay the same if I switch the order of multiplication.

So no surprise to see 32 and What I learned -- first the check that things were numerically correct, but now I've learned these eigenvectors, and actually they're about as nice as can be.

They're the best orthogonal matrix, just the identity.

Okay.

So my claim is that it ought to all fit together, that these numbers should come out right.

The numbers should come out right because the matrix multiplications use the properties that we want.

Okay. Shall we just check that?

Here's the identity, so not doing anything -- square root of 32 is multiplying that row, so that square root of 32 divided by square root of two means square root of 16, four, correct?

And square root of 18 is divided by square root of two, so that leaves me square root of 9, which is three, but -- well, Professor Strang, you see the problem?

Why is that -- okay.

Why I getting minus three three here and here I'm getting three minus three?

Phooey.

I don't know why.

It shouldn't have happened, but it did.

Now, okay, you could say, well, just -- the eigenvector there could have -- I could have had the minus sign here for that eigenvector, but I'm not happy about that.

Hmm.

Okay. So I realize there's a little catch here somewhere and I may not see it until Wednesday.

Which then gives you a very important reason to come back on Wednesday, to catch that sine difference.

So what did I do illegally?

I think I put the eigenvectors in that matrix V transpose -- okay, I'm going to have to think.

Why did that come out with with the opposite sines?

So you see -- I mean, if I had a minus there, I would be all right, but I don't want that.

I want positive entries down the diagonal of σ squared.

Okay.

It'll come to me, but, I'm going to leave this example to finish.

Okay.

And the beauty of, these sliding boards is I can make that go away.

Can I,-- let me not do it, though, yet.

Let me take a second example.

Let me take a second example where the matrix is singular.

So rank one.

Okay, so let me take as an example two, where my matrix A is going to be rectangular again -- let me just make it four three eight six.

Okay.

That's a rank one matrix.

So that has a null space and only a one dimensional row space and column space.

So actually, my picture becomes easy for this matrix, because what's my row space for this one?

So this is two by two.

So my pictures are both two dimensional.

My row space is all multiples of that vector four three.

So the whole -- the row space is just a line, right?

That's the row space.

And the null space, of course, is the perpendicular line.

So the row space for this matrix is multiples of four three.

Typical row.

Okay.

What's the column space?

The columns are all multiples of four eight, three six, one two.

The column space, then, goes in, like, this direction.

So the column space -- when I look at those columns, the column space -- so it's only one dimensional, because the rank is one.

It's multiples of four eight.

Okay.

And what's the null space of A transpose?

It's the perpendicular guy.

So this was the null space of A and this is the null space of A transpose.

Okay.

What I want to say here is that choosing these orthogonal bases for the row space and the column space is, like, no problem.

They're only one dimensional.

So what should V be?

V should be -- v_1 , but -- yes, v_1 , rather -- v_1 is supposed to be a unit vector.

There's only one v_1 to choose here, only one dimension in the row space.

I just want to make it a unit vector.

So v_1 will be -- it'll be this vector, but made into a unit vector, so four -- point eight point six.

Four fifths, three fifths.

And what will be u_1 ? u_1 will be the unit vector there.

So I want to turn four eight or one two into a unit vector, so u_1 will be -- let's see, if it's one two, then what multiple of one two do I want?

That has length square root of five, so I have to divide by square root of five.

Let me complete the singular value decomposition for this matrix.

So this matrix, four three eight six, is -- so I know what u_1 -- here's A and I want to get U the basis in the column space.

And it has to start with this guy, one over square root of five two over square root of five.

Then I want the sigma.

Okay. What are we expecting now for sigma?

This is only a rank one matrix.

We're only expecting a sigma one, which I have to find, but zeroes here.

Okay.

So what's sigma one?

It should be the -- where did these sigmas come from?

They came from $A^T A$, so I -- can I do that little calculation over here?

$A^T A$ is four three -- four three eight six times four three eight six.

This had better -- this is a rank one matrix, this is going to be -- the whole thing will have rank one, that's 16 and 64 is 80, 12 and 48 is 60, 12 and 48 is 60, 9 and 36 is 45.

Okay.

It's a rank one matrix.

Of course.

Every row is a multiple of four three.

And what's the eigen -- what are the eigenvalues of that matrix?

So this is the calculation -- this is like practicing,

now. What are the eigenvalues of this rank one matrix?

Well, tell me one eigenvalue, since the rank is only one, one eigenvalue is going to be zero.

And then you know that the other eigenvalue is going to be a hundred and twenty five.

So that's sigma squared, right, in $A^T A$.

So this will be the square root of a hundred and twenty five.

And then finally, the V^T -- the V s will be -- there's v_1 , and what's v_2 ?

What's v_2 in the -- how do I make this into an orthonormal basis?

Well, v_2 is, in the null space direction.

It's perpendicular to that, so point six and minus point

eight. So those are the V s that go in here.

Point eight, point six and point six minus point eight.

Okay. And I guess I better finish this guy.

So this guy, all I want is to complete the orthonormal basis -- it'll be coming from there.

It'll be a two over square root of five and a minus one over square root of five.

Let me take square root of five out of that matrix to make it look better.

So one over square root of five times one two two minus one.

Okay.

So there I have -- including the square root of five -- that's an orthogonal matrix, that's an orthogonal matrix, that's a diagonal matrix and its rank is only one.

And now if I do that multiplication, I pray that it comes out right.

The square root of five will cancel into that square root of one twenty five and leave me with the square root of 25, which is five, and five will multiply these numbers and I'll get whole numbers and out will come A .

Okay.

That's like a second example showing how the null space guy -- so this -- this vector and this one were multiplied by this zero.

So they were easy to deal with.

The key ones are the ones in the column space and the row space.

Do you see how I'm getting columns here, diagonal here, rows here, coming together to produce A .

Okay, that's the singular value decomposition.

So, let me think what I want to add to complete this topic.

So that's two examples.

And now let's think what we're really doing.

We're choosing the right basis for the four subspaces of linear algebra.

Let me write this down.

So v_1 up to v_r is an orthonormal basis for the row space. u_1 up to u_r is an orthonormal basis for the column space.

And then I just finish those out by $v_{(r+1)}$, the rest up to v_n is an orthonormal basis for the null space.

And finally, $u_{(r+1)}$ up to is an orthonormal basis for the null space of A transpose.

Do you see that we finally got the bases right?

They're right because they're orthonormal, and also -- again, Gram Schmidt would have done this in chapter four.

Here we needed eigenvalues, because these bases make the matrix diagonal.

A times V is a multiple of U .

So I'll put "and" -- the matrix has been made diagonal.

When we choose these bases, there's no coupling between V s and no coupling between U s.

Each A -- A times each V is in the direction of the corresponding U .

So it's exactly the right basis for the four fundamental subspaces.

And of course, their dimensions are what we know.

The dimension of the row space is the rank r , and so is the dimension of the column space.

The dimension of the null space is $n-r$, that's how many vectors we need, and $m-r$ basis vectors for the left null space, the null space of A transpose.

Okay. I'm going to stop there.

I could develop further from the SVD, but we'll see it again in the very last lectures of the course.

So there's the SVD.

Thanks.