

MITOCW | 18.06_L17

OK, here's the last lecture in the chapter on orthogonality.

So we met orthogonal vectors, two vectors, we met orthogonal subspaces, like the row space and null space.

Now today we meet an orthogonal basis, and an orthogonal matrix.

So we really -- this chapter cleans up orthogonality.

And really I want -- I should use the word orthonormal.

Orthogonal is -- so my vectors are q_1, q_2 up to q_n -- I use the letter "q", here, to remind me, I'm talking about orthogonal things, not just any vectors, but orthogonal ones.

So what does that mean?

That means that every q is orthogonal to every other q .

It's a natural idea, to have a basis that's headed off at ninety-degree angles, the inner products are all zero.

Of course if q is -- certainly q_i is not orthogonal to itself.

But there we'll make the best choice again, make it a unit vector.

Then $q_i^T q_i$ is one, for a unit vector.

The length squared is one.

And that's what I would use the word normal.

So for this part, normalized, unit length for this part.

OK. So first part of the lecture is how does having an orthonormal basis make things nice?

It certainly does.

It makes all the calculations better, a whole lot of numerical linear algebra is built around working with orthonormal vectors, because they never get out of hand, they never overflow or underflow.

And I'll put them into a matrix Q , and then the second part of the lecture will be suppose my basis, my columns of A are not orthonormal.

How do I make them so?

And the two names associated with that simple idea are Gram and Schmidt.

So the first part is we've got a basis like this.

Let's put those into the columns of a matrix.

So a matrix Q that has -- I'll put these orthonormal vectors, q_1 will be the first column, q_n will be the n -th column.

And I want to say, I want to write this property, $q_i^T q_j = 0$ for $i \neq j$, I want to put that in a matrix form.

And just the right thing is to look at $Q^T Q$.

So this chapter has been looking at $A^T A$.

So it's natural to look at $Q^T Q$.

And the beauty is it comes out perfectly.

Because Q^T has these vectors in its rows, the first row is q_1^T , the n th row is q_n^T .

So that's $Q^T Q$.

And now I want to multiply by Q .

That has q_1 along to q_n in the columns.

That's Q .

And what do I get?

You really -- this is the first simplest most basic fact, that how do orthonormal vectors, orthonormal columns in a matrix, what happens if I compute $Q^T Q$?

Do you see it?

If I take the first row times the first column, what do I get?

A one.

If I take the first row times the second column, what do I get?

Zero.

That's the orthogonality.

The first row times the last column is zero.

And so I'm getting ones on the diagonal and I'm getting zeroes everywhere else.

I'm getting the identity matrix.

You see how that's -- it's just like the right calculation to do.

If you have orthonormal columns, and the matrix doesn't have to be square here.

We might have just two columns.

And they might have four, lots of components.

So but they're orthonormal, and when we do $Q^T Q$, that $Q^T Q$ or $A^T A$ just asks for all those dot products.

Rows times columns.

And in this orthonormal case, we get the best possible answer, the identity.

OK, so this is -- so I mean now we have a new bunch of important matrices.

What have we seen previously?

We've seen in the distant past we had triangular matrices, diagonal matrices, permutation matrices, that was early chapters, then we had row echelon forms, then in this chapter we've already seen projection matrices, and now we're seeing this new class of matrices with orthonormal columns.

That's a very long expression.

I sorry that I can't just call them orthogonal matrices.

But that word orthogonal matrices -- or maybe I should be able to call it orthonormal matrices, why don't we call it orthonormal -- I mean that would be an absolutely perfect name.

For Q , call it an orthonormal matrix because its columns are orthonormal.

OK, but the convention is that we only use that name orthogonal matrix, we only use this -- this word orthogonal,

we don't even say orthonormal for some unknown reason, matrix when it's square.

So in the case when this is a square matrix, that's the case we call it an orthogonal matrix.

And what's special about the case when it's square?

When it's a square matrix, we've got its inverse, so -- so in the case if Q is square, then $Q^T Q = I$ tells us -- let me write that underneath -- tells us that Q^T is Q^{-1} .

There we have the easy to remember property for a square matrix with orthonormal columns.

That -- I need to write some examples down.

Let's see.

Some examples like if I take any -- so examples, let's do some examples.

Any permutation matrix, let me take just some random permutation matrix.

Permutation Q equals let's say oh, make it three by three, say zero, zero, one, one, zero, zero, zero, one, zero.

OK. That certainly has unit vectors in its columns.

Those vectors are certainly perpendicular to each other.

And if I -- and so that's it.

That makes it a Q .

And -- if I took its transpose, if I multiplied by Q^T , shall I do that -- and let me stick in Q^T

here. Just to do that multiplication once more, transpose it'll put the -- make that into a column, make that into a column, make that into a column.

And the transpose is also -- another Q .

Another orthonormal matrix.

And when I multiply that product I get I .

OK, so there's an example.

And actually there's a second example.

But those are real easy examples, right, I mean to get orthogonal columns by just putting ones in different places is like too easy.

So let me keep going with examples.

So here's another simple example.

Cos theta sine theta, there's a unit vector, oh, let me even take it, well, yeah.

Cos theta sine theta and now the other way I want sine theta cos theta.

But I want the inner product to be zero.

And if I put a minus there, it'll do it.

So that's -- unit vector, that's a unit vector.

And if I take the dot product, I get minus plus zero.

OK. For example Q equals say one, one, one, minus one, is that an orthogonal matrix?

I've got orthogonal columns there, but it's not quite an orthogonal matrix.

How shall I fix it to be an orthogonal matrix?

Well, what's the length of those column vectors, the dot product with themselves is -- right now it's two, right, the -- the length squared.

The length squared would be one plus one would be two, the length would be square root of two, so I better divide by square root of two.

OK. So there's a -- there now I have got an orthogonal matrix, in fact, it's this one -- when theta is pi over four.

The cosines and well almost, I guess the minus sine is down there, so maybe, I don't know, maybe minus pi over four or something.

OK.

Let me do one final example, just to show that you can get bigger ones.

Q equals let me take that matrix up in the corner and I'll sort of repeat that pattern, repeat it again, and then minus

it down here.

That's one of the world's favorite orthogonal matrices.

I hope I got it right, is -- can you see whether -- if I take the inner product of one column with another one, let's see, if I take the inner product of that column with that I have two minuses and two pluses, that's good.

When I take the inner product of that with that I have a plus and a minus, a minus and a plus.

Good. I think it all works out.

And what do I have to divide by now?

To make those into unit vectors.

Right now the vector one, one, one, one has length two.

Square root of four.

So I have to divide by two to make it unit vector, so there's another.

That's my entire array of simple examples.

This construction is named after a guy called Adhemar and we know how to do it for two, four, sixteen, sixty-four and so on, but we -- nobody knows exactly which size matrices have -- which size -- which sizes allow orthogonal matrices of ones and minus ones.

So Adhemar matrix is an orthogonal matrix that's got ones and minus ones, and a lot of ones -- some we know, some other sizes, there couldn't be a five by five I think.

But there are some sizes that nobody yet knows whether there could be or can't be a matrix like that.

OK.

You see those orthogonal matrices.

Now let me ask what -- why is it good to have orthogonal matrices?

What calculation is made easy?

If I have an orthogonal matrix.

And -- let me remember that the matrix could be rectangular.

Shall I put down -- I better put a rectangular example down.

So the -- these were all square examples.

Can I put down just -- a rectangular one just to be sure that we realize that this is possible.

let's help me out.

Let's see, if I put like a one, two, two and a minus two, minus one, two.

That's -- a matrix -- oh its columns aren't normalized yet.

I always have to remember to do that.

I always do that last because it's easy to do.

What's the length of those columns?

So if I wanted them -- if I wanted them to be length one, I should divide by their length, which is -- so I'd look at one squared plus two squared plus two squared, that's one and four and four is nine, so I take the square root and I need to divide by three.

OK. So there is -- well, without that, I've got one orthonormal vector.

I mean just one unit vector.

Now put that guy in.

Now I have a basis for the column space for a two-dimensional space, an orthonormal basis, right?

These two columns are orthonormal, they would be an orthonormal basis for this two-dimensional space that they span.

Orthonormal vectors by the way have got to be independent.

It's easy to show that orthonormal vectors since they're headed off all at ninety degrees there's no combination that gives zero.

Now if I wanted to create now a third one, I could either just put in some third vector that was independent and go to this Gram-Schmidt calculation that I'm going to explain, or I could be inspired and say look, that -- with that

pattern, why not put a one in there, and a two in there, and a two in there, and try to fix up the signs so that they worked.

Hmm.

I don't know if I've done this too brilliantly.

Let's see, what signs, that's minus, maybe I'd make a minus sign there, how would that be?

Yeah, maybe that works.

I think that those three columns are orthonormal and they -- the beauty of this -- this is the last example I'll probably find where there's no square root, the -- the punishing thing in Gram-Schmidt, maybe we better know that in advance, is that because I want these vectors to be unit vectors, I'm always running into square roots.

I'm always dividing by lengths.

And those lengths are square roots.

So you'll see as soon as I do a Gram-Schmidt example, square roots are going to show up.

But here are some examples where we did it without any square root.

OK. OK.

So -- so great.

Now next question is what's the good of having a Q ?

What formulas become easier?

Suppose I want to project, so suppose Q -- suppose Q has orthonormal columns.

I'm using the letter Q to mean this, I'll write it this one more time, but I always mean when I write a Q , I always mean that it has orthonormal columns.

So suppose I want to project onto its column space.

So what's the projection matrix?

What's the projection matrix is I project onto a column space?

OK, that gives me a chance to review the projection section, including that big formula, which used to be -- those four A s in a row, but now it's got Q s, because I'm projecting onto the column space of Q , so do you remember what it was?

It's $Q Q^T Q^{-1} Q^T$.

That's my four Q s in a row.

But what's good here?

What -- what makes this formula nice if I'm projecting onto a column space when I have orthonormal basis for that space?

What makes it nice is this is the identity.

I don't have to do any inversion.

I just get $Q Q^T$.

So $Q Q^T$ is a projection matrix.

Oh, I can't help -- I can't resist just checking the properties, what are the properties of a projection matrix?

There are two properties to know for any projection matrix.

And I'm saying that this is the right projection matrix when we've got this orthonormal basis in the columns.

OK. So there's the projection matrix.

Suppose the matrix is square.

First just tell me first this extreme case.

If my matrix is square and it's got these orthonormal columns, then what's the column space?

If I have a square matrix and I have independent columns, and even orthonormal columns, then the column space is the whole space, right?

And what's the projection matrix onto the whole space?

The identity matrix.

If I'm projecting in the whole space, every vector B is right where it's supposed to be and I don't have to move it by projection.

So this would be -- I'll put in parentheses this is I if Q is square.

Well that we said that already.

If Q is square, that's the case where Q transpose is Q inverse, we can put it on the right, we can put it on the left, we always get the identity matrix, if it's square.

But if it's not a square matrix then it's not -- we don't get the identity matrix.

We have $Q Q$ transpose, and just again what are those two properties of a projection matrix?

First of all, it's symmetric.

OK, no problem, that's certainly a symmetric

So what's that second property of a projection? matrix.

That if you project and project again you don't move the second

time. So the other property of a projection matrix should be that $Q Q$ transpose twice should be the same as $Q Q$ transpose once.

That's projection matrices.

And that property better fall out right away because from the fact we know about orthonormal matrices, Q transpose Q is I .

OK, you see it.

In the middle here is sitting $Q Q^t Q$ transpose Q , sorry, that's what I meant to say, Q transpose Q is I .

So that's sitting right in the middle, that cancels out, to give the identity, we're left with one $Q Q$ transpose, and we're all set.

OK.

So this is the projection matrix -- all the equation -- all the messy equations of this chapter become trivial when our matrix -- when we have this orthonormal basis.

I mean what do I mean by all the equations?

Well, the most important equation was the normal equation, do you remember old $A^T \hat{x} = A^T b$?

But now -- now A is Q .

Now I'm thinking I have $Q^T Q \hat{x} = Q^T b$.

And what's good about that?

What's good is that matrix on the left side is the identity.

The matrix on the left is the identity, $Q^T Q$, normally it isn't, normally it's that matrix of inner products and you've to compute all those dopey inner products and -- and -- and solve the system.

Here the inner products are all one or zero.

This is the identity matrix.

It's gone.

And there's the answer.

There's no inversion involved.

Each component of x is a Q times b .

What that equation is saying is that the i -th component is the i -th basis vector times b .

That's -- probably the most important formula in some major parts of mathematics, that if we have orthonormal basis, then the component in the -- in the i -th, along the i -th -- the projection on the i -th basis vector is just $q_i^T b$.

That number x that we look for is just a dot product.

OK.

OK, so I'm ready now for the sort of like second half of the lecture.

Where we don't start with an orthogonal matrix, orthonormal vectors.

We just start with independent vectors and we want to make them orthonormal.

So I'm going to -- can I do that now?

Now here comes Graham-Schmidt.

So -- Graham-Schmidt.

So this is a calculation, I won't say -- I can't quite say it's like elimination, because it's different, our goal isn't triangular anymore.

With elimination our goal was make the matrix triangular.

Now our goal is make the matrix orthogonal.

Make those columns orthonormal.

So let me start with two columns.

So I start with vectors a and b .

And they're just like -- here, let me draw them.

Here's a .

Here's b .

For example.

a isn't specially horizontal, wasn't meant to be, just a is one vector, b is another.

I want to produce those two vectors, they might be in twelve-dimensional space, or they might be in two-dimensional space.

They're independent, anyway.

So I better be sure I say that.

I start with independent vectors.

And I want to produce out of that q_1 and q_2 , I want to produce orthonormal vectors.

And Graham and Schmidt tell me how.

OK. Well, actually you could tell me how, we don't need -- frankly, I don't know -- there's only one idea here, if Graham had the idea, I don't know what Schmidt did.

But OK.

So you'll see it.

We don't need either of them, actually.

OK, so what I going to do.

I'll take that -- this first guy.

OK.

Well, he's fine.

That direction is fine except -- yeah, I'll say OK, I'll settle for that direction.

So I'm going to -- I'm going to get, so what I going to -- my goal is I'm going to get orthogonal vectors and I'll call those capital A and B.

So that's the key step is to get from any two vectors to two orthogonal vectors.

And then at the end, no problem, I'll get orthonormal vectors, how will -- what will those will be my q_1 and q_2 , and what will they

be? Once I've got A and B orthogonal, well, look, it's no big deal -- maybe that's what Schmidt did, he, brilliant Schmidt, thought OK, divide by the length, all right.

That's Schmidt's contribution.

OK. But Graham had a little more thinking to do, right?

We haven't done Graham's part.

This part except OK, I'm happy with A, A can be A.

That first direction is fine.

Why should -- no complaint about that.

The trouble is the second direction is not fine.

Because it's not orthogonal to the first.

I'm looking for a vector that's -- starts with B, but makes it orthogonal to A.

What's the vector?

How do I do that?

How do I produce from this vector a piece that's orthogonal to this one?

And the -- remember these vectors might be in two dimensions or they might be in twelve dimensions.

I'm just looking for the idea.

So what's the idea?

Where did we have orthogonal -- a vector showing up that was orthogonal to this guy?

Well, that was the first basic calculation of the whole chapter.

We -- we did a projection and the projection gave us this part, which was the part in the A direction.

Now, the part we want is the other part, the e part.

This part.

This is going to be our -- that guy is that guy.

This is our vector B.

That gives us that ninety-degree angle.

So B is you could say -- B is really what we previously called

e. The error vector.

And what is it?

I mean what do I -- what is B here?

A is A, no problem.

B is -- OK, what's this error piece?

Do you remember?

It's I start with the original B and I take away what?

Its projection, this P.

This -- the vector B, this error vector, is the original vector removing the projection.

So instead of wanting the projection, now that's what I want to throw away.

I want to get the part that's perpendicular.

And there will be a perpendicular part, it won't be zero.

Because these vectors were independent, so B -- if B was along the direction of A, then if the original B and A were in the same direction, then I'm -- I've only got one direction.

But here they're in two independent directions and all I'm doing is getting that guy.

So what's its formula?

What's the formula for that if -- I want to subtract the projection, so do you remember the projection?

It's some multiple of A and what's that multiple?

It's -- it's that thing we called x in the very very first lecture on this chapter.

There's an A transpose A in the bottom and there's an A transpose B, isn't that it?

I think that's Graham's formula.

Or Graham-Schmidt.

No, that's Graham.

Schmidt has got to divide the whole thing by the length, so he -- his formula makes a mess which I'm not willing to write down.

So let's just see that what I saying here?

I'm saying that this vector is perpendicular to A.

That these are orthogonal.

A is perpendicular to B.

Can you check that?

How do you see that yes, of course, we -- our picture is telling us, yes, we did it right.

How would I check that this matrix is perpendicular to A?

I would multiply by A transpose and I better get zero, right?

I should check that.

A transpose B should come out zero.

So this is A transpose times -- now what did we say B was?

We start with the original B, and we take away this projection, and that should come out zero.

Well, here we get an A transpose B minus -- and here is another A transpose B, and the -- and it's an A transpose A over A transpose A, a one, those cancel, and we do get zero.

Right.

Now I guess I can do numbers in there.

But I have to take a third vector to be sure we've got this system down.

So now I have to say if I have independent vectors A, B and C, I'm looking for orthogonal vectors A, B and capital C, and then of course the third guy will just be C over its length, the unit vector.

So this is now the problem.

I got B here.

I got A very easily.

And now -- if you see the idea, we could figure out a formula for C.

So now that -- so this is like a typical homework quiz problem.

I give you two vectors, you do this, I give you three vectors, and you have to make them orthonormal.

So you do this again, the first vector's fine, the second vector is perpendicular to the first, and now I need a third vector that's perpendicular to the first one and the second one.

Right?

This is the end of a -- the lecture is to find this guy.

Find this vector -- this vector C , that's perpendicular we n - at this point we know A and B .

But C , the little c that we're given, is off in some -- it's got to come out of the blackboard to be independent, so -- so can I sort of draw off -- off comes a c somewhere.

I don't know, where I going to put the darn

thing? Maybe I'll put it off, oh, I don't know, like that somehow, C , little c .

And I already know that perpendicular direction, that one and that one.

So now what's the idea?

Give me the Gram-Schmidt formula for C .

What is this C here?

Equals what?

What I going to do?

I'll start with the given one.

As before.

Right? I start with the vector I'm given.

And what do I do with it?

I want to remove out of it, I want to subtract off, so I'll put a minus sign in, I want to subtract off its components in the A , capital A and capital B directions.

I just want to get those out of there.

Well, I know how to do that.

I did it with B.

So I'll just -- so let me take away -- what if I do this?

What have I done?

I've got little c and what have I subtracted from it?

Its component, its projection if you like, in the A direction.

And now I've got to subtract off its component $B^T C$ over $B^T B$, that multiple of B , is its component in the B direction.

And that gives me the vector capital C that if anything is -- if there's any justice, this C should be perpendicular to A and it should be perpendicular to B .

And the only thing it's -- hasn't got is unit vector, so we divide by its length to get that too.

OK.

Let me do an example.

Can I -- I'll make my life easy, I'll just take two vectors.

So let me do a numerical example.

If I'll give you two vectors, you give me back the Gram-Schmidt orthonormal basis, and we'll see how to express that in matrix form.

OK. So let me give you the two vectors.

So I'll take the vector A equals let's say one, one, one, why not?

And B equals let's say one, zero, two, OK?

I didn't want to cheat and make them orthogonal in the first place because then Gram-Schmidt wouldn't be needed.

OK.

So those are not orthogonal.

So what is capital A?

Well that's the same as big A.

That was fine.

What's B?

So B is this b -- is the original B, and then I subtract off some multiple of the A.

And what's the multiple?

What goes in here?

B -- here's the A -- this is the -- this is the little b , this is the big A, also the little a , and I want to multiply it by that right -- that right ratio, which has A transpose A, here's my ratio.

I'm just doing this.

So it's A transpose B, what is A transpose B, it looks like three.

And what is A -- oh, my -- what's A transpose A?

Three.

I'm sorry.

I didn't know that was going to happen.

OK. But it happened.

Why should we knock it?

OK. So do you see it all right?

That's A transpose B, there's A transpose A, that's the fraction, so I take this away, and I get one take away one is a zero, zero minus this one is a minus one, and two minus the one is a one.

OK.

And what's this vector that we finally found?

This is B.

And how do I know it's right?

How do I know I've got a vector I want?

I check that B is perpendicular to -- that A and B are perpendicular.

That A is perpendicular to B.

Just look at that.

That one -- the dot product of that with that is zero.

OK. So now what is my q_1 and q_2 ?

Why don't I put them in a matrix?

Of course.

Since I'm always putting these -- so the Q, I'll put the q_1 and the q_2 in a matrix.

And what are they?

Now when I'm writing q-s I'm supposed to make things normalized.

I'm supposed to make things unit vectors.

So I'm going to take that A but I'm going to divide it by square root of three.

And I'm going to take this B but I'm going to divide it by square root of two to make it a unit vector, and there is my matrix.

That's my matrix with orthonormal columns coming from Gram-Schmidt and it sort of it -- it came from the original one, one, one, one, zero, two, right?

That was my original guys.

These were the two I started with.

These are the two that I'm happy to end with.

Because those are orthonormal.

So that's what Gram-Schmidt did.

It -- well, tell me about the column spaces of these matrices.

How is the column space of Q related to the column space of

A ? So I'm always asking you things like this, and that makes you think, OK, the column space is all combinations of the columns, it's that plane,

right? I've got two vectors in three-dimensional space, their column space is a plane, the column space of this matrix is a plane, what's the relation between the planes?

Between the two column spaces?

They're one and the same, right?

It's the same column space.

All I'm taking is here this B thing that I computed, this B thing that I computed is a combination of B and A , and A was little A , so I'm always working here with this in the same space.

I'm just like getting ninety-degree angles in there.

Where my original column space had a perfectly good basis, but it wasn't as good as this basis, because it wasn't orthonormal.

Now this one is orthonormal, and I have a basis then that -- so now projections, all the calculations I would ever want to do are -- are a cinch with this orthonormal basis.

One final point.

One final point in this chapter.

And it's -- just like elimination.

We learned how to do elimination, we know all the steps, we can do it.

But then I came back to it and said look at it as a matrix in matrix language and elimination gave me -- what was elimination in matrix language?

I'll just put it up there.

A was LU.

That was matrix, that was elimination.

Now, I want to do the same for Gram-Schmidt.

Everybody who works in linear algebra isn't going to write out the columns are orthogonal, or orthonormal.

And isn't going to write out these formulas.

They're going to write out the connection between the matrix A and the matrix Q.

And the two matrices have the same column space, but there's some -- some matrix is taking the -- and I'm going to call it R, so $A = QR$ is the magic formula here.

It's the expression of Gram-Schmidt.

And I'll -- let me just capture that.

So that's the -- my final step then is $A = QR$.

Maybe I can squeeze it in here.

So A has columns, let's say a_1 and a_2 .

Let me suppose n is two, just two vectors.

OK.

So that's some combination of q_1 and q_2 .

And times some matrix R.

They have the same column space.

This is just -- this matrix just includes in it whatever these numbers like $\frac{3}{3}$ and $\frac{1}{\sqrt{3}}$ and $\frac{1}{\sqrt{2}}$, probably that's what it's got.

One over square root of three, one over square root of two, something there, but actually it's got a zero there.

So the main point about this $A = QR$ is this R turns out to be upper triangular.

It turns out that this R is upper triangular.

We could see why.

Let me see, I can put in general formulas for what these

This I think in here should be the inner product of a_1 with q_1 . are.

And this one should be the -- the inner product of a_1 with q_2 .

And that's what I believe is zero.

This will be something here, and this will be something here with inner -- $a_1^T q_2$, sorry $a_2^T q_1$ and $a_2^T q_2$.

But why is that guy zero?

Why is $a_1^T q_2$ zero?

That's the key to this being -- this R here being upper triangular.

You know why $a_1^T q_2$ is zero, because a_1 -- that was my -- this was really a and b here.

This was really a and b .

So this is $a^T q_2$.

And the whole point of Gram-Schmidt was that we constructed these later q -s to be perpendicular to the earlier vectors, to the earlier -- all the earlier vectors.

So that's why we get a triangular matrix.

The -- result is extremely satisfactory.

That if I have a matrix with independent columns, the Gram-Schmidt produces a matrix with orthonormal columns, and the connection between those is a triangular matrix.

That last point, that the connection is a triangular matrix, please look in the book, you have to see that one more time.

OK.

Thanks, that's great.