[SQUEAKING]

[RUSTLING]

[CLICKING]

ALAN EDELMAN: So a quick problem. You might want to grab a pen and paper, and-- I think it's not that hard, but let's just ask.

Let's say we have this function, which is just the 2-norm of x From Rn to R. So everybody knows this function, of course. It's the square root of x1 squared up to xn squared. So a vector goes in, a scalar comes out. And my simple question is what is the gradient of this function?

And the reason why I'm putting this up in part is because I suspect everybody in the room can do this the 1802 way with indices. And if you'd like to do it that way because that's where your comfort is, that's fine. But I think the real challenge is if you're-- to exercise what you've learned so far is to see if you can do this without indices, what I call the grown up way.

There's no right and wrong. If you get the right answer, that's fine. But of course, I think that the comfortable way is, well, comfortable for everybody, and the grown up way takes some practice. So let me put this up and I'll go through it in a moment.

So a simple question, the gradient of this function. And I think the 1802 way might be to start with the first component, but pretty quickly, you realize that every component's the same. So you might start calculating, for example, d dx1 of the square root of x1 squared plus x1 squared.

And using all the calculus you remember, the derivative of something to the one half is the same thing to the negative one half. And then you'd get 2x1, so this would be 1 times 2x1. And you would have x1 over the square root of x1 squared plus x1 squared.

And it wouldn't take you long to realize that if you did x2, there's no real difference. And before you knew it, you would find out that the gradient, of course, would be the vector x divided by its length. That gives you every-- and that would be what I would call the 1802 way.

And it's not so bad. But in this class, we're trying to see if we can avoid using indices. Sometimes, you can't. Many times you can. And so we want to emphasize the ability to not use indices.

And so I don't know, has anyone already written it out for the-- or convinced themselves they could do it without indices yet? Has anyone done it yet? But you think you can do it?

So I'm going to write down r is equal to the length of x. And therefore, r squared is x transpose x. This is the linear algebra way to set things up.

And then I'm just going to take the differential of both sides and get to 2rdr equals, and we know this differential, so I can just write it down. But if you forgot it, you can get it yourself.

Sorry, this is d r. It's 2x transpose dx. And so therefore, dr is equal to x transpose over rdx, and then the gradient is the transpose of this, which is x over r.

So no indices. That was kind of the trick here, just to be able to do this thing without any indices. Kind of like-- it's sort of like superpowers for vectors and matrices. Like I said, this isn't hard.

What happens, though, is as things get a little more complicated, being good at this non-index approach gets more and more valuable. Because there's no one right way to do it. You could grind it through, no matter how complicated.

I bet you all can do it. But this is the fun way to do it. there's this thing over here. Any questions about that? OK.

So I don't know if you've looked at the homework yet, but the first problem has the transformations of the plane with the corgis, and we're looking for 2 by 2 Jacobian matrices. And in one case, there's-- in part D where there's the warp, there's a matrix function of x that's multiplied by x. If the matrix didn't depend on x, the Jacobian would just be the matrix.

But because the matrix depends on x, there's going to be two terms. There's going to be roughly D matrix of x times x plus-- this is kind of the template. And somehow, a dx is going to come out of this term as well. And so this term is the easy one if the matrix was constant and this one takes a little bit more work. So that's problem D on your homework in the first part.

Good. Well, then I'm going to go back and do what I was starting to do last time. And I think it just makes sense to go back to the beginning since we've had a whole weekend since last time. And so matrix Jacobian.

So this Julia notebook here-- and by the way, if anybody has questions about Julia, feel free to ask about that as well. But the goal here-- this blackboard example is a scalar function of a vector. And I'm really excited about functions where the input and output are actually matrices themselves. So matrix squared, matrix inverse we've seen, matrix cubed.

I like to think of the LU decomposition as a function of matrices. I don't know if you've ever thought about it. But the LU decomposition, or even the eigendecomposition of a matrix is a function from matrices to pieces of matrices, in effect. So we'll see how that goes through.

So let me go back to the beginning here. Just very quickly, we're using Julia's symbolics, which is kind of fun to do. And this shows up later. I don't know why it's at the beginning. So we'll just skip right past it.

But here's the matrix that we'll work with a lot. This is the symbolic matrix that has p, q, r, s going in the column major order. And so I just declared p, q, and r and s to be variables, as well as theta. So Julia lets me put them in as elements of matrices. OK, so simple thing.

And last time, I introduced the vec command, which basically squashes a 2 by 2 matrix, or any-- in fact, any matrix down to a vector. All you do with vec, in case you missed it last time, is you take the first column and stack it on the second column, stack it on the third column, and so forth.

And this is, again, what we show students because it's concrete. But as far as I can tell, you never really need to vec things other than maybe in your mind. I mean, it's another one of these comfort things that I don't think you really have to do in practice. But it does have a way of allowing you to use one notation for many, many things.

So there's the vec command. It just squashes a matrix. And here's the matrix square of a 2 by 2 matrix. All of you can do this. Just take that p, q, r, s matrix and do that dot producty thing, and you get the matrix square, which we can vec. And maybe I'll make this just a little bit smaller so you can get a little more real estate. Yeah. OK, good.

So this is the vec of the matrix square. And so one way to think of matrix square is a function from r4 to r4. It's the input are the variables p, q, r, and s, and the output are these four functions of p, q, r, and s, the p squared plus qr all the way down to qr plus s squared.

So that that's a function that takes r4 to r4. But really we all know it's a function from 2 by 2 matrices to 2 by 2 matrices. r4 is just a crutch. It's really 2 by 2 matrices to 2 by 2 matrices. And so we can calculate the Jacobian of this thing.

And I'll just remind you that the Jacobian is nothing other than all the derivatives of each of these four expressions with respect to all of the four input variables p, q, r, and s. So this output p squared plus qr belongs to the first row. And we'll take the derivative with respect p, and then q, and then r, and then s.

And if you do that, you'll get this first row, and so forth. So the inputs are the columns, the outputs are the rows, and that's the Jacobian matrix. And Julia is happy to compute it for you, but you could have all done this yourself as well. So there's the Jacobian.

And our goal is to-- our goal is really to find a notation that lets you write this matrix without writing the matrix. That's what we're trying to do here. We'd like to write this matrix, but without writing the matrix. That's what we're heading for.

But I love to check things over. I like to do things with symbols, if I can. I like to do things with numbers, if I can. It makes me happy when things come out equal.

So let's do this numerically. So here's-- my inputs are going to be 1, 2, 3, 4. And I'm going to make some perturbations of the 1, 2, 3, 4 by just taking 1, 2, 3, 4 divided by 10,000. And I'm going to put-- into this Jacobian, I'm just going to put in these numbers, this Jacobian.

So this is the Jacobian of the square function at the matrix 1, 2, 3, 4. So if I perturb the matrix a little bit, like this E here, I should know how the matrix perturbs. And so all I should have to do is take this Jacobian, which is-- this is the same expression as here, and multiply that by the vec of my perturbation.

And of course, here's what we get as the answer. And we can check that by just doing a finite difference with the matrices. And you could see the first order, it just works.

So when I see this I'm happy. Other people trust mathematics. Me, I make too many errors. I have to double check it, just to see these numbers, this 0.0014, and the 0.0030, and the 0.0020, and the 0.0044 just matching up with the numbers above.

Again, this is just the finite differences. This is what the theory tells you. And I don't know if you guys need to check your work, but I love checking my work. Invariably, I lose factors of two or pi. So this is the best way to find them.

So rather than vecing and flattening, another way to basically express the Jacobian of a square, and you'll see why it works in a minute, is to express it as a linear transformation. And one of the things that if it didn't-- if you weren't-- if you weren't hit with this with your linear algebra class when you first took linear algebra, then this is a good time to recognize that being able to write down a linear transformation doesn't mean you have to write down a matrix. You could write down linear transformations without writing down a matrix.

For example, you think of M as a constant here. So M is going to be our position where we're taking derivative of whatever it is. It's a constant. And I just want to point out that this is a linear transformation of matrices.

So if M is a 2 by 2 matrix and E is a 2 by 2 matrix, this expression, without writing down the matrix, is a linear transformation. What do you have to check if it's a linear transformation? Well, let's see. If I multiply E by a number like 2 or pi, does the output get multiplied by the same number like 2 or pi?

Yep, the scalar would pull right out. And if I input E1 plus E2, would the output be the output of the 1 plus the output of the 2? You check those two things and that's what you need to check that the function is linear. You don't need a matrix.

Again, matrices are very concrete, but you don't need them when you're grown up. So this is a linear transformation on E and it's going to turn out to be exactly the right linear transformation for the very matrix we have at hand. And so with M being the matrix 1, 2, 3, 4 and E-- here, let's remind you what these matrices are if you've already forgotten.

But if M is the matrix 1, 2, 3, 4 and E is the perturbation matrix, the same numbers divided by 10,000, this linear transformation applied to E gives you the derivative without-- somehow, the information in here is the same as the information of the Jacobian, and yet I never form the Jacobian. That's a key, key point.

Never form the Jacobian, but the information to get the answer is embodied in this expression. So who needs to build a 4 by 4 matrix if you can just have this expression? Questions? Anything? Does it make sense? Doesn't make sense?

You want to see why that works? So I find the Kronecker product notation is a good way to be ready to set this thing up. So let me tell you about the Kronecker product notation and I'm going to set up these variables a, b, c, and d. And here's just a pair of matrices.

And what I want you to notice is that we have-- oh, am I sharing my screen? It just occurred to me that I'm not even sure I'm sharing my screen.

AUDIENCE: You're not sharing it on Zoom, though.

ALAN EDELMAN: Oh, yeah. Sorry about that. Just hit me that you don't see what I'm doing. All right, there we go.

And the reason I realized that is because I wanted to use this pen here anyway. So I want everybody to see the top half versus the bottom half of this matrix. So here what I'm doing is I'm taking the Kronecker product of this 2 by 1 matrix, Kronecker the-- see if I can do this.

It's hard to-- I'm using just the mouse. So p, q, r, and s. Is this coming out on the screen? Can you see the green?

So p, q, r, and s. And I just want you to see-- yes. Oops. It's not doing it now. there we go.

OK. So I want you all to see that it's taking the whole second matrix p, q, r, s, and multiplying it by a on top. And then this is taking the whole second matrix, pqrs, and multiplying by b on the bottom. And so what it does is, in effect. It takes-- you'll see that every possible product of something over here on the left and something over here on the right shows up, and it shows up in a very consistent order.

Once you see a couple of more examples, you'll see how the pattern goes. So this is the Kronecker product. So let's see. So here's an example where we take abcd with pqrs. And I'm sure you can all see that there's a pqrs times the a, pqrs now on the bottom left times the b. In the upper right, the pqrs is times the c and times the d.

So the pattern is-- one way to look at it is you take every element on the left and multiply it by the entire matrix on the right, and then put it in the same order. And so sometimes it's easier to see it with these pictures. And so here what we have is this 2 by 3 matrix, which is abcdef. So I can put it over here if you want to see it.

So I'm taking the Kronecker product of abcdef this way and I'm taking the Kronecker product with-- there's the pizza. What do we have? The alien. The pizza, the alien, the smiley cat, and what do we have here? The panda.

So Julia's perfectly comfortable with doing that sort of thing. And you could see that we have a times the pizza. Did I put the alien in the wrong order? Yep. The alien's over here.

But you get the idea. So you have a times pizza alien panda cat, b times pizza alien panda cat, c times pizza alien panda cat, d times those, e times those, f times those. I think you get the pattern.

You could do it in the other order. Let's see, what else did we do? Oh, here it is with the identity. So this one is pizza times 1, pizza times 0, pizza times 0, pizza times 1. Alien times 1, alien times 0, alien times 0, alien times-- yeah, I think you all get it. But it's still fun to watch.

Here it is in the other order. So this is 1 times pizza alien panda cat, 0 times pizza alien, 0 times the 2, and then again 1 times all of these funny symbols. OK, you get it.

Yeah. So the Kronecker product with the identity is a rather important special case. So here it is with just the pqrs. Here's the Kronecker product of the transpose with the identity, just putting everything in the other order. And what I want to demonstrate is that if you form the sum of these two Kronecker products as matrices and compare it with the Jacobian that you've already seen for the square function, you'll notice that you get exactly the same expression.

So one way to write the Jacobian without writing the Jacobian is to just write these Kronecker products. So you don't-- just because you write down the Kronecker product, it doesn't mean you actually have to multiply it all out, though. You could write it as an expression.

And what underlies everything, an identity that years ago, I remember I wrote this in big thick letters on an index card, and I placed it on my bookshelf above my desk to make sure that I would never forget it. So this is absolutely worth memorizing. That if you have the right compatibility so that you can multiply b times c times a transpose so the number of columns of has to be the number of rows of c, et cetera, that the Kronecker product of a and b, when multiplied by the vec of c, always is identically equal to the vec of bca transpose. So I would recommend memorizing that because it's just so handy to always know that.

And the mnemonic that I still have in my mind, I mean, you could just say to yourself, bca transpose. But the way I look at it, and if this works for you, that's fine, and if it doesn't work, you can have your own way. But the way I look at this is I see the b is closest to the c on the left. So I say bc.

And then I say, ah, the a, then, should go-- this is, again, a little memory thing. It doesn't mean anything. But the a should go around the other way. And since it went around, it deserves to be transposed. So if that works for you, that's fine. The b is just to the left of the c. The a has to come around so it gets transposed. v And that always works.

And so you can-- in fact, here is some random-- here you can see it with some random numbers, where-- and in fact, I could type this till the cows come home. It'll always come out to be a true identity. So here we can-- I could hit it again, but it'll always be true.

So you see it's bca transpose. OK, everybody convinced? So the Kronecker product of two 5 by 5 matrices would be 25 by 25, just to show you. And here are some useful identities worth knowing about.

So if you take the transpose of a Kronecker product, it's the Kronecker product of the transposes. And it's not in the other order. If you were thinking of ordinary matrix multiply, you'd have AB transpose. You know that would be B transpose A transpose.

But when you do the Kronecker product, the order remains the same. So don't get confused with matrix multiplication. Again, with inverse, inverse, if this wasn't a Kronecker product, if it was a matrix multiply product, again, you would put this in the other order. But with Kronecker product, it stays in that order.

Another useful thing to know is if a and are square matrices, if A's n by n and B is m by m, then the determinant is the determinant of a to the m. So a is the n by n1. It picks up the m from the b. And then B is the m by m1. It picks up the n from the a. So there's a formula.

The trace of a Kronecker product is the product of the traces. A Kronecker B is orthogonal, if A and B are both orthogonal. The matrix multiply of a Kronecker B and C Kronecker D is the matrix multiply of A and C with that of B and D. And yeah, so A Kronecker B and B Kronecker A have the same eigenvalues, which are actually the products of the eigenvalues of A and B separately, and the eigenvectors are just transposed of each other.

And so maybe this is worth writing out on the board just to see it, but this is how I like to do this. So if somebody tells me to talk about the derivative-- this is my favorite way of seeing the derivative of x squared.

So I'm going to go over here and I'm going to say, the derivative of x squared, which we've all known to-- we've seen this now a few times. I'm going to use this white chalk. The derivative of x squared, of course, is going to be the matrix x times dx plus dx x from the product rule. But I'm going to look at this and I'm going to say in my mind, this is xdxi. And this is idxx.

I normally wouldn't write that out, but just for purposes of showing you folks, I'll put the identity matrix in. And then I would say that-- which Kronecker product do I need? Well, let's see.

The thing that's on the left is going to be closest to the dx, so I'll put the x here. And then this I would be transposed. But since it's the identity, I don't have to write that out. And then here, you see the same game. The thing that's to the left of the x is the I. This thing then has to be x transpose.

And so one thing you could write is the vec-- but I don't really like this. But this would be reasonable to write. The vec of dx squared is this times the vec of x. And this is expressing-- it's just another notation for what's over there.

But what I really like to write is I like to think of this as a linear operator, and maybe I'll use the square brackets that Stephen likes to use. And I think I would-- the one I would enjoy the most to write, if I have to put everything on the-- you see, there's something nice about putting the dx on the right. So I think this is the notation that I would like the best.

This is equal to this, as an operator acting on dx. And by that, what I mean by that is if I ever have A Kronecker B acting on anything, I'll just call it M, I'm going to have that mean A M-- sorry, BMA transpose. So that would be my definition of a Kronecker bracket with a square input.

So if you want to have the dx all on one side, this is how I would do it. Yes?

AUDIENCE: Wouldn't it be vec dx in the third one, or the second one?

ALAN EDELMAN: vec of dx. Thank you. You are quite right. Good. Thanks for the catch. Perfect.

So that's basically what I'm writing here in this part of the notes over here. And again, this is-- I often don't even put the square bracket. So maybe it's not a bad idea to avoid confusion. So following Stephen's notation, it would be like-- this would be the way to do it, as this Kronecker operator acting on dx.

I find that there's never any real confusion. And so I don't write the vec, I don't put a square bracket. I just do this. I think I'm the only one who does this, so there's no reason why you should follow me.

But I find it quite nice to just write it this way and drop the vec. I think there's never any confusion. This is an operator acting on the dx in just the way we say. So you get to make your own choice. That's how I like to do it.

So just to show off a little bit in Julia, so we can get these Jacobians in Julia using the automatic differentiation. And we're going to talk about automatic differentiation. I would imagine-- I don't know if it's later this week, or-- when are we talking about add, Stephen. Do we know? Later this week or maybe next week?

STEPHEN: Yeah, I think probably by Friday or something like that.

ALAN EDELMAN: All right. Maybe by Friday we'll talk about automatic differentiation. So the key introduction, for those of you who've never seen AD, is that it's not symbolic like Mathematica and it's not numerical finite differences. It's none of those things.

It's something-- I mean, of course, it's mathematically equivalent, but it's-- when I first saw it, I was kind of surprised as to exactly how it works. So you might be surprised, too.

But right now, we're just going to be users. I'm just going to grab-- here's that j matrix. And I'm just going to point out that if-- it'll compute the Jacobian of the square function at the point M and we could check that it gives the right answer.

And in fact, it'll work with symbols, as well as-- so in some sense, here we're getting the Jacobian at every point because the elements are variables.

Let's do everything we just did, but let's move up to the matrix cube function. So here's the matrix cube function for 2 by 2 matrices written out in all of its glory. This is the thing that if I did it on the blackboard, just writing it out would take a little too long. But the computer doesn't mind, right? So there's the matrix cube function.

I mean, the computer wouldn't care if I did x to the fourth, I assume. Let's see. At what point does it get hairy? Maybe at the fifth? It's kind of-- well, you'd have to scroll, but there it is.

So let's do the cube. So there's the matrix cube function. And we can get the Jacobian. Again, it's just a matter of-- just to remind you that there's nothing-- I'm just letting the computer do the work, but you all could do it.

If I took the derivative of this with respect to p, you'd get this expression, the derivative with respect to q, you'd get rs plus 2pr You can all do this. Derivative with respect to s and the derivative with respect to t, and then you move down and do the same for this one, and then this one, then this one. But let the computer do the work. And so here's all the 16 derivatives right here in our faces.

And here-- let's see. So these are just two different ways to basically, I guess, write the same thing. Of course, we would like to use the product rule. So it would be dx times x squared plus xdx plus dxx squared. So Stephen showed you that, I think, the other day.

And this cannot be simplified, really. I mean, you could change notation, but you can't combine any of these terms. So of course, if it was a scalar-- if everything was a scalar, this would amount to 3x squared dx, which is the ordinary 1801 kind of answer. But for matrices, this is all you get.

And so if you want to write this as a function-- as a linear function where E is the dx and M is the x, you see that-- so here I guess I'm just using another notation where E and M-- oops. E and M for dx and x. So I'm just using these variables.

This is the linear function. And again, sometimes it takes a little getting used to. You might look in this and say, wait, this is not a linear function. I mean, I've got an M squared. Or M times E to-- this is not linear.

But why do I get to tell you that this is a linear function? You've got squaring of M. That's not linear. Squaring is not linear. What?

AUDIENCE: It's linear in E.

ALAN EDELMAN: It's linear in E, bingo. So this function is linear in E, and that's all that ever matters.

And I'll even remind you in ordinary freshman calculus that the derivative of x cubed, 3x squared, is not linear in x. The derivative actually is very nonlinear in x. It's quadratic in x. But what is linear is that if you make a perturbation E to the x, then the output is perturbed by 3x squared times E. And that is linear. So this is what's going on in matrix land.

Here's a numerical finite difference. Here is this function applied to E. And as usual, you see that to a reasonable number of digits, you get the same answer. So this is Julia's function, this linear function right here, applied to the perturbation E.

So E here is a dummy variable. It's just the dummy variable for the function. And then here, E is the numerical perturbation that I'm making. So E's having two roles here.

And we could check against the symbolic answer, and of course, I don't know how easy it is to see it. Maybe I'll just flatten up here. But you could see that these numbers that are about to go off screen here are exactly the same here. So everything checks out. It all works out. You get the right answer.

So how do we do this with the Kronecker notation? Well, I think-- here, let's just do this one on the blackboard as well, just to do it. So I'm just going to do the very same thing here, but I'll do it for x cubed. Just to get the practice.

So the x cubed version of what's here is going to be x squared dx plus xdxx plus dxx squared. And I'm going to look at this and say, aha, I Kronecker x squared, because x squared is multiplying dx on the left.

Here, I'm going to look at this and say, there's an x and an x transpose because x is on the left and x is on the right, but it comes around and it gets transposed. And this one, I'm going to say, is x transpose squared Kronecker I because it's on the right. And so whatever is on the right here goes on the left here and gets transposed.

And what I would write is this. But if you want to put in the square brackets to emphasize this is an operator, that's fine by me. Or you could, of course, put everything with vecs if you want to as well.

So your choices are you can write this. This is the operator notation. You could just write the same thing over here with a dx, which is that this is Stephen's notation. I guess this is Alan's notation. So you get to pick. And then finally most people's notation, which I think is horrible, is that the vec of the x cubed is equal to all this times the vec of the x. So this is the most standard notation.

So you get to take your pick. There's no one right answer. Yeah?

AUDIENCE: Maybe you have [INAUDIBLE] looked in detail, but it seems like further up in the presentation, you have dx times x squared plus x dx x plus dx squared instead of having the x squared out in [INAUDIBLE].

ALAN EDELMAN: Is it just-- did I write it in some other order? I mean, 1 plus 2 is 2 plus 1 kind of thing. Is that what we're talking about here?

Yeah, I've got the M squared on the-- yeah. If you look at this one-- oh yeah. Yeah, that doesn't matter, though. Yeah. Yeah. Or even up here. Yeah. But of course--

STEPHEN: So you have E times M times M twice. Is one of those supposed to be M times M times E?

ALAN EDELMAN: Oh, how did that happen? Oh, is that what you were noticing?

AUDIENCE: Yeah.

ALAN EDELMAN: Oh. That's not good news. How did I get the right answer?

Oh, I happened to pick an E that commuted with M, which is probably not a good check. Oh my gosh. That's exactly why I got the right answer. Don't ever take-- so I didn't really have a good general test. I see. So let's--

STEPHEN: Another one below, it looks like.

ALAN EDELMAN: Yeah, I probably copied and pasted it. You're right. Oh, good catch. Thank you. Why did I do that? So yes, we should fix the thing online, not just what I'm doing here.

So where did I define E? Or let's-- did E get defined earlier and earlier? There's a way in Pluto to-- if you go below the E and I think you hit Command or option-- oh. Command click to jump to the definition.

So let's see, if I command click, there's my-- OK. So this is where-- yeah. Because E commuted with M, I got away with murder. So let's do the following.

Let's not do that. Let's go round of 2 divided by 1-- 10,000 or something. So here's a perturbation. And I'll point out that with this one, I really did have the wrong answer, and thank you for catching that. So let's see. So where were before we had all the wrong answers? So here.

So if I did this-- so actually, it doesn't matter right here. But if I did it over here, so you see we're getting the right answer. Now but if I put back what I had at the beginning, you'd see I'd have the wrong answer. So it really did matter and I'm glad you found it.

Nobody found that last year. And I think I even showed this in another class. But I had a notebook that I showed for about four or five years until somebody found the mistake, so we're doing better. OK, good. Any other questions about-- all right.

So let's move on to something that might seem a little bit even more interesting, something that I like to think of as a matrix function, but however you want to think about it is fine, the LU decomposition. So let's think about this for a minute. Does everybody know what LU even means, just to make sure?

So the LU decomposition says that a matrix, which, by the way, it doesn't always exist, but it often exists. A matrix A can be written as a unit lower triangular, which means one's on the diagonal, times an upper triangular matrix, no assumptions about the diagonal.

And so this is called the LU decomposition of a matrix. And you've all seen it? You've all heard the word LU? It's equivalent to Gaussian elimination.

And one thing that's worth pointing out is that if a is an n by n matrix, so it has n squared elements, if you look at the amount of memory you need, the number of numbers that you have to write out L as a matrix, what you need is-- well, let's see. How many numbers are there sitting inside of a-- I mean, we don't care about the zeroes.

Numbers with information in them. How many numbers are in an upper triangular n by n? n times n plus 1 over 2, exactly.

And then you don't even have to store the ones. In fact, all the linear algebra routines like LAPACK, which is what it's underneath Julia, or Python, or MATLAB, or whatever you like to use. Everybody's using the same LAPACK software under the hood.

They don't even store the ones because you don't have to. I mean, why waste memory that you implicitly know it's a one. So what are the numbers? How many are there here that has information? How many numbers on the outside? Sorry?

AUDIENCE: Same one but a minus.

ALAN EDELMAN: Same, but?

AUDIENCE: But a minus.

ALAN EDELMAN: But a minus. Oh, I like the way you said that. The same, but a minus. OK, n times n minus 1 over 2. Good.

And if you add those two numbers up-- so if you-- I mean, you can do it visually, if you like. You could add them up. This one's below the ones. What do you get?

You can add it visually or just do it algebraically, whatever works for you. What do you get?

AUDIENCE: n squared.

ALAN EDELMAN: n squared. So I don't know if you ever stopped to think about this, but the LU decomposition is nothing other than a function from n squared variables to n squared variables. It's just a function from n squared variables to n squared variables.

And so therefore, it has a Jacobian. It has a Jacobian that, if you write it as a matrix, will be n squared by n squared. Just as much as the matrix square, or matrix cube, or the matrix inverse function has a Jacobian that's n squared by n squared, so does the LU decomposition.

And in fact, here's the 2 by 2 version. It gets more complicated when you go above 2 by 2. But here are L and U explicitly. So there's the-- so 2 by 2 matrix has four elements, and there's this one element in L and these three elements in U. And we could think of the LU decomposition as the function from the four elements pqrs to these four numbers, the 1 in L and the 3 in U.

So here's the LU decomposition in Julia, just checking that L times U is pqrs. And so here are the four entries, the L entry over here and then the three U entries I'm just writing out. And of course, Julia would happily take the Jacobian of that function and give you the 4 by 4 matrix Jacobian.

And so I hope everybody remembers how this works. I've said it a few times. You take the first one, q over p, and then take respect to p, that's minus q over p squared. p respect to q would be 1 over p. And then there's no r or s here, and you can-- and then you can do the same for the second, third, and fourth entry.

So in fact, I'm writing this as an exercise, but hopefully you can all see this, you can just do this with the product rule and get the right answer. So dLU is going to be dLU plus LdU. And did I write the answer here? No.

I'll have you guys shout it out. So let's see if you're getting the hang of this Kronecker product notation. And so the product rule says that dLU is going to equal LdU plus dLU. Or did I write it the other way? Not that it matters. I wrote it the other way. It doesn't matter, though. Plus dLU.

The thing that does matter is the L is always to the left of the U. It doesn't matter which term you write first, though. So anybody getting good at this? What Kronecker product is this? There's an I and an L or something. This one will have an I and a U. You're getting good at it yet?

This one's I Kronecker L. And then this one? What goes here? What goes here? U transpose Kronecker I. Exactly. You guys are good.

So you could write this like this. And that, I think, is a perfectly good way to write down the Jacobian. Of course, this is in matrix form and it's not in separate form, but it has all the right information. Yes?

AUDIENCE: [INAUDIBLE]?

ALAN EDELMAN: Yeah, that gets a little tricky. But yeah, that does get a little tricky. So I'm just going to-- I'm just going to say that it is-- so the underlying vector space is-- yeah, this is a little tricky.

So I like to think of the DL as 0 0 0 and then dL21. After all, the ones-- the derivative of the ones on the diagonal is zero and then everything else is zero anyway. And the dU is 0 dU11 dU12 and dU22.

So in some sense, if you really wanted to do it the standard notation way, the one that would be like this, you would have to write down-- you would have to write down something that worked with dL21 and dU11, dU12, and dU22.

But again, I feel like that's just following the crowd. Me, I would rather just say that I understand that my vector space consists of lower triangular perturbations to L and upper triangular perturbations to U. And so let's just look at it that way.

And that's how I like to see it. But maybe that's-- maybe I'm pushing the notation too far. Maybe you would prefer-- maybe for the purposes of this class, I would be better off doing it like this.

But even this is slightly-- yeah. Maybe that would be better for now. But you see, I like to put all my perturbations on the right. So again, nobody's fully standardized all this notation. So once again, you have your choice.

So here I'll show you just a couple of more matrix factorizations just quickly. I'm not going to go into too much detail, But I want you to just see that you can still do this for other situations. So one favorite of mine is a two-parameter example in a 2 by 2 matrices have four parameters, but there's only two real parameters here, which traceless symmetric matrices.

So the fact that the 2 by 2 matrix is symmetric means you have the s over here and over here. And the fact that it's traceless says that the trace is zero. So if you have a p over here, you need a minus p to make the sum of the diagonals zero. And so this is a two-parameter family.

And this is, by the way, the q that showed up on top. So if you write down-- you all know that symmetric matrices have orthogonal eigenvectors. So you can write it with cosines and sines. And as a little bit of setup, I'm using one half theta as opposed to just using theta. And here's my eigenvalues.

The eigenvalue matrix-- the only thing that we know about the eigenvalue matrix for starters is that the sum of the eigenvalues are zero. So let's call this the positive 1 and this is the negative 1. And what we have to equate is this is the eigendecomposition. So these two are equal, these two are opposite. And so it all becomes-- let's see if we can get this right here.

Yeah. So this is the matrix. And this is-- the matrix has two parameters, p and s. The eigendecomposition has two parameters. Maybe I should have emphasized that.

So the eigenvectors have one parameter, theta, and the eigenvalue has one parameter, r. So the eigendecomposition here, this little 2 by 2 eigendecomposition, is a map from the two variables ps to the two variables r theta.

And computing the eigenvalues of this 2 by 2 matrix is just a map from r2 to r2. And we can get the Jacobian. And here's the Jacobian of that map. So this is the actual 2 by 2 Jacobian of that. And sometimes, it's fun to compute the Jacobian determinant, which is r in this case.

And some of you might recognize that this eigendecomposition change of coordinates is exactly the same mathematical change of coordinates of going from Cartesian to polar coordinates. And this determinant, R, is exactly the-- many of you probably have memorized that the area element in Cartesian coordinates dx dy is r dr d theta. Well, here's-- we just, in effect, let Julia derive that for us.

But let's go further. I'll do one more example, the full 2 by 2 symmetric eigenproblem. But any questions about this restricted eigenproblem? It's a map from r2 to r2 and it's exactly the map-- it's exactly the map-- that's why I chose the theta over 2. It's exactly the map from Cartesian to polar coordinates by-- I don't know whether you call that a coincidence or what, but I think it's intriguing that it's the same map.

Let's do the full symmetric. So in the full case, a symmetric matrix is a three-dimensional vector space. So you've got three parameters, the two diagonals and then the off-diagonals are the same number. So we can refer to every 2 by 2 symmetric matrix with the variables p, r, and s.

And the three variables are-- since the symmetric matrix has an orthogonal eigenvectors, you still have one parameter, theta, for the eigenvectors, but we have two parameters, lambda 1 and lambda 2, for the eigenvalues. And so the symmetric eigenvalue problem is, once again, mapping this time from r3 to r3. You can go backwards and forwards.

You can go from the eigenvalues and eigenvectors to the matrix, which, by the way, is the more convenient direction, or you can go from the matrix entries to the eigenvalues and eigenvectors. So everybody in first-year linear algebra learns how to go this way. You set up the characteristic polynomial, you write down the quadratic, you solve it.

And in one way or another, you can get the eigenvector as well. So if you look at the null space of a minus lambda I or something. So this left direction, you've all done in your first-year linear algebra class.

Turns out it's a little more convenient to go the other way. This is just a matrix multiply where I set up the eigenvalues and eigenvectors and then I compute the matrix. And so here is the Jacobian calculation.

So it's a 3 by 3 Jacobian. There it is in all of its glory. A little bit painful to do by hand, but completely doable. Julia, of course, just happily gives you the answer without all the hard work.

And I'll just mention that if you calculate the determinant of this matrix, you get the volume element. So just to make sure everybody knows how that works, the volume of the Jacobian is how much a little cube scales. If I go back to that picture of the corgi being warped, I could look at his nose and then in the image, maybe it gets squashed.

The ratio is the determinant of the transformation. It's actually determinant of the Jacobian. And if you do that, you actually get lambda 1 minus lambda 2. And people like to say that symmetric-- the eigenvalues don't want to-- so if the determinant being lambda 1 minus lambda 2 is an indication of repulsion between eigenvalues, but maybe that's neither here nor there.

All right. Well, I think my timing is perfect. I went through this notebook and it's 12:01 exactly. So should we take a five minute break, and then Stephen, you'll be on board for the next set and stuff?

STEPHEN: That's right.

ALAN EDELMAN: All right, so 12:06 come on back. We'll start up again.