[SQUEAKING]

[RUSTLING]

[CLICKING]

ALAN EDELMAN: Today, the plan would be I'm going to talk a little bit-- I've got this notebook from last year talking about taking the derivative of such matrix quantities as eigenvalues. And so what's going on here is that we are-- the big story is orthogonal matrices may play a role if you have symmetric eigenvalue problems. And so we have constrained problems, which is something we haven't really talked about yet.

So we'll talk about differentiating under constraints a little bit. Then I will use the blackboard. And you might remember from over a week ago, I think it was, where I was talking about forward and reverse mode differentiation on the blackboard. And we made it through forward, but we never made it through backwards, or reverse mode. And so the plan is to kind of remember where we were and go back and do that.

And Steven had the idea of, the last 10 minutes-- at the end of class, we'll talk about the things that we haven't talked about. We've already figured out that, in principle, this class could go on for another week or two. But of course, IEP is coming to an end this week. But there's so much more that we could do.

So here, let me just kind go through this a little bit quickly. So ultimately, I want to think about orthogonal matrices, right? You all know what orthogonal matrices are. q transpose q is the identity. But to go slowly and as a bit of a warm up, let me just talk about differentiating on the unit sphere. OK?

So the unit sphere-- right? The sphere is like a baby version of an orthogonal matrix, right? It's one column of an orthogonal matrix. So let's talk about being on the sphere. OK?

And just as a reminder, if you are restricted--

[BARK]

Philly, what's going on? There's something about the sunbeams. He likes the-- oh, it's somebody's reflection. Is it mine? Or somebody is reflecting their phone. And Philip is going to chase it. [CHUCKLES] He-- he's really part cat somewhere genetically or something. Yeah, he will chase it. [CHUCKLING] OK. Thanks for-- all right. Where was I? Yes.

So you all know that if you move along on a circle, then tangents on the circle or [? quasar ?] are orthogonal, right? If you think of the vector of where you are from the center, right, tangents of course are orthogonal. And that's true in n dimensions-- that if you move around along a sphere in n dimensions, your position vector is from the origin. And then the velocity is always going to be orthogonal or tangent to the sphere, right?

And so all of a sudden now, we're talking about-- up until now, derivatives were free to be in any direction, right? All of our little dx's that we've seen-- or da's or whatever you want to call it-- have always been free to be in any direction you like. But if your point is restricted to a surface like the sphere, then somehow your dx is restricted to being tangent to the sphere now, right? It's not in any direction anymore, right?

We're no longer interested in radial directions because they don't even exist, right? We're only interested in tangent directions because they describe movements on the constrained surface, right? And so we need to figure out how to accommodate these sorts of tangential directions, right?

And so here, I'm kind of just showing you roughly what it would look like. By the way, is the sun drowning this out? Or is it OK? I'm wondering if this one should close-- or at least the little-- or partially close. It looks like there's a lot of glare from the sun. I'm happy to see the sun. Don't get me wrong. We'll take as much of it as we can get this. But a little-- yeah, that might be a little bit better.

Right. So just to kind of show you how you construct sort of the right kind of dq here, right-- by the way, do people know the trick for generating a random vector on the sphere? So if you-- so the question is, on the sphere, you would like to pick a point uniformly, right? "Uniformly" means that, no matter how you rotate the sphere, how you're picking it doesn't matter, right?

And the trick to doing that is to take x to be a standard normal. Like, here, I'm doing it in five dimensions, right? So x is a standard normal. And then all you have to do is normalize, which makes it a unit vector.

And properties of the normal distribution have it that it's invariant on the rotations. And so q is invariant on the sphere, right? And so that's the simplest way to generate a random vector of unit one that actually is uniform on the sphere. If you took Rand, which is-- you wouldn't get the right answer. It's very special to take Rand n, which is the normal distribution.

So q is on the sphere. And if you take a little dx-- you guys know me by now. I like to type 0.0001, something like that. If you take a random direction, of course, it won't be tangential to the sphere.

So what I'm doing is I'm normalizing it and subtracting it from q to kind of see-- so this is a random nearby point from x on the sphere. And I'm going to see its distance from q. And I will observe that q transpose dq is heading towards 0. OK?

So that's basically saying that-- I mean, I hope you all see the picture. But here, let's draw the picture. Maybe that'd be even better. So here's the picture, right? I've got my q over here. And I've got my q plus dq just a little bit off from here.

And so this is basically the dq. And what I'm showing you is-- numerically, the things that's easy to believe in your mind is that this is a right angle, right-- that q transpose dq is 0, right? So this right angle over here is the same thing as saying q transpose dq is 0.

Any questions about that? That's all pretty intuitive, right, that tangents are orthogonal to the radii no matter what the dimensions? OK?

And there's a math way of describing this. You don't just need your geometric intuition to see it. You just do the obvious thing. If you're restricted to the sphere, if x transpose x equals 1, then we could take our usual rules to differentiate, right? So if we differentiate, we have dx transpose x plus x transpose dx.

And Steven spent a lot of time in the first week reminding you that x transpose dx and dx transpose x are the same thing. They're just dot products. So we can combine them to get 2x transpose dx is the derivative of the constant 1, which is 0.

And so here, we get the-- this is a constraint on the perturbation, right? It says that dx is orthogonal to x. Those are the only dx's we're allowed. We're not allowed to go in any direction. Only those that are orthogonal to x are what we're going to consider if we want to restrict ourselves to a sphere. OK?

So that's working with constraints. And this is our first example. You can even do it-- here's just a little exercise just to kind of check. Suppose x is restricted to being on a circle, right? So x is a vector in two dimensions, but it's parameterized by theta over here. And so it's living on just a pure circle.

If you wanted to calculate x transpose dx as a kind of quick check just to show that it all-- I just like seeing that it works. I mean, again, there's nothing fancy here. But x is the vector cosine theta sine theta. dx just by taking simple, ordinary derivatives is minus sine theta d theta and cosine theta d theta. We can pull out the d theta. And you can see that this dot product is obviously minus cosine sine plus sine cosine. It's obviously 0. OK? So you--

And then there's a little bit of math gossip-- maybe just as a side story, if you'd like to hear a little bit of math gossip. So the x vector with its two coordinates is sometimes called an extrinsic vector, right? It's living in this plane, right?

The single theta coordinate, the one-dimensional-- the perimeter of a circle is a one-dimensional object, right? That's what's called intrinsic coordinates-- that, here, we are just parameter-- like, from the theta point of view, there's nothing but the circle, right?

x has two coordinates. And we have to constrain it. And mathematicians started out doing everything extrinsically. And then they discovered that intrinsic was better. And nowadays, many pure mathematicians will tell you you've got to do things intrinsically, it's so much better.

But I'll tell you, from computational viewpoints, I think extrinsic is actually better in the end. So I'm not convinced of what the pure mathematicians will tell you. So you get to decide. But whatever variables work is fine for you.

OK. So now what I want to do is move into matrix land. We want to grow up. This was simple circles. Let's go into matrix land. And wait, where are we going? What are we doing here? Let me just remember. Oh, right. OK.

So yeah. So we're taking another baby step into matrix land for the moment. All right. So I'm going to restrict myself to symmetric matrices-- not for any particular reason, but just for simplicity. And I'm going to look at essentially the gradient of the quadratic form x transpose ax, which we've seen enough times in this class. And as you know, the gradient is basically ax, right? And you've seen that before.

But now what I'd like to do is, ask what happens if you take the gradient of that very same quadratic form, but we restrict ourselves to being on the sphere, right? So what kind of change does that happen? So to do that, here's kind of the trick. We want to use the fact that the dx's we're allowing are orthogonal to x.

Those are the only dx's that we want to allow now. And one way to do that, one way to kind of say the same thing is that, if you use this projection matrix which projects orthogonal to x, then within this projection matrix times dx is still dx, right? So if you're-- right?

So here's the picture. I'll just do it with my hands. Here's your radius. Here's the plane perpendicular to the radius. And I want to project-- of course, the plane goes through the origin. I want to project onto that plane. And if I'm orthogonal, that's kind of a NOOP, right? If I'm in the radial direction, it's 0. And if I'm orthogonal to the radial direction, it doesn't change anything.

And so one way to describe the perturbations that are tangential is in this way. And so if you now calculate x transpose adx, I get to put this projection matrix in, right? And then I do a little bit of manipulation.

And then what you see is that the gradient is now the projection of ax, right? And so if you're moving around the sphere, this is the way to actually get the gradient. Because we want the gradient to live on that tangent space to be perpendicular to the radius.

So what's really going on could be said more simply perhaps, which is, in general, here you are on the sphere. And you want to figure out the best direction to go in to minimize that quadratic form. But you're only restricted to the sphere. So go ahead and project that gradient down to the sphere. That's really all that's happening. But I just thought I'd show it to you formally. OK?

So what did we just do? We needed two things. We needed a linearization of the function that's correct on tangents and a direction that itself is tangent. That's kind of what we just did to be able to do this sort of thing. OK? Here's a-- in general-- this wouldn't surprise anybody. But if you had some general function on the sphere, then, again, you would end up just projecting the original gradient so that it's in the tangent space. OK.

Let me get to the thing that's more fun. Let's go to real matrices. Let's differentiate orthogonal matrices. OK? That's the more fun part. And so this is where I want to go now. OK. So what am I doing here? So let's see.

So what I wanted this to look like is a complete analog of what I just did on the sphere, right? I just want to do it in this bigger orthogonal matrix land. So instead of taking a vector, I'm now going to get a 5-by-5 random matrix with standard normal entries and a perturbation that's also a 5-by-5 matrix with my 0, 0, 1's.

But what I'd like to do is-- I don't want to work with general matrices. I want to work with orthogonal matrices. So one quick way to get your hands on a random orthogonal matrix is to do the QR factorization of a and grab the q component.

AUDIENCE: That's not quite uniform.

ALAN EDELMAN: Not quite uniform-- yeah, I have written that down. I've written a whole thing about that. But if you randomize the signs, then it would be. Yeah, there's a whole story. I contacted the LAPACK people years ago about that-- yeah, not quite uniform among-- so there's this thing called Haar measure on the orthogonal group, which is the uniform measure that's invariant not quite because of certain reasons. But it doesn't matter for this purpose, right?

So q will be random, just not uniformly random, right? And then this dq will be a little bit of a change. OK? And what I want to do is look at the same kind of thing we looked at before. What is the relationship between dq and q? On the sphere, it was orthogonality. And if you look at-- I'm going to divide by roughly the size of my da there so that we can see.

And I wonder, if you don't know what q transpose dq is-- if you've known this before, I don't want you to say anything. But can anybody look at these numbers and guess what is going to be the differential version of being tangent to the-- what does it mean to be tangent to an orthogonal matrix, right? What's a small change to an orthogonal matrix?

So what do you see when you look at this 5-by-5 set of numbers? What do you think it's trying to point you towards? So it's obviously not the zero matrix. So-- yeah, what do you think you see?

AUDIENCE: Skew symmetric.

ALAN EDELMAN: Skew symmetric. That's exactly right. So the diagonal wants to be 0. That's all those 10 to the minus 6 and 1/7. And if you look, for example, at this entry versus this entry, you could see that they're the opposite sign with the same number, essentially.

So yes, this-- and so what this little numerical experiment is showing you is that, if you want to make a small change to a matrix q, if you make a small change dq, the constraint of being orthogonal at the differential level is that q transpose dq will be skew symmetric, or antisymmetric, right?

And of course, we could do this with-- here's the proof. The proof is almost the same proof as when you're on the sphere-- that, let's just differentiate the constraint. So this is the very definition of a matrix being orthogonal. q transpose q is the identity. That is the entire constraint of being orthogonal.

And so if you differentiate, you get q transpose dq plus dq transpose q equals 0. OK? But now you don't really get to combine these like we did on the sphere. OK? But what we can notice is that this is the same thing as saying that this plus its transpose is 0. OK? But that's just saying the very definition of being antisymmetric, or skew symmetric, is that a matrix plus its transpose is 0. And so therefore, q transpose is an antisymmetric differential. OK?

Now, if "antisymmetric differential" sounds like a weird thing to say, from my point of view, from a practical point of view, it's just like if you-- in the limit of making little perturbations, q transpose dq will be a little antisymmetric matrix, right? And then you don't have to worry about it being an antisymmetric differential. OK?

Let's talk about this. You all know that if you're in n dimensions and I ask you what is the dimensionality of the-- let's talk about the sphere first. If I have a sphere in n dimensions, what is the dimensionality of the sphere itself? Like, do you understand the question? Yeah. So yeah, the surface of the sphere.

And so in 3 dimensions, the surface is two, right? When you look at the world locally, it's kind of flat, right? I mean, if the world were a perfect sphere and you only could look at a small distance from where you are, right, it looks like a flat plane locally, right? I mean, it curves but from a small-- so in n dimensions, what's the dimensionality of a sphere?

AUDIENCE: n minus 1.

ALAN EDELMAN: n minus 1, right? There's only one constraint. And so that takes you down from n dimensions to n minus 1, right? The only constraint is that the sum of squares is 1. OK?

So let's now talk about-- generally speaking, n-by-n matrices live in what dimensional space totally? Just general n-by-n matrices for starters just to get our feet wet. n-squared, right? So every matrix is a point in n-squared dimensional space, right? Mathematicians have no trouble with high dimensions. We talk about it all the time. We don't get into philosophical discussions about time or anything, right? We just write down n-squared dimensional space. OK?

Now, the set of orthogonal matrices is some sort of blob in that space, right? It may be hard to imagine, but there's some constraints. And so now my question is, what is the dimension of that? Anybody want to attack that right now or want to-- maybe we should go slowly. Let's see.

So the 2-by-2 orthogonal matrices-- how many parameters are there? So let's just focus on rotations for starters. So we're talking about rotations and reflections. How many parameters for rotations?

So we're in four-dimensional space-- all possible matrices. And then we're looking at rotations only as some sort of blob in four-dimensional space. Dimensionality? One, right? Just the angle of rotation, right? So that's one. There's the cosine theta sine theta. Reflections also are kind of like that.

So for 2-by-2 matrices, it's one. For 3-by-3 matrices-- people who fly airplanes know the answer to this, believe it or not. Do you know how many parameters describe an orthogonal matrix in-- I mean, you don't need all nine numbers of a-- right? There's constraints. Do you want to guess?

AUDIENCE: 2.

ALAN EDELMAN: It's not 2. So it's not 6, but 6 is-- I feel like 6 is on the right track in a sort of backwards kind of way.

AUDIENCE: Is it 4?

ALAN EDELMAN: 4? We're running out of numbers. No, it's not 4.

AUDIENCE: Is it 3?

ALAN EDELMAN: 3. A good guesser here. Give this man a prize. So roll, pitch, and yaw are what the airplane people seem to know. That's what they've named it for the 3.

So the general answer-- anybody want to guess the general-- oh, yeah, anybody want to guess the general answer before I show you? You've got two data points. That may be not enough to guess with. It's one for 2-by-2. And it's 3 for 3-by-3. Anybody want to-- can you name that song in how many notes? All right. I'll just kind of tell you. So--

AUDIENCE: n-squared minus n choose 2.

ALAN EDELMAN: That sounds like the right backwards answer. So it's the complement of that. It's n choose 2. So somebody had said 6. You said 6. And the answer was 3. I feel like you're on the same wavelength. You said the 6 number when n equals 3.

And so there's a couple of ways to see this. But in general, the answer is n times n minus 1 over 2. One way to see it is to take the n-squared free parameters and count the constraints of q transpose q as the identity, right?

And the constraints-- as a picture, you could kind of count the constraints as the ones on the diagonal. The ones on the diagonals say that your sum of squares is 1, right? And then you could pick, say, the upper triangular because the lower is the same, right? This one tells you that the i'th column is-- right? This one says the i'th column has norm 1. This one says the i and jth column are orthogonal, right?

And so the total number is sort of an upper-- oh, wait. Yeah, the number of constraints is n times n plus 1 over 2. And so you have to subtract that from n-squared. And you get-- what's left is these numbers here, right? This is n times n minus 1 over 2-- or n choose 2, as this man was saying. OK? So that's one way to count it.

Another way to count it-- there are lots of ways to count it, actually. You could think about the QR factorization. And the R, which is upper triangular, has the same n times n plus 1 over 2. Leaving n times n minus 1 over 2 free parameters for the q is another way to say that.

You could also think about the symmetric eigenvalue problem, if you like, where a symmetric matrix has n times n plus 1 over 2. And the eigenvalues eat up n of them. And so now you have n times n minus 1 over 2.

So there are a lot of ways to-- you could also look at the SVD. That's another way to see it. For square matrices, this is n-squared. This one has n. And so these two orthogonal matrices each have n squared minus n over 2, right? So there's lots and lots of fun ways to check that this is the right answer. OK?

So now what I want to do is sort of talk about the subject of this class, which is to differentiate matrix functions, right? So this is what the buildup has been for the last 30 minutes or so. So I want to differentiate the symmetric eigenvalue problem. And I want to show you how to do it. OK?

So in fact, I guess we might somewhere derive this-- you had a name I had never heard before of something that I think is much older. But the physics-- the something Feynman theorem.

AUDIENCE: Hellmann-Feynman.

ALAN EDELMAN: Hellmann-Feynman. I'm sure it's much older than Hellman and Feynman. So that sounds like the-- was Hellmann a physicist?

AUDIENCE: I don't know who Hellmann was.

ALAN EDELMAN: OK. But we all know who Feynman was. All right. So the symmetric eigenvalue problem as, you all know, d you know that the eigenvalues of a symmetric matrix are real. And the eigenvectors can be put into an orthogonal matrix, right? And so the symmetric eigenvalue problem can be written as factoring a matrix s into q lambda q transpose. OK?

And so we can differentiate. And we get the three terms. This is just the product rule. And it turns out to be handy to kind of spin around the differential. It kind of makes you look like you're at a diagonal matrix. And if you spin it around-- I'm just putting q transpose on the left and q on the right. Let me move it up to make sure everybody can see this. But yeah. Let's move everything up. Yeah.

So this rotated version of the change to your symmetric matrix is q transpose dq lambda minus lambda q transpose dq plus d lambda. OK? So that's the derivative of-- so what this says is, if I perturb my symmetric matrix this much, then my eigenvectors will be perturbed by this much. And my eigenvalues will be perturbed this much.

And you know me by now. I don't believe any theorem unless I can check it numerically, right? So without a computer, I wouldn't be able to do math. So let's do that.

Let's create a random 5-by-5 matrix and also a random perturbation. But what I'd like to do is I want to symmetrize. So I'm going to call this s. And I'll also symmetrize my perturbation. And I'm just going to do the obvious. I'm going to take the eigendecomposition of s and the eigendecomposition of my perturbed s.

And so here's my eigenvalues. And here's the perturbed eigenvalues. So let's call dq to be the perturbed eigenvectors. d lambda is the perturbed eigenvalues. And let's just do a comparison of-- let's see. So how do I see the comparison? Did I stack them on top of each other? Yes.

I was stacking them on top of each other. I'm not sure why I did that, but I did. I stacked them on top of each other. So you should be able to see that, if I look at q transpose dq versus the math thing that I say it has to equal, I guess you could see that if you look at the top five rows-- what would happen if I didn't stack them on top of each other?

What if I just went comma? Would that not be easy? Maybe 5-by-5 is too big. Maybe that's why I did that. I just want to see why I did that. Oh, that's a bad-- and if I do this, that's good. That's better. All right Let's do it that way. OK. I think that's easier on the eye.

So yeah, you can see these two matrices are to enough digits the same. And you can believe the math right now, right? So I perturbed my symmetric matrix. And I look at this. And I have an identity that connects the perturbation of the eigenvalues and the perturbation of the eigenvectors.

And that Hellmann-Feynman theorem is just the diagonal part of this equation here. So the Hellmann-Feynman theorem would have said-- and I'll just write it in this notation. But the Hellmann-Feynman theorem is the diagonal, which says that essentially qi transpose dsqi-- so qi being the eigenvector-- is the diagonal.

So tell me. There's three terms there. There's three terms. What is the diagonal? So what's the diagonal of the first term? So capital lambda is a diagonal matrix, right? q transpose dq you tell me is an antisymmetric or a skew symmetric matrix. So what's the diagonal of that term there, q transpose dq lambda? Well, let me ask an easier question. What's the diagonal of q transpose dq? That's an easy question.

AUDIENCE: 0.

ALAN EDELMAN: It's 0, right? That's just the fact that any skew symmetric matrix has diagonal 0. OK? Now, if you take a-- what happens if you multiply it on the right by a diagonal matrix? What does that do to the diagonal elements? Right? So now I want to take that q transpose dq and multiply it by that capital lambda, which is diagonal. What does it do to the diagonals? So if I have an antisymmetric times diagonal, what's the diagonal of that?

AUDIENCE: 0.

ALAN EDELMAN: It's just 0, right? It doesn't change anything. So the only term that matters is the d lambda, right? And so basically, this is actually just saying that the change to the eigenvalue is exactly this. You just take the inner product of the change to your matrix in the direction of the eigenvector. And there you get it, right? So that's one quick derivation of that theorem.

STEVEN JOHNSON: Fun fact is that the Hellmann-Feynman theorem does not require the perturbation to be symmetric. So even if ps is not symmetric, you can-- which is really useful. Because then you can take a symmetric matrix. But then you look at a non-symmetric perturbation and you can still find the effect.

ALAN EDELMAN: That is a fun fact. Yeah, and of course, the same--

STEVEN JOHNSON: The two terms cancel. So you actually didn't need the diagonals to be 0.

ALAN EDELMAN: That's true.

STEVEN JOHNSON: In the Q transpose dQ Because the first term-- if you look at it, the diagonals cancel.

ALAN EDELMAN: That is true. Yeah, I can see that right there, that-- and of course, you could have taken a non-symmetric matrix. But then you would need both eigenvectors to play a role. So lots of generalizations. Yeah, I never noticed that. But I never perturbed symmetric matrices-- you perturb them?

STEVEN JOHNSON:

ALAN EDELMAN: I feel like it's violating the--

STEVEN JOHNSON: Physics often-- a symmetric system is like a lossless system, like a lossless vibrating system. But real systems, of course, often have losses. But the losses are small. So you want to analyze the losses--

ALAN EDELMAN: You want to analyze-- that makes sense.

STEVEN JOHNSON: Where it's easy. And then, you want to put in the losses perturbatively.

ALAN EDELMAN: That's cool, yeah.

STEVEN JOHNSON: and so it's nice.

ALAN EDELMAN: The other part of this actually tells you what happens to eigenvectors, right? That's living on the off diagonal. So here's the theorem. I just wrote it on the board because I forgot it was here. But yeah.

So if s depended on a parameter, one way to say it is that the derivative of the eigenvalue is this. And we can use this to, for example, get the gradient of a single eigenvalue because the-- oh, I wonder if I should mention-- is this obviously the same thing? No, it's not.

Let me-- there's another theorem that I think hasn't come up yet, but it's a very useful fact in linear algebra. And maybe, Steven, you could tell me if you did mention it. But I don't think--

STEVEN JOHNSON: The cyclic property of the trace?

ALAN EDELMAN: The cyclic property of the trace.

STEVEN JOHNSON: Yeah. So they're actually deriving this in homework. But yeah. We've used the cyclic property quite a bit.

ALAN EDELMAN: Oh, OK. All right. I didn't remember that you had done it. But all right. Then I don't have to say a word about it then. And the fact that the cyclic property kind of holds-- there's another one of these things where, when it's a scalar, you might be sort of a little bit surprised that it works, right? Because you don't have to write the word "trace."

When you have a scalar, it's like a 1-by-1 matrix, right? You don't have to say the word "trace" because it's just that. But if you pretend the word "trace" is here, then you can move it around. And now it's a big matrix, right? Was that another thing? I mean, you pointed out that it's kind of like dot products commuting--

STEVEN JOHNSON: We've used that a couple of times.

ALAN EDELMAN: OK.

STEVEN JOHNSON: Yeah, this just kind of gives away one of the homework problems. But that's OK.

ALAN EDELMAN: Oh, I see.

STEVEN JOHNSON: But hey probably already looked at the homework already. So.

ALAN EDELMAN: All right. Well, it doesn't hurt if it helps, I think.

STEVEN JOHNSON: So physicists call this first order perturbation theory which is the same thing as-- a first order perturbation--

ALAN EDELMAN: Is the same thing as computing the derivative.

STEVEN JOHNSON:

ALAN EDELMAN: Right, yeah.

STEVEN JOHNSON:

ALAN EDELMAN: It's just physics speak for the same thing. OK. And we can also get information about the eigenvectors as well.

So off the diagonal, you can write down this equation. And it basically says that this divided by this will give you the change of the eigenvectors. And all sorts of havoc happens if two eigenvalues are equal. And thing is not exactly differentiable anymore and kind of revealing itself. OK?

And so I think-- let's see. So-- oh, and by the way, here's second order perturbation theory, while we're at it. So yeah. So one can follow all the way through. Maybe I won't make a big deal about it right now.

But if you know the first order perturbation of the eigenvectors, then you can use that to get the second order perturbation of the eigenvalues. And here's your whole perturbation theory. Should I have have divided by 2? Oh, no, there is a 2. So this is correct. OK.

All right. So I just wanted to kind of show you this far for the derivatives of eigenvalues. You could do essentially the same game and get derivatives of singular values, all sorts of other things. But I think this is kind of just enough of a peak for the time we have in this class.

Any questions about all this? Anything you had always wondered about differentiating eigenvalues, or eigenvectors, and stuff like that? OK then. All right. Well, then I'm going to switch gears. And-- oh, yes. There is a question-- two questions, actually.

AUDIENCE: What if-- for example, sometimes I want to take the derivative confined to something like the the positive semidefinite matrices.

ALAN EDELMAN: So that--

AUDIENCE: Do I have to also check, am I inside the thing, so I can go anywhere? And then I have to on the boundary so that then I can do this?

ALAN EDELMAN: I think so. That's a good question. So this is now a-- it's kind of like an inequality constraint. And I was kind of talking about an equality constraint. So yeah. I don't know any other way.

STEVEN JOHNSON: Well, you could use intrinsic coordinates because--

ALAN EDELMAN: I'm not a fan.

STEVEN JOHNSON: But you could write semidefinite-- any--

ALAN EDELMAN: Use Cholesky factor, yeah.

STEVEN JOHNSON: well, just as b transpose b.

ALAN EDELMAN: Right.

STEVEN JOHNSON: And so if you parameterize it in terms of some kind of factor-- it doesn't have to be a Cholesky factor, but you can--

ALAN EDELMAN: That's true. It could be the square root, right, or any--

STEVEN JOHNSON: It's probably the easiest way.

ALAN EDELMAN: Maybe for that case that's probably true.

STEVEN JOHNSON:

ALAN EDELMAN: But in general--

STEVEN JOHNSON: Yeah. Well, yeah. In general, I would say that semidefinite matrices don't fall down from the sky. So they usually-- if you stare at it hard enough, they came from a b transpose b. So--

ALAN EDELMAN: So maybe the b is handy. But that's actually a good point.

STEVEN JOHNSON:

ALAN EDELMAN: That is a good point. OK. I don't know how general that is, though-- that it's easy to find an intrinsic coordinate system.

STEVEN JOHNSON: Yeah.

ALAN EDELMAN: So that might be the tricky part. But that is a good answer. OK. And I think there was another question over there? OK.

AUDIENCE: I had a question where you take x transpose A x and the constraint x transpose x is 1. So if you just like the df assuming that for x transpose x, you still get the dot product of Ax and x and dx, right?

ALAN EDELMAN: Right. But we wouldn't be going-- like, our search direction is kind of not physical, if you will. It's not on the sphere anymore.

AUDIENCE: Yeah. That's why you have to multiply the projection.

ALAN EDELMAN: Right. So to put the-- one way to look at it is the projection matrix is just, like, slapping it back onto the sphere, like whack-a-mole right down back onto the tangent space of the sphere. But we also see that it gives the correct answer and that it's moving in the steepest descent direction on the sphere. Because we don't want any directions that go off the sphere at this point.

So that's right. Was there another question? Did I-- was there another hand up or just those two? All right.

STEVEN JOHNSON: Do you want the blackboard now?

ALAN EDELMAN: Yeah, I'm going to do the blackboard now. And we're going to switch gears. Thanks, Steven. And I'm going to pull back these notes here. So I guess I can close this.