PROFESSOR: Well, OK. So first important things about the course, plans for the course. And then today I'm going to move to the next section of the notes, section 2, or part 2, I should say. And actually I'll skip for the moments section 2-1 and go to section 2-2, and all of chapter 2 will come to you probably today or latest tomorrow. So that's where we're going next.

I'm following the notes pretty carefully, except I'm going to skip the section on tensors until I learn more basically. Yeah. Yeah. I could say a little about tensors, but this flows naturally using the SVD. So it's just a terribly important problem, least squares. And of course, I know that you've seen one or two ways to do least squares. And really the whole subject comes together.

Here I want to say something, before I send out a plan for looking ahead for the course as a whole. So there's no final exam. And I don't really see how to examine you, how to give tests. I could, of course, create our tests about the linear algebra part. But I don't think it's-- it's not sort of the style of this course to expect you quickly to create a proof for something in class.

So I think, and especially looking at what we're headed for, and moving quite steadily in that direction, is all the problems that this linear algebra is is aimed at, right up to and including conjugate gradient descent and deep learning, the overwhelmingly important and lively, active research area. I couldn't do better than to keep the course going in that direction.

So I think what I would ask you to do is late in sort of April, May, the regular homeworks I'll discontinue at a certain point. And then instead, I'll be asking and encouraging a project-- I don't know if that's the right word to be using-- in which you use what we've done. And I'll send out a message on Stellar listing five or six areas and only-- I mean, one of them is the machine learning, deep learning part. But they're all the other parts, things we are learning how to do. How to find sparse solutions, for example, or something about the pseudo inverse. All kinds of things.

So that's my goal, is to give you something to do which uses the material that you've learned. And look, I'm not expecting a thesis. But it's a good chance. So it will be more than just, drag in some code for deep learning and some data matrix and do it. But we'll talk more as the time comes.

So I just thought I'd say, before sending out the announcement, I would say it's coming about what as a larger scale than single one week homeworks would be here before. Any thoughts about that? I haven't given you details. So let me do that with a message, and then ask again. But I'm open to-- I hope you've understood-- I think you have-- that if you make suggestions, either directly to my email or on Piazza or whatever, they get paid attention to.

OK. Shall I just go forward with least squares? So what's the least squares problem, and what are these four ways, each bringing-- so let me speak about the pseudo inverse first. OK, the pseudo inverse of a matrix. All right. Good.

So we have a matrix A, m by n. And the pseudo inverse I'm going to call A plus. And it naturally is going to be n by m. I'm going to multiply those together. And I'm going to get as near to the identity as I can. That's the idea, of course, of the pseudo inverse, The word pseudo is in there, so no one's deceived. It's not an actual inverse.

Oh, if the matrix is square and has an inverse, of course. Then if A inverse exists, which requires-- everybody remembers it requires the matrix to be square, because I mean inverse on both sides. And it requires rank n, full rank. Then the inverse will exist. You can check it. MATLAB would check it by computing the pivots in elimination and finding n pivots.

So if A inverse exists, which means A times A inverse, and A inverse times A, both give I, then A plus is A inverse, of course. The pseudo inverse is the inverse when there is one. But I'm thinking about cases where either the matrix is rectangular, or it has zero eigenvalues. It could be square, but it has a null space, other than just the 0 vector. In other words, the columns are dependent.

What can we do then about inverting it? We can't literally invert it. If A has a null space, then when I multiply by a vector x in that null space, Ax will be 0. And when I multiply by A inverse, still 0. That can't change the 0. So if there is an x in the null space, then this can't happen. So we just do the best we can. And that's what this pseudo inverse is.

And so let me draw a picture of the picture you know of the row space and the null space. OK,

and it's there, you see. There is a null space. And over here I have the column space and the null space of A transpose. OK. So this is the row space, of course. That's the column space of A transpose, and there is the column space of A. OK.

So which part of that picture is invertible, and which part of the picture is hopeless? The top part is invertible. This is the r-dimensional row space, r-dimensional column space. A takes a vector in here, zaps it into every-- you always end up in the column space. Here I take a vector in the row space-- say, x-- and it gets mapped to Ax.

And between those two spaces, A is entirely invertible. You get separate vectors here, go to separate vectors in the column space, and the inverse just brings it back. So we know what the pseudo inverse should do. It will take A will go that way, and A plus, the pseudo inverse will be just-- on the top half of the picture, it'll give us A plus. We'll take Ax back to x in the top half.

Now, what about here? That's where we have trouble, when we don't have-- that's what spoils our inverse. If there is a null space vector, then it goes where? When you multiply by A, this guy in the null space goes to 0. Usually along a straighter line than I've drawn. But it goes there. It gets to 0.

So you can't raise it from the dead, so to speak. You can't recover it when there's no A inverse. So we have to think, what shall A inverse do to this space here, where nobody's hitting it? So this would be the null space of A transpose. Because A-- sorry-- yeah, what should the pseudo inverse do? I said what should the inverse do? The inverse is helpless.

But we have to define A plus. I've said what it should do on that guy, on the column space. It should take everything in the column space back where it came from. But what should it do on this orthogonal space, where-- yeah, just tell me, what do you think? If I have some vector here-- let's call it V r plus 1. That would be like-- so here I have a nice basis for the column space. I would use V's for the ones that come up in the SVD. They're orthogonal, and they come from orthogonal U's.

So the top half is great. What shall I do with this stuff? I'm going to send that back by A plus. And what am I going to do with it? Send it to-- nowhere else could it go. 0 is the right answer. All this stuff goes back to 0.

I'm looking for a linear operator, a matrix. And I have to think, once I've decided what to do with all those and what to do with all these, then I know what to do with any combination. So

I've got it. I've got it. So the idea will be, this is true for x in the row space. For x in the row space, if x is in the row space, Ax is in the column space, and A inverse just brings it back as it should.

And in the case of an invertible matrix A, what happens to my picture? What is this picture looking like if A is actually a 6 by 6 invertible matrix? In that case, what's in my picture and what is not in my picture? All this null space stuff isn't there. And null space is just a 0 vector. But all that I don't have to worry about. But in general, I do have to say.

So the point is that A plus on the-- what am I calling this? It's the null space of A transpose, or whatever on V r plus 1 to Vn, all those vectors, the guys that are not orthogonal to the column space. Then we have to say, what does A plus do to them? And the answer is, it takes them all to 0.

So there is a picture using what I call the big picture of linear algebra, the four spaces. You see what A plus should do. Now, I need a little formula for it. I've got the plan for what it should be, and it's sort of the natural thing. So A plus A is, you could say it's a projection matrix. It's not the identity matrix because if x is in the null space, A plus A will take it to 0. So it's a projection.

A plus A is the identity on the top half, and 0 on the bottom half. That's really what the matrix is. And now, I want a simple formula for it. And I guess my message here is, that if we're looking for a nice expression, start with the SVD. Because the SVD works for any matrix. And it writes it as an orthogonal matrix times a diagonal matrix times an orthogonal matrix.

And now I want to invert it. Well, suppose A had an inverse. What would that be? This is if invertible, what would be the SVD of A inverse? What would be the singular value decomposition, if this is good? So when is this going to be good? What would I have to know about that matrix sigma, that diagonal matrix in the middle, if this is truly an invertible matrix?

Well, no. What's its name? Those are not eigenvalues. Well, they're eigenvalues of A transpose A. But they're singular values. Singular value, that's fine. So that's the singular value matrix. And what would be the situation if A had an inverse? There would be no 0's. All the singular values would be sitting there, sigma 1 to sigma n.

What would be the shape of this sigma matrix? If I have an inverse, then it's got to be square n by n. So what's the shape of the sigma guy? Also square, n by n. So the invertible case would

be-- and I'm going to erase this in a minute-- the invertbile case would be when sigma is just that. That would be the invertible case.

So let's see. Can you finish this formula? What would be the SVD of A inverse? So I'm given the SVD of A. I'm given the U and the sigma is cool and the V transpose. What's the inverse of that? Yeah, I'm just really asking what's the inverse of that product of three matrices.

What comes first here? V. The inverse of V transpose is V. That's because V is a orthogonal matrix. The inverse of sigma, just 1 over it, is just the sigma inverse. It's obvious what that means. And the inverse of U would go here. And that is U transpose. Great. OK.

So this is if invertible. If invertible, we know what the SVD of A inverse is. It just takes the V's back to the U's, or the U's back to the V's, whichever. OK. OK. Now we've got to do it, if we're going to allow-- if we're going to get beyond this limit, this situation, allow the matrix sigma to be rectangular. Then let me just show you the idea here.

So now I'm going to say, now sigma, in general, it's rectangular. It's got r non 0's on the diagonal, but then it quits. So it's got a bunch of 0's that make it not invertible. But let's do our best and pseudo invert it. OK. So now help me get started on a formula for using-- I want to write this A plus, which I described up there, in terms of the subspaces. Now I'm going to describe A plus in terms of U, sigma, and V, the SVD guys.

OK. So what shall I start with here? Well, let me give a hint. That was a great start. My V is still an orthogonal matrix. V transpose is still an orthogonal matrix. I'll invert it. At the end, the U was no problem. All the problems are in sigma. And sigma, remember, sigma-- so it's rectangular. Maybe I'll make it wide, two wide. And maybe I'll only give it two non-zeros, and then all the rest. So the rank of my matrix A is 2, but the m and n are bigger than 2. It's just got two independent columns, and then it's just sort of totally singular.

OK. So my question is, what am I going to put there? And I've described it one way, but now I'm going to describe it another way. Well, let me just say, what I'll put there is the pseudo inverse of sigma. I can't put sigma inverse using that symbol, because there is no such thing. With this, I can't invert it. So that's the best I can do.

So I'm almost done, but to finish, I have to tell you, what is this thing? So sigma plus. I'm now going to tell you sigma plus. And then that's what should sit there in the middle. So if sigma is this diagonal matrix which quits after two sigmas, what should sigma plus be? Well, first of all,

it should be rectangular the other way. If this was m by n column, n columns and m rows, now I want to have n rows and m columns.

And yeah, here's the question. What's the best inverse you could come up with for that sigma? I mean, if somebody independent of 18.065, if somebody asks you, do your best to invert that matrix, I think we'd all agree it is, yeah. One over the sigma 1 would come there. And 1 over sigma 2, the non zeros.

And then? Zeros. Just the way up there, when we didn't know what to do, when there was nothing good to do. Zero was the right answer. So this is all zeros. Of course, it's rectangular the other way. But do you see that if I multiply sigma plus times sigma, if I multiply the pseudo inverse times the matrix, what do I get if I multiply that by that? What does that multiplication produce? Can you describe the-- well, or when you tell me what it looks like, I'll write it down.

So what is sigma plus times sigma? If sigma is a diagonal, sigma plus is a diagonal, and they both quit after two guys. What do I have? One? Because sigma 1 times 1 over sigma 1 is a 1. And the other next guy is a 1. And the rest are all zeros. That's right. That's the best I could do. The rank was only two, so I couldn't get anywhere. So that tells you what sigma plus is.

OK. So I described the pseudo inverse then with a picture of spaces, and then with a formula of matrices. And now I want to use it in least squares. So now I'm going to say what is the least squares problem. And the first way to solve it will be to involve-- A plus will give the solution.

OK. So what is the least squares problem? Let me put it here. OK, the least squares problem is simply, you have an equation, Ax equals b. But A is not invertible. So you can't solve it. Of course, for which-- yeah, you could solve it for certain b's. If b is in the column space of A, then just by the meaning of column space, this has a solution. The vectors in the column space are the guys that you can get.

But the vectors in the orthogonal space you cannot get. All the rest of the vectors you cannot get. So suppose this is like so, but always A is m by n rank r. And then we get A inverse when m equals n equals r. That's the invertible case. OK.

What do we do with a system of equations when we can't solve it? This is probably the main application in 18.06. So you've seen this problem before. What do we do if Ax equal b has no solution? So typically, b would be a vector of measurements, like we're tracking a satellite, and we get some measurements. But often we get too many measurements. And of course,

there's a little noise in them. And a little noise means that we can't solve the equations.

That may be the case everybody knows is, where this equation is like expressing a straight line going through the data points. So the famous example of least squares is fit a straight line to the b's, to b1, b2. We've got m measurements. We've got m measurements. The physics or the mechanics of the problem is pretty well linear. But of course, there's noise.

And a straight line only has two degrees of freedom. So we're going to have only two columns in our matrix. A will be only two columns, with many rows. Highly rectangular. So fit a straight line. Let me call that line Cx plus D. Say this is the x direction. This is the b's direction. And we've got a whole bunch of data points. And they're not on a line.

Or they are on the line. Suppose those did lie on a line. What would that tell me about Ax equal b? I haven't said everything I need to, but maybe the insight is what I'm after here. If my points are right on the line-- so there is a straight line through them-- the unknowns here-- so let me-- so Ax-- the unknowns here are C and D. And the right hand side is all my measurements. OK.

Suppose-- without my drawing a picture-- suppose these points are on the line. Here's the different x's, the measurement times. Here is the different measurements. But if they're on a line, what does that tell me about my linear system, Ax equal b? It has a solution. Being on a line means everything's perfect. There is a solution.

But will there usually be a solution? Certainly not. If I have only two parameters, two unknowns, two columns here, the rank is going to be two. And here I'm trying to hit any noisy set of measurements. So of course, in general the picture will look like that. And I'm going to look for the best C and D. So I'll call it Cx plus D. Yeah, right. Sorry. That's my line. So those are my equations.

Sorry, I often write it C plus dx. Do you mind if I put the constant term first in the highly difficult equation here for a straight line? So let me tell you what I'm-- so these are the points where you have a measurement-- x1, x2, up to xn. And these are the actual measurements, b1 up to bm, let's say .

And then my equations are-- I just want to set up a matrix here. I just want to set up the matrix. So I want C to get multiplied by ones every time. And I want D to get multiplied by these x's-- x1, x2, x3, to xm, the measurement places. And those are the measurements. Anyway.

And my problem is, this has no solution. So what do I do when there's no solution? Well, I'll do what Gauss did. He was a good mathematician, so I'll follow his advice. And I won't do it all semester, as you know. But Gauss's advice was, minimize-- I'll blame it on Gauss-- the distance between Ax and b squared, the L2 norm squared, which is just Ax minus b transpose Ax minus b. It's a quadratic. And minimizing it gives me a system of linear equations.

So in the end, they will have a solution. So that's the whole point of least squares. We have an unsolvable problem, not no solution. We follow Gauss's advice to get the best we can. And that does produce an answer.

So this is-- if I multiply this out, it's x transpose, A transpose, Ax. That comes from the squared term. And then I have probably these-- actually, probably I'll get two of those, and then a constant term that has derivative 0 so it doesn't enter. So this is what I'm minimizing. This is the loss function.

And it leads to-- let's just jump to the key here. What equation do I get when I look for-- what equation is solved by the best x, the best x? The best x solves the famous-- this is regression in statistics, linear regression. It's one of the main computations in statistics, not of course just for straight line fits, but for any system Ax equal b.

That will lead to-- this minimum will lead to a system of equations that I'm going to put a box around, because it's so fundamental. And are you willing to tell me what that equation is? Yes, thanks.

AUDIENCE: A transpose A.

PROFESSOR: A transpose A is going to come from there-- you see it-- times the best x equals A transpose b. That gives the minimum. Let me forego checking that. You see that the quadratic term has the matrix in it. So it's derivative. Maybe the derivative of this is 2 A transpose Ax, and then the 2 cancels that 2. And this could also be written as x transpose A transpose b. So it's x transpose against A transpose b. That's linear. So when I take the derivative, it's that constant.

That's pretty fast. 18.06 would patiently derive that. But here, let me give you the picture that goes with it, the geometry. So we have the problem. No solution. We have Gauss's best answer. Minimize the 2 norm of the error. We have the conclusion, the matrix that we get in. And now I want to draw a picture that goes with it. OK.

So here is a picture. I want to have a column space of A there in that picture. Of course, the 0 vector's in the column space of A. So this is all possible vectors Ax. Right? You're never forgetting that the column space is all the Ax's.

Now, I've got to put b in the picture. So where does this vector b-- so I'm trying to solve Ax equal b, but failing. So if I draw b in this picture, how do I draw b? Where do I put it? Shall I put it in the column space? No. The whole point is, it's not in the column space. It's not an Ax. It's out there somewhere, b. OK. And then what's the geometry that goes with least squares and the normal equations and Gauss's suggestion to minimize the error? Where will Ax be, the best Ax that I can do?

So what Gauss has produced is an A here. You can't find an x. He'll do as best he can. And we're calling that guy x hat. And this is the algebra to find x hat. And now, where is the picture here? Where is this vector Ax hat, which is the best Ax we can get?

So it has to be in the column space, because it's A times something. And where is it in the column space? It's the projection. That's Ax hat. And here is the error, which you couldn't do anything about, b minus Ax hat. Yeah. So it's the projection, right.

So all this is justifying the-- so we're in the second approach to least squares, solve the normal equations. Solve the normal equations. That would be the second approach to least squares. And most examples, if they're not very big or very difficult, you just create the matrix A transpose A, and you call MATLAB and solve that linear system. You create the matrix, you create the right hand side, and you solve it.

So that's the ordinary run of the mill least squares problem. Just do it. So that's method two, just do it. What's method three? For the same-- we're talking about the same problem here, but now I'm thinking it may be a little more difficult. This matrix A transpose A might be nearly singular.

Gauss is assuming that-- yeah, when did this work? When did this work? And it will continue to work in the next three-- this works, this is good, if assuming A has independent columns. Yeah, better just make clear. I'm claiming that when A has-- so what's the reasoning?

If A has independent columns-- but maybe not enough columns, like here-- it's only got two columns. It's obviously not going to be able to match any right hand side. But it's got independent columns. When A has independent columns, then what can I say about this

matrix? It's invertible. Gauss's plan works. If A has independent columns, then this would be a linear algebra step. Then this will be invertible. You see the importance of that step.

If A has independent columns, that means it has no null space. Only x equals 0 is in the null space. Two independent columns, but only two. So not enough to solve systems, but independent. Then you're OK. This matrix is invertible. You can do what Gauss tells you.

But we're prepared now-- we have to think, OK. So what do I really want to do? I want to connect this Gauss's solution to the pseudo inverse. Because I'm claiming they both give the same result. The pseudo inverse will apply. But we have something-- A is not invertible. Just keep remembering this matrix. It's not invertible. But it has got independent columns.

What am I saying there? Just going back to the picture. If A is a matrix with independent columns, what space disappears in this picture? The null space goes away. So the picture is simpler. But it's still the null space of A transpose. This is still pretty big, because I only had two columns and a whole lot of rows. And that's going to be reflected here.

So what am I trying to say? I'm trying to say that this answer is the same as the pseudo inverse answer. We could possibly even check that point. Let me write it down first. I claim that the answer A plus b is the same as the answer coming from here, A transpose A, inverse A transpose b, when I guess the null space is 0, the rank is all of n, whatever you like to say.

I believe that method one, this two within one quick formula-- so you remember that this was V sigma plus U transpose, right? That's what A transpose was. That this should agree with this. I believe those are the same when the null space isn't in the picture.

So the fact that the null space is just a 0 vector means that this inverse does exist. So this inverse exists. But A A transpose is not invertible. Right? No inverse. Because A A transpose would be coming-- all this is the null space of A transpose. So A transpose is not invertible.

But A transpose A is invertible. How would you check that? You see what I'm-- it's taken pretty much the whole hour to get a picture of the geometry of the pseudo inverse. So this is the pseudo inverse. And this is-- that matrix there, it's really doing its best to be the inverse. In fact, everybody here is just doing their best to be the inverse.

Now, how well is this-- how close is that to being the inverse? Can I just ask you about that, and then I'll make this connection, and then we're out of time. How close is that to being the inverse of A? Suppose I multiply that by A. What do I get?

So just notice. If I multiply that by A, what do I get? I get, yeah? I get I. Terrific. But don't be deceived to thinking that this is the inverse of A. It worked on the left side, but it's not going to be good on the right hand side. So if I multiply A by this guy in that direction, I'll get as close to the identity as I can come, but I won't get the identity that way.

So this is just a little box to say-- so what's the point I'm making? I'm claiming that this is the pseudo inverse. Whatever. Whatever these spaces. The rank could be tiny, just one. This works when the rank is n. I needed independent columns.

So when the rank is n-- so this is rank equal n. That Gauss worked. Then I can get a-- then it's a one-sided inverse, but it's not a two-sided inverse. I can't do it. Look, my matrix there. I could find a one-sided inverse to get the 2 by 2 identity. But I could never multiply that by some matrix and get the n by n identity out of those two pathetic columns. OK.

Maybe you feel like just checking this. Just takes patience. What do I mean by checking it? I mean stick in the pseudo SVD. Just put it in the SVD and cancel like crazy. And I think that'll pop out. Do you believe me? Because it's going to be a little painful. 3 U sigma V transpose, all transposed, and then something there and something there. I've got nine matrices multiplying away. But it's going to-- all sorts of things will produce the identity. And in the end, that's what I'll have.

So this is a one-sided true inverse, where the SVD-- this fit formula is prepared to have neither side invertible. It's still-- we know what sigma plus means. Anyway. So under the assumption of independent columns, Gauss works and gives the same answer as the pseudo inverse.

OK. Three minutes. That's hardly time, but this being MIT, I feel I should use it. Oh my god.

Number three. So what's number three about? Number three has the same requirement as number two, the same requirement of no null space. But it says, if I could get orthogonal columns first, then this problem would be easy. So everybody knows that Gram-Schmidt is a way-- boring way-- to get from these two columns to get two orthogonal columns.

Actually, the whole idea of Gram-Schmidt is already there for 2 by 2. So I have two minutes, and we can do it. Let's do Gram-Schmidt on these two columns-- I don't want to use U and V-- column y and z. OK.

Suppose I want to orthogonalize those guys. What's the Gram-Schmidt idea? I take y. It's

perfectly good. No problem with y. There is the y vector, the all 1's. Then this guy is not orthogonal probably to that. It'll go off in this direction, with an angle that's not 90 degrees.

So what do I do? I want to get orthogonal vectors. I'm OK with this first guy, but the second guy isn't orthogonal to the first. So what do I do? How do I-- in this picture, how do I come up with a vector orthogonal to y?

Project. I take this z, and I take its projection. So z has a little piece-- that z vector has a big piece already in the direction of y, which I don't want, and a piece orthogonal to it. That's my other piece. That's my other piece. So here's y. And here's the-- that is z minus projection, let me just say. Whatever. Yeah.

I don't know if I even drew that picture right. Probably I didn't. Anyway. Whatever. The Gram-Schmidt idea is just orthogonalize in the natural way. I'll come back to that at the beginning of next time and say a word about the fourth way. So this least squares is not deep learning. It's what people did a century ago and continue to do for good reason. OK. And I'll send out that announcement about the class, and you know the homework, and you know the new due date is Friday. Good. Thank you.