The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high-quality educational resources for free. To make a donation or to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.


#### Abstract

GILBERT STRANG: OK, let me make a start. On the left, you see the topic for today. We're doing pretty well. This completes my review of the highlights of linear algebra, so that's five lectures.


I'll follow up on those five points, because the neat part is it really ties together the whole subject. Eigenvalues, energy, A transpose A, determinants, pivots-- they all come together. Each one gives a test for positive and definite matrices. That's where I'm going.

Claire is hoping to come in for a little bit of the class to ask if anybody has started on the homework. And got Julia rolling, and got a yes from the auto grader. Is anybody like-- no. You're taking a chance, right? Julia, in principle, works, but in practice, it's always an adventure the first time.

So we chose this lab on convolution, because it was the first lab last year, and it doesn't ask for much math at all. Really, you're just creating a matrix and getting the auto grader to say, yes, that's the right matrix.

And we'll see that matrix. We'll see this idea of convolution at the right time, which is not that far off. It's signal processing, and it's early in part three of the book.

If Claire comes in, she'll answer questions. Otherwise, I guess it would be emailing questions to-- I realize that the deadline is not on top of you, and you've got a whole weekend to make Julia fly.

I'll start on the math then. We had symmetric-- eigenvalues of matrices, and especially symmetric matrices, and those have real eigenvalues, and l'll quickly show why. And orthogonal eigenvectors, and I'll quickly show why.

But I want to move to the new idea-- positive definite matrices. These are the best of the symmetric matrices. They are symmetric matrices that have positive eigenvalues. That's the easy way to remember positive definite matrices. They have positive eigenvalues, but it's certainly not the easy way to test.

If I give you a matrix like that, that's only two by two. We could actually find the eigenvalues, but we would like to have other tests, easier tests, which would be equivalent to positive eigenvalues. Every one of those five tests-- any one of those five tests is all you need.

Let me start with that example and ask you to look, and then I'm going to discuss those five separate points. My question is, is that matrix s? It's obviously symmetric. Is it positive, definite, or not?

You could compute its eigenvalues since it's two by two. It's energy-- l'll come back to that, because that's the most important one. Number two is really fundamental.

Number three would ask you to factor that. Well, you don't want to take time with that. Well, what do you think? Is it positive, definite, or not? I see an expert in the front row saying no. Why is it no?

The answer is no. That's not a positive definite matrix. Where does it let us down? It's got all positive numbers, but that's not what we're asking. We're asking positive eigenvalues, positive determinants, positive pivots. How does it let us down? Which is the easy test to see that it fails?

AUDIENCE: Maybe determinant?

GILBERT Determinant. The determinant is 15 minus 16, so negative. So how is the determinant

STRANG:

## AUDIENCE: [INAUDIBLE]

GILBERT
STRANG:

It's the product. So the two eigenvalues of s , they're real, of course, and they multiply to give the determinant, which is minus 1 . So one of them is negative, and one of them is positive. This matrix is an indefinite matrix-- indefinite.

So how could I make it positive definite? OK. We can just play with an example, and then we see these things happening. Let's see.

OK, what shall I put in place of the 5 , for example? I could lower the 4 , or I can up the 5 , or up the 3. I can make the diagonal entries. If I add stuff to the main diagonal, I'm making it more positive. So that's the straightforward way.

## AUDIENCE: 6.

GILBERT $\quad$ 6. OK. 6 would be safe. If I go up from 5 to 6 , I've gotta de-- so when I say here "leading STRANG: determinants," what does that mean? That word leading means something. It means that I take that 1 by 1 determinant-- it would have to pass that. Just the determinant itself would not do it.

Let me give you an example. No for-- let me take minus 3 and minus 6 . That would have the same determinant. The determinant would still be 18 minus 16-- 2 . But it fails the test on the 1 by 1 .

And this passes. This passes the 1 by 1 test and 2 by 2 tests. So that's what this means here.

Leading determinants are from the upper left. You have to check n things because you've got n eigenvalues. And those are the n tests.

And have you noticed the connection to pivots? So let's just remember that small item. What would be the pivots because we didn't take a long time on elimination?

So what would be the pivots for that matrix, 3-4-4-6? Well, what's the first pivot? 3, sitting there-- the 1-1 entry would be the first pivot. So the pivots would be 3 , and what's the second pivot? Well, maybe to see it clearly you want me to take that elimination step. Why don't I do it just so you'll see it here?

So elimination would subtract some multiple of row 1 from row 2 . I would leave 1 one alone. I would subtract some multiple to get a 0 there. And what's the multiple? What's the multiplier?

AUDIENCE: In that much--

GILBERT $4 / 3.4 / 3$ times row 1 , away from row 2 , would produce that 0 . But $4 / 3$ times the 4 , that would
STRANG: be $16 / 3$. And we're subtracting it from $18 / 3$. I think we've got $2 / 3$ left.

So the pivots, which is this, in elimination, are the 3 and the $2 / 3$. And of course, they're positive. And actually, you see the immediate connection.

This pivot is the 2 by 2 determinant divided by the 1 by 1 determinant. The 2 by 2 determinant, we figured out-- 18 minus 16 was 2 . The 1 by 1 determinant is 3 . And sure enough, that

This is not-- so by example, I'm illustrating what these different tests-- and again, each test is all you need. If it passes one test, it passes them all. And we haven't found the eigenvalues.

Let me do the energy. Can I do energy here? OK. So what's this-- I am saying that this is really the great test.

That, for me, is the definition of a positive definite matrix. And the word "energy" comes in because it's quadratic, [INAUDIBLE] kinetic energy or potential energy. So that's the energy in the vector x for this matrix.

So let me compute it, $x$ transpose $S x$. So let me put in $S$ here, the original $S$. And let me put in of any vector x , so, say xy or x 1 . Maybe-- do you like $\mathrm{x}-\mathrm{-} \mathrm{xy}$ is easier.

So that's our vector x transposed. This is our matrix S . And here's our vector x .

So it's a function of $x$ and $y$. It's a pure quadratic function. Do you know what I get when I multiply that out? I get a very simple, important type of function. Shall we multiply it out?

Let's see. Shall I multiply that by that first, so I get $3 x$ plus $4 y$ ? And $4 x$ plus $6 y$ is what I'm getting from these two. And now l'm hitting that with the xy.

And now I'm going to see the energy. And you'll see the pattern. That's always what math is about. What's the pattern?

So I've $x$ times $3 x$, $3 x$ squared. And I have y times $6 y$. That's $6 y$ squared. And I have $x$ times $4 y$. That's for $4 x y$. And I have $y$ times $4 x$. That's 4 more $x y$.

So I've got all those terms. Every term, every number in the matrix gives me a piece of the energy. And you see that the diagonal numbers, 3 and 6 , those give me the diagonal pieces, $3 x$ squared and 6y squared.

And then the cross-- or I maybe call them the cross terms. Those give me $4 x y$ and $4 x y$, so, really, $8 x y$. So you could call this thing $8 x y$.

So that's my function. That's my quadratic. That's my energy. And I believe that is greater than 0.

Let me graph the thing. Let me graph that energy. OK. So here's a graph of my function, $f$ of $x$
and y .

Here is $x$, and here's $y$. And of course, that's on the graph, $0-0$. At $x$ equals 0 , $y$ equals 0 , the function is clearly 0 .

Everybody's got his eye-- let me write that function again here-- $3 x$ squared, $6 y$ squared, $8 x y$. Actually, you can see-- this is how I think about that function.

So $3 x$ squared is obviously carrying me upwards. It will never go negative. 6y squared will never go negative. $8 x y$ can go negative, right? If $x$ and $y$ have opposite signs, that'll go negative.

But the question is, do these positive pieces overwhelm it and make the graph go up like a bowl? And the answer is yes, for a positive definite matrix. So this is a graph of a positive definite matrix, of positive energy, the energy of a positive definite matrix.

So this is the energy $x$ transpose $S x$ that I'm graphing. And there it is. This is important. This is important. This is the kind of function we like, $x$ transpose $S x$, where $S$ is positive definite, so the function goes up like that.

This is what deep learning is about. This could be a loss function that you minimize. It could depend on 100,000 variables or more.

And it could come from the error in the difference between training data and the number you get it. The loss would be some expression like that. Well, I'll make sense of those words as soon as I can.

What I want to say is deep learning, neural nets, machine learning, the big computation-- is to minimize an energy-- is to minimize an energy. Now of course, I made the minimum easy to find because I have pure squares. Well, that doesn't happen in practice, of course.

In practice, we have linear terms, $x$ transpose $b$, or nonlinear. Yeah, the loss function doesn't have to be a [INAUDIBLE] cross entropy, all kinds of things. There is a whole dictionary of possible loss functions.

But but this is the model. This is the model. And I'll make it the perfect model by just focusing on that part.

Well, by the way, what would happen if that was in there? I shouldn't have X'd it out so quickly
since I just put it up there. Let me put it back up. I thought better of it.

OK. This is a kind of least squares problem with some data, b. Minimize that. So what would be the graph of this guy? Can I just draw the same sort of picture for that function?

Will it be a bowl? Yes. If I have this term, all that does is move it off center here, at x equals 0 .

Well, I still get 0 . Sorry. I still go through that point. If this is the 0 vector, I'm still getting 0 . But this, we'll bring it below.

That would produce a bowl like that. Actually, it would just be the same bowl. The bowl would just be shifted. I could write that to show how that happens.

So this is now below 0 . That's the solution we're after that tells us the weights in the neural network. I'm just using these words, but we'll soon have a meaning to them.

I want to find that minimum, in other words. And I want to find it for much more complicated functions than that. Of course, if I minimize the quadratic, that means setting derivatives to 0 .

I just have linear equations. Probably, I could write everything down for that thing. So let's put in some nonlinear stuff, which way to wiggles the bowl, makes it not so easy.

Can I look a month ahead? How do you find-- so this is a big part of mathematics-- applied math, optimization, minimization of a complicated function of 100,000 variables. That's the biggest computation. That's the reason machine learning on big problems takes a week on a GPU or multiple GPUs, because you have so many unknowns. More than 100,000 would be quite normal.

In general, let's just have the pleasure of looking ahead for one minute, and then l'll come back to real life here, linear algebra. I can't resist thinking aloud, how do you find the minimum? By the way, these functions, both of them, are convex. So that is convex.

So I want to connect convex functions, f-- and what does convex mean? It means, well, that the graph is like that.

## [LAUGHTER]

Not perfect, it could-- but if it's a quadratic, then convex means positive definite, or maybe in the extreme, positive semidefinite. I'll have to mention that. But convex means it goes up.

But it could have wiggles. It doesn't have to be just perfect squares in linear terms, but general things. And for deep learning, it will include non-- it will go far beyond quadratics. Well, it may not be convex. I guess that's also true.

Yeah. So deep learning has got serious problems because those functions, they may look like this but then over here they could go nonxconvex. They could dip down a little more. And you're looking for this point or for this point. Still, I'm determined to tell you how to find it or a start on how you find it.

So you're at some point. Start there, somewhere on the surface. Some x , some vector x is your start, x0-- starting point.

And we're going to just take a step, hopefully down the bowl. Well of course, it would be fantastic to get there in one step, but that's not going to happen. That would be solving a big linear system, very expensive, and a big nonlinear system.

So really, that's what we're trying to solve-- a big nonlinear system. And I should be on this picture because here we can see where the minimum is. But they just shift.

So what would you do if you had a starting point and you wanted to go look for the minimum? What's the natural idea? Compute derivatives. You've got calculus on your side. Compute the first derivatives.

So the first derivatives with respect to $x$-- so I would compute the derivative with respect to x , and the derivative of $f$ with respect to $y$, and 100,000 more. And that takes a little while.

And now I've got the derivatives. What do I do?

## AUDIENCE: [INAUDIBLE]

GILBERT I go-- that tells me the steepest direction. That tells me, at that point, which way is the fastest STRANG: way down. So I would follow-- I would do a gradient descent. I would follow that gradient.

This is called the gradient, all the first derivatives. It's called the gradient of $f--$ the gradient. Gradient vector-- it's a vector, of course, because f is a function of lots of variables.

I would start down in that direction. And how far to go, that's the million dollar question in deep learning. Is it going to hit 0 ? Nope. It's not. It's not.

So basically, you go down until it-- so you're traveling here in the x , along the gradient. And you're not going to hit 0 . You're all going here in some direction. So you keep going down this thing until it-- oh, I'm not Rembrandt here.

Your path down-- think of yourself on a mountain. You're trying to go down hill. So you take-as fast as you can. So you take the steepest route down until-- but you have blinkers. Once you decide on a direction, you go in that direction.

Of course-- so what will happen? You'll go down for a while and then it will turn up again when you get to, maybe, close to the bottom or maybe not. You're not going to hit here. And it's going to miss that and come up. Maybe I should draw it over here, whatever.

So it's called a line search, to decide how far to go there. And then say, OK stop. And you can invest a lot of time or a little time to decide on that first stopping point.

And now just tell me, what do you do next? So now you're here. What now? Recalculate the gradient. Find the steepest way down from that point, follow it until it turns up or approximately, then you're at a new point.

So this is gradient descent. That's gradient descent, the big algorithm of deep learning of neural nets, of machine learning-- of optimization, you could say.

Notice that we didn't compute second derivatives. If we computed second derivatives, we could have a fancier formula that could account for the curve here. But to compute second derivatives when you've got hundreds and thousands of variables is not a lot of fun. So most effectively, machine learning is limited to first derivatives, the gradient.

OK. So that's the general idea. But there are lots and lots of decisions and-- why doesn't that-how well does that work, maybe, is a good question to ask. Does this work pretty well or do we have to add more ideas?

Well, it doesn't always work well. Let me tell you what the trouble is. I'm way off-- this is March or something. But anyway, l'll finish this sentence.

So what's the problem with this gradient descent idea? It turns out, if you're going down a narrow valley-- I don't know, if you can sort of imagine a narrow valley toward the bottom. So here's the bottom. Here's your starting point.

And this is-- you have to have think of this as a bowl. So the bowl is-- or the two eigenvalues, you could say-- are 1 and a very small number. The bowl is long and thin. Are you with me? Imagine a long, thin bowl.

Then what happens for that case? You take the steepest descent. But you cross the valley, and very soon, you're climbing again. So you take very, very small steps, just staggering back and forth across this and getting slowly, but too slowly, toward the bottom.

So that's why things have got to be improved. If you have a very small eigenvalue and a very large eigenvalue, those tell you the shape of the bowl, of course. And many cases will be like that-- have a small and a large eigenvalue. And then you're spending all your time. You're quickly going up the other side, down, up, down, up, down. And you need a new idea.

OK, so that's really-- so this is one major reason why positive definite is so important because positive definite gives pictures like that. But then, we have this question of, are the eigenvalues sort of the same size? Of course, if the eigenvalues are all equal, what's my bowl like?

Suppose I have the identity. So then x squared plus y squared is my function. Then it's a perfectly circular bowl. What will happen?

Can you imagine a perfectly circular-- like any bowl in the kitchen is probably, most likely circular. And suppose I do gradient descent there. I start at some point on this perfectly circular bowl. I start down.

And where do I stop in that case? Do I hit bottom? I do, by symmetry.

So if I take x squared plus y squared as my function and I start somewhere, I figure out the gradient. Yeah. The answer is I'll go right through the center.

So really positive eigenvalues, positive definite matrices give us a bowl. But if the eigenvalues are far apart, that's when we have problems. OK. I'm going back to my job, which is this-because this is so nice. Right.

Could you-- well, the homework that's maybe going out this minute for middle of next week gives you some exercises with this. Let me do a couple of things, a couple of exercises here. For example, suppose I have a positive definite matrix, S, and a positive definite matrix, T. If I add those matrices, is the result positive definite? So there is a perfect math question, and we hope to answer it.

So S and T-- positive definite. What about S plus T? Is that matrix positive definite? OK. How do I answer such a question?

I look at my five tests and I think, can I use it? Which one will be good? And one that won't tell me much is the eigenvalues because the eigenvalues of $S$ plus $T$ are not immediately clear from the eigenvalues of $S$ and $T$ separately. I don't want to use that test. This is my favorite test, so I'm going to use that.

What about the energy in-- so look at the energy. So I look at x transpose, S plus T x. And what's my question in my mind here? Is that a positive number or not, for every $x$ ? And how am I going to answer that question?

Just separate those into two pieces, right? It's there in front of me. It's this one plus this one. And both of those are positive, so the answer is yes, it is positive definite. Yes.

You see how the energy was right. I don't want to compute the pivots or any determinants. That would be a nightmare trying to find the determinants for $S$ plus $T$. But this one just does it immediately.

What else would be a good example to start with? What about S inverse? Is that positive definite? So let me ask S positive definite, and I want to ask about its inverse.

So its inverse is a symmetric matrix. And is it positive definite? And the answer-- yes. Yes. I've got five tests, $20 \%$ chance at picking the right one. Determinants is not good. The first one is great.

The first one is the good one for this question because the eigenvalues. So the answer is yes. Yes, this has-- eigenvalues. So what are the eigenvalues of S inverse? 1 over lambda? So-yes, positive definite, positive definite. Yep.

What about-- let me ask you just one more question of the same sort. Suppose I have a matrix, S, and suppose I multiply it by another matrix. Oh, well. OK. Suppose-- do I want to ask you this? Suppose I asked you about $S$ times another matrix, M. Would that be positive definite or not?

Now I'm going to tell you the answer is that the question wasn't any good because that matrix is probably not symmetric, and I'm only dealing with symmetric matrices. Matrices have to be symmetric before I know they have real eigenvalues and I can ask these questions. So that's
not good.

But I could-- oh, let's see. Let me put it in an orthogonal guy. Well, still that's not symmetric. But if I put the-- it's transpose over there. Then I made it symmetric. Oh, dear, I may be getting myself in trouble here.

So I'm starting with a positive definite S. I'm hitting it with an orthogonal matrix and its transpose. And my instinct carried me here because I know that that's still symmetric. Right? Everybody sees that?

If I transpose this, Q transpose will come here, $\mathrm{S}, \mathrm{Q}$ will go there. It'll be symmetric. Now is that positive definite? Ah, yes. We can answer that.

Can we? Is that positive definite? So remember that this is an orthogonal matrix, so also, if you wanted me to write it that way, I could. And what about positive-definiteness of that thing? Answer, I think, is yes. Do you agree? It is positive definite?

Give me a reason, though. Why is this positive definite? So that word similar, this is a similar matrix to S? Do you remember what similar means from last time? It means that sum M and its inverse are here, which they are.

And so what's the consequence of being similar? What do I know about a matrix that's similar to S? It has--

## AUDIENCE: Same [INAUDIBLE]

GILBERT Same eigenvalues. And therefore, we're good. Right? Or I could go this way. I like energy, so

STRANG:

AUDIENCE: y [INAUDIBLE]

GILBERT $\quad \mathrm{y}$ transpose. So this energy would be the same as y transpose, Sy. And what do I know about
STRANG: let me try that one. $x$ transpose, $Q$ transpose, SQx-- that would be the energy. And what am I trying to show? I'm trying to show it's positive. So, of course, as soon as I see that, it's just waiting for me to-- let Qx be something called y , maybe. And then what will this be? that? It's positive because that's an energy in the $y$, for the $y$ vector. So one way or another, we get the answer yes to that question. OK. OK.

Let me introduce the idea of semidefinite. Semidefinite is the borderline. So what did we have?

We had 3,4 , 4 . And then when it was 5 , you told me indefinite, a negative eigenvalue. When it was 6 , you told me 2 positive eigenvalues-- definite.

What's the borderline? What's the borderline there? It's not going to be an integer. What do I mean? What am I looking for, the borderline?

So tell me again?

## AUDIENCE: 16 over--

GILBERT $16 / 3$, that sounds right. Why is that the borderline?
STRANG:

## AUDIENCE: [INAUDIBLE]

## GILBERT Because now the determinant is--

STRANG:

## AUDIENCE: 0.

GILBERT $\quad 0$. It's singular. It has a 0 eigenvalue. There's a 0 eigenvalue. So that's what semidefinite STRANG: means. Lambdas are equal to 0 .

Wait a minute. That has a 0 eigenvalue because it's determinant is 0 . How do I know that the other eigenvalue is positive? Could it be that the other ei-- so this is the semidefinite case we hope. But we'd better finish that reasoning.

How do I know that the other eigenvalue is positive?

## AUDIENCE: Trace.

GILBERT The trace, because adding 3 plus 16/3, whatever the heck that might give, it certainly gives a
STRANG: positive number. And that will be lambda 1 plus lambda 2. That's the trace.

But lambda 2 is 0 . We know from this it's singular. So we know lambda 2 is 0 . So lambda 1 must be 3 plus $5--5$ and $1 / 3$. The lambdas must be 8 and $1 / 3$, 3 plus 5 and $1 / 3$, and 0 . So that's a positive semidefinite.

If you think of the positive definite matrices as some clump in matrix space, then the positive semidefinite definite ones are sort of the edge of that clump. There the boundary of the clump,
the ones that are not quite inside but not outside either. They're lying right on the edge of positive definite matrices.

Let me just take a-- so what about a matrix of all 1s? What's the story on that one-- positive definite, all the numbers are positive, or positive semidefinite, or indefinite? What do you think here? 1-1, all 1.

## AUDIENCE: Semi--

| GILBERT | Semidefinite sounds like a good guess. Do you know what the eigenvalues of this matrix would |
| :--- | :--- |
| STRANG: | be? |

## AUDIENCE: 0 [INAUDIBLE]

## GILBERT <br> 3,0 , and $0-$ - why did you say that?

## STRANG:

AUDIENCE: Because 2 [INAUDIBLE]

GILBERT Because we only have-- the rank is?
STRANG:

## AUDIENCE: 1.

GILBERT Yeah, we introduced that key where the rank is 1 . So there's only one nonzero eigenvalue.

## STRANG:

 And then the trace tells me that number is 3 . So this is a positive semidefinite matrix.So all these tests change a little for semidefinite. The eigenvalue is greater or equal to 0 . The energy is greater or equal to 0 . The A transpose A-- but now I don't require-- oh, I didn't discuss this. But semidefinite would allow dependent columns.

By the way, you've got to do this for me. Write that matrix as A transpose times A just to see that it's semidefinite because-- so write that as A transpose A. Yeah. If it's a rank 1 matrix, you know what it must look like.

A transpose A, how many terms am I going to have in this? And now I'm thinking back to the very beginning of this course if I pulled off the pieces. In general, this is lambda 1 times the first eigenvector, times the first eigenvector transposed.

AUDIENCE: Would it just be a vector of three 1s?

GILBERT
STRANG:

AUDIENCE: Normalize them.

GILBERT

## STRANG:

## AUDIENCE:

GILBERT
STRANG:
[INAUDIBLE]

Yeah, it would just be a vector of three 1s. Yeah. So this would be the usual picture. This is the same as the Q lambda, Q transpose. This is the big fact for any symmetric matrix.

And this is symmetric, but its rank is only 1 , so that lambda 2 is 0 for that matrix. Lambda 3 is 0 for that matrix. And the one eigenvector is the vector 1-1-1.

And the eigen-- so this would be 3 times 1-1-1. Oh, I have to do-- yeah. So I was going to do 3 times 1-1-1, times 1-1-1. But that gives me 3-3-3. That's not right.

I have to normalize them. That's right. Yeah. So that's a vector whose length is the square root of 3 . So I have to divide by that, and divide by it. And then the 3 cancels the square root of 3 s , and I'm just left with 1-1-1, 1-1-1. Yeah.

So there is a matrix-- one of our building-block type matrices because it only has one nonzero eigenvalue. Its rank is 1 , so it could not be positive definite. It's singular. But it is positive semidefinite because that eigenvalue is positive.

OK. So you've got the idea of positive definite matrices. Again, any one of those five tests is enough to show that it's positive definite. And so what's my goal next week? It's the singular value decomposition and all that that leads us to. We're there now, ready for the SVD.

OK. Have a good weekend, and see you-- oh, I see you on Tuesday, I guess. Right-- not Monday but Tuesday next week.

