PROFESSOR: So we started last week on the big topic for the rest of the semester, optimization. Maybe can you close that door or just part way anyway. Thanks. Great. That's perfect, just like that.

So, since it was a few days ago, I wanted to recap what I did in the first lecture about optimization, which was to pick out the least squares problem as a beautiful model problem. I followed that through -- the input is a matrix a rectangular, a right hand side b probably measurements. We would like to get au equal b but we can't. We got too many measurements. We do the best possible, which is the solution u hat of this normal equation. Then I rewrote the normal equations -- well, I also drew the picture that you see over here on the left. The picture that shows the two optimization problems at the same time.

Here is the vector b. Here are all possible au's -- this is the column space, this is all au's. The best au was the projection. The error e was what we couldn't get right, the part that's perpendicular to the column space we can't help. It's the solution to another projection problem that e is the same -- of course, it's the same over here -- as projecting b onto this perpendicular line, which is the line of all vectors, a tranpose e says -- what that says in words is e is perpendicular to the columns of a. I've drawn it the best I could as a perpendicular to the columns. So that we have a plane of columns -- this is like a 3 x 2 matrix. We have a plane with the two columns and the perpendicular line.

Then just near the end, I said that it would be great to have -- this model would be almost all we need except it should have one more matrix and I call that matrix c. I want just to show you quickly where that comes. So this is all section 7.1 of the notes, except that these words are not yet typed. So they'll, as soon as possible, and an updated 7.1 will include this, what I want to do. Because I think it's the very

best introduction I could give to these pair of optimization problems. You might think who needs optimization. Your main activity might be solving differential equations.

So, can I just take a time out here because I happened to see the homework upcoming. I think it's 16930, so it's a course a little bit like this, only the first word in the course title is "Advanced," and our first word is "Introduction," but, of course, it's the same course or same stuff. This is the homework that I don't think has even been assigned, even been handed out yet. But I just thought it's a great example of a applied, so that you see where optimization appears and what's involved. So differential equations are involved, but also you got something that you want to optimize. So I just wrote it on this board and I simplified it a little over the homework that they'll actually get.

so here's the problem. We want a certain distribution of heat. So I could draw a picture. We want a heat distribution for whatever reason that maybe goes like this over the integral zero to 1. What do we have at our disposal to get the heat to be that way? Well, we've got sources of heat. But we don't have a continuous source, we only have n parameters to play with. I mean right away you recognize an optimization problem. We're trying to get this function here u knot of x. We're trying to match a function, but we've only got n parameters. Those will be the right hand side. So what we're allowed to choose, we can put in little space heaters, and we can turn them to the temperatures we want to -- temperatures s1, s2, s3. That was probably a stupid choice to put an s for down there because I don't even know if negative heat is allowed. Anyway, we wouldn't want it if we're aiming for that distribution.

Now you understand the s's are not supposed to match that u knot. The s's are the sources of heat, and the u's are the outputs, the distribution, and they are controlled by a differential equation. What we control is the right hand side of the equation, but we only control n parameters. Then we have to solve the equation and find out what distribution that gives. So that's all like differential equations, we know how to do it. It's one dimensional in this example, so straightforward.

But now comes the optimization part. We take this result and we compare it with the desired result. So the actual result from some source of heat might be something like that. We want this u of x, the heat distribution that we're actually producing to be as close as possible to u knot. Close could mean different things, but if we measure closeness in this integral square mean square sense, then we're going to have nice problem. In fact, we're going to have a linear -- everything's linear right now. Well, everything's quadratic here, so when we minimize it's going to give us a linear equation for the s's.

I just think that's a good model problem. A good model of what we might do. I think actually in the problem to be assigned -- well, it's not a big deal for us. There is convection as well. So this is a pure diffusion problem, just the second derivative. We know very well that if there was a first derivative in there, the stuff's convecting, passing through the region. So if I put on a convective term, cdudx, how does that change the problem? Not in a big, big major way, but one thing we can guess is that now that is not a symmetric term, right. We've seen the difference between first differences and second differences, first derivatives and second derivatives.

So now it's a little trickier and it's not symmetric. So somehow there will be a primal problem, this one, and there will be a adjoint problem, dual problem, perpendicular problem, whatever name you want to give it, just as there is in our very first model over there. So that's my sort of quick look to kind of put on the board one example that I didn't invent, that came from applications and gives a sort of typical of what you have to do. You control an input, you get an output, so that's the analysis problem. Find the output. But then comes the optimization problem -- make that output close to something that you wish. So what's the typical algorithm going to do? It's going to make a choice of s, it's going to solve the analysis problem for the u, it's going to look and see what the error is, it's going to figure out probably the gradient somehow.

What's the steepest way to make it closer? That's going to lead us to a change of s. We use a change of s, the new s, solve that, and iterate. That would be a typical algorithm. We might be able to shortcut it in a model problem like this. But that's

totally the typical optimization idea, is an analysis problem and then figure out a gradient to get a better source back to the analysis problem with the new source. What the algebra, the math, has to do is those two steps of figuring out OK, how do we improve the error, how do we reduce the error, what's the steepest direction? Somehow we got to compute a derivative. Actually, that's what this month is about. Derivatives that are not just like the derivative of x cube or something.

I often wondered how many presidents could take the derivative of x cube and I'm not sure. Anybody occur to you who you could count on being able to take the derivative of x cubed? I don't think the current president would know what it meant. But I think Carter could have done it, because he went to the Naval Academy. Jefferson was probably -- he knew everything. I don't know. Anybody else has another candidate they can tell me. So there is our problem -- finding derivatives that would be definitely beyond the capacity of the White House.

Now I want to stay with this model a little more because it's the perfect model. So this was the like model 1, unweighted ordinary least squared, and it produced the identity matrix in there. I mentioned last time what I want to do, the more general model has another symmetric positive-definite matrix in there, but not necessarily i, and it comes from weighted least squares. So that's what I'm going to talk about.

So what are weighted least squares. Well, you've got these measurements and you think they maybe don't all, maybe they're not independent, maybe they're not equally reliable. So you weight them by how reliable they are. A more reliable one you would put a heavier weight, w, on because you want that to be more important. So, you change to this problem. But it looks practically the same. The only difference is a has become wa, b has become wb. So the equations will be the same, but a is now wa, b is now wb, and those are the equations. I guess I should call, just to make clear that u is a different u, that the best u now depends on the choice of weights. I should really be calling that u -- somehow indicate that it depends on the weight.

Now they key nice thing is that if I write this out as a tranpose -- can you just write

this out. You get the a tranpose, and the a is over here. But what's in the middle? What's this matrix c -- I jumped the gun and called that matrix in the middle c, but how is it connected to w? You can see it here as c is w tranpose w, right? It's just sitting there in the middle. That's great that the fact that the combination w tranpose w is all you need to know. So we can forget w in favor of -- I'll just put it here -- w tranpose w is now given the name c, and this matrix is symmetric positive-definite. So it's a great matrix and it's exactly the one we want. This is exactly the equation we want.

So if I go back to writing it in this -- you know, this was the equation. You remember the point that we could go directly to one equation for the best u, or we could keep our options open and have two equations that led to the same one. They're totally equivalent. But the two equations will give us not only u, but the error, b minus au, as an other unknown. That's what we did up there. So we had two equations, and if we eliminated e we got to this. Now I want to have two equations, and if I eliminate e, I get to that. So let me see what those would be.

Well, here's one of them. a tranpose c -- e is zero -- you see, that's the only difference really. That the weighted normal equation, I just took a tranpose c, and then I took the b minus au together, and this is what I'm calling e. So one way to do it is just the new guy is just now e plus auw is still b. But now a tranpose ce is zero. Good deal. I mean that's quite nice. It isn't absolutely perfect though, because I've lost the symmetry. I'm putting a c in there and I don't really want it. I want a c but I don't want it there. So, I just make a small change. I'm going to introduce a new unknown that I'll call little w, apologies for the fact that it's also a w, it just happened to fit. That'll be the ce.

So I'm just calling this a new name here. So that now my e -- of course, if I just invert, e is c inverse w. So now I'm just going to write these equations with w instead of e because I like it better. So this equation is now a tranpose w equals zero. This equation e is disappeared in favor of c inverse w plus au equals b. So that's the system that I really like. That's the saddle point system, the Kuhn-Tucker system, the primal dual system, the fundamental system of the whole subject in this linear

matrix case. We haven't got functions, we got vectors, and we've got symmetry, and we've got linearity, and we've got a saddle point matrix that's now the s -- well, let me just change it here. It's just changed to this, c inverse. That's the fundamental matrix of the whole subject. So, s is the saddle point matrix.

So I wanted to get that far. You see that the whole picture was elementary linear algebra. Let me come back to the elementary figure that illustrates what's the geometry. How was the geometry affected by introducing this guy w, this weighting matrix or the c equal w transport w? Well, here was the picture from last time. A right angle picture. This was a right triangle. This line was perpendicular to that plane, but not anymore now. The second one it's a tranpose ce is zero. That's still a line, but it's not any longer perpendicular to the -- this is still all the au's. This plane is still the column space of all the au's.

We have the same b, but you see the problem is it lost its 90 degree angle, because the projection is now a projection, it's now an oblique projection, it's slanted. This is the best au, and if I occasionally keep up-to-date I'll put that uw there. There's still an error e, and this is still a parallelogram but it's not a rectangle anymore. Forgive my enthusiasm. I'm sort of happy that the picture and the algebra both come out so neatly. I totally agree that at this point I'm asking you to follow a model without giving you an application, and that's one reason I threw in this mention of a specific application that came from somewhere else. But this is the picture there.

So I'll say one more word about the picture. I said that we lost the right angle. We lost perpendicularity, and, of course, literally speaking we did. This is no longer -- this is not a right angle anymore. This is not a right angle anymore. It's not a right angle in the usual meaning of right angles. It is a right angle in the inner product that's associated with c. In other words, right angles here mean x transpose y equals zero. That's the idea of a right angle, right? x perpendicular to y, and they have different letters here.

Now over here I still have perpendicular, but I don't -- this is not the right inner product anymore. It should be a weighted inner product, weighted with this c in the

middle. So that's really what I mean by c orthogonal. Maybe I'll put those words down. So this weighted thing is -- if I can squeeze it, I doubt if I can -- is c orthogonality, weighted orthogonality. So let me circle the whole thing. The c is the w tranpose w. Just to say that we aren't giving up on dot products and perpendiculars and good equations, we're just changing them by inserting c every time in a product. What it means is that this is the natural inner product for the particular problem. This is the natural inner product for Euclid, right? But then from some specific application like this one or a million others, they have their own natural inner product, and the inner product for that particular problem would be one of these guys with some kind of a matrix c showing up.

So least squares and weighted least squares, that's my example one. Now I'd like to give a second example, a more mechanic -- will come closer to mechanics. Because this is least squares, statistics, algebra. But let me put on the middle board an application out of mechanics. It will be, say, I'll make it small and just a couple of springs with a mass between them and fixed at both ends. So this spring extends by some amount. This spring extends by some amount. There's a force on this mass. So there's a force on there mass, maybe just gravity, f. That's the external force from the mass of that's here in between the springs.

Then also acting on that mass are spring forces. This spring is pulling it up, right? There's a spring force, w1. Here, do you want me to draw -- really which direction is this? I'm going to draw it this way just to show the w2 drawn that way would actually be negative, because I think that this spring would get compressed, right -- this mass is pushing it down. This spring would be under compression. It would be pushing the mass back up. So the w2 in that picture would be negative. So this w1 will actually be positive and go the way the arrow is showing. This w2 would actually be negative and go not that way the arrow is showing.

But what's the -- oh, what equations have we got then? What's our optimization problem? Well actually, we have a choice. We could work with equations, period. Actually, one of the equations is pretty obvious. This mass is an equilibrium. So w1 is equal to w2 plus s. So it doesn't move. Or you might prefer me to write it, I would

rather write it, with w's on one side and source terms on the other. So that's the equilibrium equation.

So what decides, we want to know what these w's are, and these springs are extended. So that first spring is expanded by an amount e1, and this second spring is extended by an amount e2. Stretched or compressed. e1 is probably going to be positive here -- that spring's going to be stretched. This spring is going to be compressed, so that e2 is probably going to be negative. What's the mechanics here? Well, I can state it two ways, as I said. I can state the mechanics in terms of equations -- force and stretching elastic constant. That's how we did it in the first semester, 18085. It's a little clearer because simple equation [UNINTELLIGIBLE PHRASE]. Or I can state the problem as a minimization of energy. That's what I want to do today in 1806. So I want to minimize 8.6. I want to minimize the total energy. The energy in these springs. Subject to the constraint -- this constraint, equilibrium.

So that's the optimization statement of the mechanical problem. Well, I guess all that remains is -- well, I guess what remains? First of all, I need an expression for this energy. If the springs are governed by Hooke's law, then it'll be pretty simple. If they're real springs that don't quite obey Hooke's law then there'll be none -- no, there'll be fourth degree, sixth degree , whatever, terms in the energy. It's like the energy in the first spring plus the energy in the second spring, anyway. e in the first spring, and the energy in a second spring, and, of course, the two springs could have different spring constants. So those e's -- I'll make life easy in solving, if I choose Hooke's law, if I choose the energy to be just a square. The constraint is linear, so that'll be the model problem. And actually, that'll be the kind of problem I've got here.

One reason for introducing a new example was to get some mechanics into the lecture, but also to get a problem where we're doing a minimization but we've got a condition on the w's. The question is how do you find a minimum? You can't just set derivatives of the energy to zero. You would discover w1 equal w2 equals zero. Nothing happening. That would be the minimum. But that minimum is ruled out, that

solution is ruled out because we have this constraint. We've got to balance the external force. So this is the question and you'll maybe have met this question in other courses, but it's essential. How do you deal with minimizing when there's a constraint? I guess in some way we had it over here. We were minimizing something -- well, the minimum would be, take u equal u knot. But no, there was a constraint that u had to satisfy a certain equilibrium equation. Here it was a differential equation so that problem is a little harder than this one where the equation is just discrete, one simple equation.

So how are you going to do it? Well, actually the quickest way would be -- that's such an easy constraint that I could say hey, w2 is w1 minus f. So I could just, if I wanted to really like shortcut this whole lecture, I could say well, w2 is w1 minus f. Now I've accounted for the constraint. I've removed w2 from the problem. I have a minimization, an ordinary minimization with an unknown w1. I take the derivative. I solve derivative equals zero. I find w1, then I go back, I get that w2. That's the fast way. Of course, gets the right answer. But there's another way that in the end turns out to be better. It's not necessarily better for this simple problem, but it's better for the general approach to constrained optimization. So I'm going to not do the simple deal of solving for w2, but I'm going to keep the constraint around. It's the idea of Lagrange multipliers. You've heard those words and probably seen it happen.

So what is Lagrange multiplier? What is Lagrange's idea? Lagrange's idea is he constructed a function of w, which is this to work with, which is the same energy. But he's going to include a multiplier. Now the next question is what letter shall I use for that multiplier, Lagrange's multiplier. Books on optimization often call it Lambda -- Lambda's sort of like for Lagrange. So, Lagrange obviously wasn't Greek, but anyway -- close. Lambda for us always means eigenvalue -- For me always means eigenvalue. So I'm reluctant to use Lambda. Sometimes in books on economics, which is we're doing this all the time, the multiplier's called pi because it turns out to be a price. But let me use u for the Lagrange multiplier. So that'll be the Lagrange multiplier.

So what do I do with it? I multiply this equation by u and I build it in to this. So this

thing is this part -- I'll just copy that down there -- plus or minus, depending what I want, depending which sign I want to end up with, u the multiplier times the constraint -- w1 minus w2 -- let me put it like that. So the constraint is that this should be zero. So you can say I haven't added anything. I've added zero. But that will only be true at the end when I have a specific w1 and w2. Right now what I've done is I've built in the constraint into the function. Lagrange's brilliant idea was that now I've got a function that I can -- my derivatives I can take. I could set the derivatives to zero. dldw1 is going to be zero. dldw2 is going to be zero. And dldu is going to be zero. So we got now three equations, three unknown. Instead of going from two to one, we've gone from two up to three. But it's so much more systematic that it's the right thing to do.

What is this last equation? What's the u derivative of an expression? Well, u doesn't appear here, the derivative is just w1 -- this just leads us to w1 minus w2 minus f, which is the constraint that we wanted. So the constraint is showing up as this equation. Just the way the constraint showed up as this equation here. I guess what I want to say is when I wrote this down, when we did this first example, I didn't say here's the constraint. I mean that equation kind of came out from the normal equations here. So what we're doing new here is the constraint equation is sort of part of the mechanics and we're asking the question how do I deal with it. Here it came out of the geometry, so it wasn't there at the beginning so we didn't have to say how do we deal with this equation. It just emerged. Here it's really forced on us right away.

So anyway, we've still got to figure out the derivatives of those, and I guess I do have to finally now say what choice -- yeah. So this is -- what is this, the derivative of e1 minus u. If I take the w1 derivative, I've got the derivative of v1 with respect to w1. Then w1 doesn't appear there but it appears here, so it's minus u. This would be the derivative of v2 with respect to w2, that second spring minus -- oh maybe plus cu.

Well, those are the three equations. Let me move now to the linear case, just so we see the beautiful pattern. So if I make the equations linear, what's the energy in

Hooke's law spring here if the extension is e1 and if it produces a force of w1, I want to know what is this e1 here. So let me just remember Hooke's law. I think Hooke's law would say that -- well, there's some elastic constant c1, so there's a c1 and a c2 that tell us how hard or soft the springs are. So these are physical constants. If I remember right, the energy -- so I'm going to erase a little here just to -- well no. So what is the energy in that first spring? You remember, there's a 1/2 of an c1 e1 squared. That's the energy in a spring with constant c1 and the stretch e1.

But now I really want it not in terms of e1, I want it in terms of w1, just the way I did here. I've got to do the same thing here. I want to get to w, because the constraint is in terms of w. What does Hooke's law say? Hooke's law says w equals ce, that's Hooke, Hooke's law. The fourth is the elastic constant times the stretch. So in place of the e I have w over c. So this is 1/2, e1 squared will be w1 squared over c1 squared, and then a c1 cancels -- that's what it looks like. I guess I'm certainly happy to see that I'm coming up with c inverse. This c is showing up in the denominator, and that's exactly the -- so this is what I mean, this is what I want for my energy is a 1/2, w1 squared over c1, and a 1/2, w2 squared over c2, and now I was doing a minus sign because that kind of goes well with mechanics where a plus sign would go well with all the other applications.

So now I've got explicit energy. So now I can say what this thing really -- the derivative with respect to w1. OK. The derivative with respect to w1 is just w1 over c1. Is that what I'm getting? This is zero. This says w2 over c2 minus u is zero, and this one says w1 minus w2 equals f. Our three equations. Yes, so I will help kill these equal signs and just look at -- oh, that's a plus. We've got our saddle point matrix again. That's the nice thing here. This is a problem with a 3 x 3 matrix, with three unknowns, w1, w2, and u. With right hand side zero zero and f. With what matrix? That looks like a 1 over c1, zero and a minus 1. This looks like zero, a 1 over c2 and a plus 1. This looks like -- oh, wait a minute. I don't want this. w1 minus w2 equal f. I can live with 1 and minus 1 there, but it's not really what I wanted. I wanted the signs to -- you know what I want here. I want the transpose of that to be here, and zero to be in that block. So that's what my saddle point matrix would look like.

Well, let me just say that I could live with either. I was aiming for this one because it's symmetric. But a lot of people would rather have the opposite signs and have the 1 minus 1 there. I don't care which sign f has, of course. Some people these days are liking this form better because then it has a symmetric part and an anti-symmetric part. I mean the thing is at some point we're going to get problems like this with thousands of unknowns, and we're going to think how do we solve them and maybe some iteration. So we might want the matrix to be symmetric but indefinite, or we might want a positive-definite symmetric part and an anti-symmetric part.

What we can't have is positive definite-symmetric. That's like asking for what can't happen here. The combination of the problems is producing a saddle point and we can play with that sign, but we can't make that zero something -- oh, we could actually. What I was going to say is we can't make that zero something different, but you could. That would be a possible way to -- it's another way people thought of of solving these problems is artificially throw in a big number there, or even a small number, and push things towards positive. Anyway, my purpose today is essentially completed, that we're getting out of physical application after I linearized and it became a linear equation and it had that saddle point form. So, saddle point form here, saddle point form here, saddle point forms everywhere. I mean we'll have the saddle point forms for differential equations, as well as for matrix equations.

So those are the examples to sort of hang on to, and it's section 7.1 that has a big part of it. Ah -- I always have one last thing to say.

What's the meaning of u? So, u was a Lagrange multiplier. Lagrange just like helped us out by saying OK, there was a constraint by using one of my multipliers. But the point is that the multiplier always has a real meaning. I mention prices before. Here, what's the meaning of u? What's the physical meaning of the Lagrange multiplier. It turns out to be the displacement of the math. It's the dual variable, so it always has some interpretation. In this case, with mechanics it's the amount, the mass comes down when the force asks. What's more, it also has a

derivative interpretation. Turns out -- I'll just put turns out that the Lagrange multiplier u turns out to be the derivative of the minimum energy in the system with respect to the source term. It's the sensitivity of the problem somehow. I'll just use sensitivity.

You often want to know -- it's actually this quantity that we need over there. We want to know how much does the answer depend on the source, and that's what the Lagrange multiplier tells us. So, if I computed the solution here, so the notes do this -- maybe I'll leave it for the notes. The notes solve this little problem. That's easy to do. They figure out what the energy is in the springs. It depends on the right hand side, f, which is just a number here. The energy turns out to be quadratic. You can take its derivative and you find out that it's u. So that Lagrange multiplier, that's really a key message, is an important quantity in itself. Here it happens to mean displacement, which is obviously crucial quantity, and in general it tells us the change in the minimum energy with respect to a change in the input. That sensitivity's a natural word to use for that.

So that's a final word about -- well, a near to final word about Lagrange multipliers.

So I'll see you Wednesday and by that time I'll know more about the projects and we'll be moving onward with optimization. Good. Thanks.