PROFESSOR: For me this is the third and last major topic of the course. The first one was initial value problems -- stabililty, accuracy. Topic two was solving large linear systems by iterative methods and also by direct methods like re-ordering the equations.

Now, topic three is a whole world of optimization. In reality it means you're minimizing or possibly maximizing some expression. That expression could be a function of several variables. We could be in the discrete case, so we've got -- maybe I just emphasized that we have discrete optimization. So that's in rn, discrete in n dimensions, and that will include some famous areas as single, as small subsets, really, of the big picture. For example, one subset would be linear programming. So that's a very special but important case of problems where the cost is linear, the constraints are linear, and has its own special method. So I think that's worth considering on its own at a later point in a lecture.

Another bigger picture -- for us actually bigger will be quadratic programming where the quantity that's being minimized is a quadratic function. Now what's good about a quadratic function? Its derivatives are linear. So that leads us to linear equations, but always with constraints. We don't have a free choice of any vector in rn. We have constraints on the vectors, they have to -- maybe they solve a linear system of their own. We might be in 100 dimensions and we might have 10 linear equations as the unknowns have to solve. So in some way we're really in 90 dimensions, but it might -- you know, how should we treat those constraints? Well, you know that Lebesgue multipliers play a role.

So there's a big area, very big area, of discrete optimization. Then also there is the continuous problem where the unknown is a function, is a function u of x I'll say, or u of x and y. It's a function. That's why I refer to that area as continuous optimization. Well, first you always want to get an equation, which is in some way going to be derivative equals zero, right. When we learn about minimization in elementary calculus, somewhere along the line is going to be an equation that has something like derivative equals zero.

But, of course, we have to account for the constraints. We have to ask what is the derivative of what when our unknown is a function. I'm just going to write down a topic within mathematics that this is often expressed as. Calculus -- that would be derivative. But in the case for functions it's often called the calculus of variations. So that's just so you see that word. There are books with that title. The idea of what is that derivative when our unknown is a function and the objective that we're trying to minimize is the integral of the function and its derivatives -- all sorts of possibilities there.

So that that's a very quick overview of a field that we'll soon know a lot about. I was trying to think where to start. I think maybe it better be discrete. I want to get to the system of equations that you constantly see. So let me use as an example the most basic problem which comes for -- maybe I'll start over here -- the problem of least squares, which I'll express this way. I'm given a matrix a, and a right hand side b, and I want to minimize the length squared of au minus b. So that would be a first problem to which we could apply calculus, because it's a straight minimization and I haven't got any constraints in there yet. We could also apply linear algebra, and actually linear algebra's going to throw a little extra light. So, just what am I thinking of here? I'm thinking of a as being m by n with m larger than n. If a is the square matrix, then this problem is the same as solving au equal b. Of course, we'll always reduce to that k'th, if m equals n.

But to focus on the problems that I'm really thinking about, I'm thinking about the case where this is n is the number of unknowns. It's the size of u. m, the larger number, is the number of measurements, the number of the data, so it's the number of equations, and it's the size of b. So we have more equations than unknowns. You've met these squares before. I hope that maybe even in these few minutes, a little new light well be shed on these squares.

So here's our problem, and calculus could lead us to the equation for the best u. So, u stands for u1, u2, up to un. There are n components in that vector u. Maybe you know the equation -- I guess I hope you know the equation, because it's such a key to so many applications. If I just write it, it will sound as if problem over. Let me write it though. So that the key equation, and then this comes up in statistics, for example. Well, that's one of 100 examples. But in statistics, this is the topic of linear regression in statistics. Let me write down -- they gave the name normal equation to the equation that gives you [UNINTELLIGIBLE]. Do you remember what it is? The normal equation? The equation for the minimizing u, which we could find by calculus? It involves the key matrix a transpose a. Let me call u hat the minimizer, the winner in this competition. The right hand side of the equation is a transpose b.

So I won't directly go back to derivitals, so probably I'm going to end up deriving it, because you can't help approaching that equation from one side or another. As I say, one way to approach it would just be to write out what that sum of squares is, take its derivative and you would get linear equation. So again, u hat stands for the u that gives the minimum. This a transpose a, of course, so I'm putting in a little bit of linear algebra. This matrix a transpose a is obviously symmetric. It's important property, beyond the symmetry, is that it's positive-definite. Well, I have to say positive-definite -- there's always a proviso, of course. I haven't eliminated the degenerate case yet. a has m, many, many rows, a smaller number of columns, and let's assume that those columns are linearly independent so that we really do have n unknown.

If those columns were, say if all the columns where the same, then au would just be multiplying that same column and there would really be only one unknown. So I'm going to say that a has rank n by which I mean n independent columns. In that case, that's what guarantees that this is positive-definite -- let me try to draw an arrow there -- this is the same statement. If I say about a that the columns are independent, then I'm saying about a transpose a that it is positive-definite. That means all its eigenvalues are positive, it's vertible certainly. All its pivots are positive. It's the great class of matrices.

But I don't really want to start with that equation. Here's my point. Optimization -- a key word that I better get on the board maybe up here to show that it's really important, is plus the idea of duality. The effect of duality, if I just give a first mention to that word, is that very often optimization problems, there are really two problems. Two problems that don't look identical, but in some important way they both, each problem is a statement of the task ahead of us. What are the two problems, the two dual problems in this basic example of these squares.

All right, here's a good picture. Here's a good picture. Let me put it on this board so I can recover it. So, minimize au minus b. So I think of the vector b as being where it's in m dimension. So it's a picture, I'm in m dimensions here. Now what about au? au -- where will au go in this picture? So, au is -- all the candidates au is multiply a by any vector u, so that means au is a combination of the columns of a, the possible vectors au lie in a subspace. So this is a subspace of all possible vectors au. It's only n dimensional. This is an n dimensional subspace because I have only n parameters in u, only n columns in a. So the set of all au's I think of as a, you could say a plane, an n dimensional plane within the bigger space rm.

Another name for that subspace, that plane, is the -- in 1806 I would call it the column space of a, or the range of s is another expression that you--. All the possible au's and here's b, which isn't one of the possible au's. So where is u hat? Where is the best au -- the best au now? The one that's closest to b is -- now comes another central word in this subject. If I draw it, I'm going to draw it here. That will be my best au, which I'm calling au hat. That's the, if the picture seems reasonable to your eye, this is the vector that's in the plane closest to b. What's the geometry here? See, that's what I wanted to see -- a little geometry and a little algebra, not just calculus.

So the geometry is that this vector b, what's the connection between b and that vector -- that's the closest vector, right? We're minimizing the distance. This distance here, I might call that the error vector e. This is as small as possible. That's being minimized. That's the difference between -- it's a pythagoras here. Of course, when I say it's pythagoras, I'm already saying the most important point that this a right angle here. That's a right angle. The closest au, which is au hat, which is this vector, the way geometrically we know it's closest is that the line from b to the plane, that's where the line from b, perpendicular to the plane. This line, this error vector e is perpendicular to the plane.

There's a good word that everybody uses for this vector. Take a vector b that's not on a plane, what's the word to look for the nearest vector in the plane?

AUDIENCE: Projection.

PROFESSOR: Projection. So this is the projection orthogonal projection, if I wanted to really emphasize the fact that that's a right angle. So that would give me a geometric way to see the least squares problem.

Now comes the point to see the dual problem. The dual problem will be here if I draw the perpendicular subspace. So that's a subspace of what dimension? This contains all the vectors perpendicular to the plane. So I have it as -- if m is 3, so we're in three dimensions, and our plane is an ordinary two dimensional plane, then the dimension is one -- that's the perpendicular line. But thinking bigger, if we're in m dimensions and this plane is m dimensional, than this is going to have -- the true dimension of this is m minus n, which could be pretty substantial. But anyway, that's the perpendicular subspace.

If this is the column space of a, I can figure out what vectors are perpendicular to the columns of a. That's really what I mean. This contains -- I've drawn it as a line, but I've written up there its dimension so that you see -- I just don't know how to draw like a bigger subspace. Yet you would have to see that all the vectors in it were perpendicular to all of these vectors. You see what I'm saying? If we were stuck in thinking in three dimensions, if I make this a plane I can't make that a plane. If I make m equal 3, and I make n equal to two, I'm only got a line left to be perpendicular. But in higher dimensions there are lots of dimensions left.

So what's my dual problem? My dual problem is find the vector e in this plane closest to b. In other words, by the same reasoning, what I'm saying is take the vector b, project it over to this plane, project it orthogonally -- that same right angle is going to be there. This plane -- I haven't said what's in this -- I've said what's in this plane but I haven't written it yet. But you could tell me already what is this? What's that vector? One answer would be it's the projection of b onto this perpendicular. So you see we're really taking the vector b and we're separating it into two components. One in the column space, the other perpendicular to the column space. So just tell me what that vector is. It is e. Same guy. In other words, e is the solution to the dual problem -- maybe I call this projection p for the best vector in the plane. e is the best vector in this subspace, and they add up to--. So, we're really taking the vector b, and we're splitting it into a part p in this space, and a part e in the perpendicular space. If I just write down the equations for that, I'll see what's cooking.

Well, I guess what I have to do is remember what are the equations to be in this subspace, to be perpendicular to the column. So I can't go further without remembering what's in that subspace. So everything in that subspace is perpendicular to the columns of a. Let me just write down what that means. Let me use the letter maybe y for the vectors in that subspace, and e for the winning vector, the projection. So y will be the vectors in that subspace. So those vectors are perpendicular -- so this is the subspace of y, all the y's in here. Now, what's the condition? So y -- that y in this perpendicular subspace. What do I mean? I mean that y is perpendicular to the columns of a. How shall I write that? Perpendicular means inner product zero. So I want to change the columns into rows, take the inner product with y and get zeros. Zero zero zero.

So, this is column 1 transpose to be a row. I'm trying to express this requirement in terms of a matrix, so to be perpendicular to the first column, I know that means that the inner product of the first column with y should be zero. The inner product with the second column -- the second column with y should be zero. The nth column, it's inner product with y should be zero. So what matrix have I got here? What's the condition on y's, simple and beautiful? What matrix is that? It's a transpose. So that perpendicular thing -- this is completely expressed by the equation a transpose y equals zero. That tells me the y's, and, of course, e is going to be one of the y's. It's going to be, I could say y hat, but I've already named it e. It's the particular one that's closest to b -- the y's are everybody all along here -- this is the null space of a transpose. So in words, I would call it that perpendicular thing is the null space of a transpose. So when you did linear algebra, you'll remember that. That the null space of a transpose -- let me write it. It is perpendicular to the column space of a.

The fundamentals theorem of linear algebra right there. Now we're just using it again to see what are the two dual problems here. So the prime problem, the one that we stated first, was this one. So I'll call this the primal -- p for primal. What is the dual problem? The dual problem is the problem about the y's. Not about the u's at all. That's the beauty of this duality. One problem is about u's, and it's a problem that ends up projecting on the column space. The second problem is about y's, it's the problem that ends up projecting onto this perpendicular space, and it was a projection. So the dual problem is just minimize the distance from b to y, but with the constraint with -- and now I get to use that word constraint -- with a transpose y equals zero. So there is the other problem.

So I hope your eye can travel between the -- well, let me write underneath it the primal again. Minimize au minus b square. This is the one whose solution is e, and this is the one whose solution is, well, u, and the projection u at, and the projection p is au hat, and I guess what I'm trying to say is that somehow there's a very important connection between the two problems. First of all, the two problems use the same data. They use the same vector b, they use the same matrix a. Notice that in one problem it's a, and in the other problem it's the transpose appears. That's very common -- we'll see that always. But there's something a little different about the two problems. This problem was unconstrained, any u was allowed. This problem was constrained, only a subset of y's, only that subspace of y's was allowed. This is a problem with n unknown. This is a problem with m minus n unknown. m minus n unknown variables, once we've accounted for the constraint. This is one thing I'm thinking about. Often, the problem will come with a constraint. Maybe I'll do a physical example right away.

The problem comes to us with a constraint. In other words, suppose you were given this problem. How would you deal with it? That's like the first question in optimization, or one of the central questions. How do you deal with a constraint? If we minimize this, of course, the minimum would be when y equals b. But that's failing to take into account the constraint on y. So how do you take constraints into account and end up with an equation? We can see in this picture that somehow or other we ended up with this normal equation, but actually I would rather end up with a primal dual equation. I'd like to end up with an equation for the best u and the best y. So what will that be?

So I need now two equations that will connect the best u and the best y, and probably this is going to be the key. This b is au hat, right. So this will be one of my equations, and this will be the other. Let me see if I -- well, OK. I don't know what to do now. Here I've called it y. Over here it's e. I've got myself in a corner. Maybe I should call e y hat, would you like that? We have in mind that it's e, the error in the primal problem. But just to make the notation for the two problems consistent, let me call the winner here y hat, the winner here u hat. What's the relation? So let me just -- here I'll write down the relation between the two. Well, it's over there. Let's see, is that right? Yes? y hat plus a u hat is b, and a transpose y hat is zero. That's it. That's it.

Here we have -- that's the pair of equations that solves, that connects the primal and the dual, solves them both, solves each one, and is really, it's a system -- you could say it's a block equation. A block matrix being identity a, a transpose zero, the unknown being the y and then u. The right hand side being the data, which in this case was the b. I guess what I want to do is emphasize in what's coming for the month of April. The importance of this class of problems. It's dealing with two -- it's dealing with the primal and the dual at the same time. It's important for so many reasons I can't say them all on the first day. That would be a mistake to try to say everything the first day. But let me just say something -- that linear programming, which is just one example, and it doesn't fit this because it has inequality constraints. But you maybe know that the number one method to solve linear program is called the simplex method. Well, it was the number one method for many years. For many problems it's still the right way to do it. But a new method called the primal dual -- at least that's part of its name, primal dual. It is essentially soving the primal and the dual problems at once, and there are inequality constraints, of course. I'm going to stop there with linear programming and give it its turn later.

In this perfect example here, we have only equations. How many do we have? We have m plus n equations, because y is m unknowns, u is n unknowns. I have all together m plus n equation -- m y's, and n u's, and they come together. Now, could you just to connected back with what we absolutely know that it's a normal equation, where is this normal equation coming from? So here's the normal equation. We know that that's gotta come, right, out of the [? paper. ?] How does it come? Suppose I have a block system, 2 x 2. How do I solve it? Well actually, that's, in a way, the big question.

But one way to solve it, the natural way to solve it would be elimination. Multiply this first row by a suitable matrix. Subtract from the second row to produce a zero there. In other words, eliminate y and get an equation for u hat. So what do I do? How do I do it? I multiply -- would you rather look at equations or matrices? I've tried to keep the two absolutely together. Let me look at the equation. What shall I multiply that equation by and subtract from this -- I just want to eliminate. I want to get y hat out of there and leave just an equation for u hat that we will totally recognize.

So what do I do? I multiplying that first equation by a transpose, thanks. So I multiply this first equation by a transpose. Let me just do it this way. Now what? a transpose y is zero. Now I use the second -- well, this is one way to do it. a transpose y is zero. What am I left with? The normal equation. Well, it shouldn't be a surprise. We had to end up with a normal equation. Maybe you would rather -- and actually, what I intended was to make it sound more like Gaussian and elimination. Multiply this row by a transpose, subtract from this row. I still have identity a up here -- when I do that subtraction I get the zero, that was the point. Here I get a transpose a, subtracted from zero is minus a transpose a, y hat, u hat. Of course, I had to do the same thing to the right hand side, and when I subtracted this the b was still there, but it was minus a transpose b.

So, now I've got in the second equation -- the second equation only involves u hat, and, of course, when I changed the signs, it's our friend -- a transpose au hat equal a transpose b. So this is -- maybe you would say the natural way to solve this type of system, but I want to emphasize, throw it away. I really want to emphasize the importance of these -- let me clean it back up again to what it was -- of these block systems.

Now, they need a name. We have to give some name to this type of two field problem. I guess then in the next month I'm going to find examples of it everywhere. So here I've found the first example of it in ordinary old-fashioned least square. So what am I going to call this? I'll call it -- I'll give it a couple of names. Saddle point equation, saddle point system maybe I should say. I have to explain why do I call it saddle point system. In optimization, I could call it the optimality equation -- just meaning it's the equations for the winners. In the world of optimization, the names of Kuhn and Tucker are associated with these equations -- Kuhn-Tucker equations, and there are other names we'll see. But let me just say for a moment why saddle point. Why do I think of this as a saddle point problem? See, the point about a transpose a was that it was positive-definite. This is a transpose a.

Now what's the corresponding issue for this matrix? So this is my matrix that I'm constantly gonna look at. Matrices of that form are going to show up all the time. I've said probably in 18085 where these appear but then we don't do much with them. Now we're ready to do something. I didn't appreciate their importance until I realized in going to lectures on applied mathematics that if I waited a little while that matrix would appear. That block matrix. It just shows up in all these applications. One of our issues will be how to solve it. Another issue that comes first is what's the general form of this? Can I jump to that issue just so we see something more than this single problem here.

Let me point in the matrix as it comes in applications. Some matrix a, rectangular, right? It's transpose. A zero. Often a zero. But what's up here is not always the identity. I want to allow something more general. I want to allow, for example, weighted least squares. So weighted least squares -- if you've met least squares it's very important to meet its extension to weighted least squares. When the equation au equal b are not given the same weight, there's a weighting matrix, often it's a covariance matrix. I'm going to call the matrix that goes in here c inverse. So this will be then an important class of application. This is pretty important already when the identity is there, but many, many applications produce some other matrix that is usually -- it very, very often is symmetric positive-definite in that corner, like the identity is. But key point, which I have to make today. Is the whole matrix either this one or this one. It is symmetric, right? That matrix is symmetric. Is it or isn't it positive-definite? If I do elimination do I get all positive pivots? It's a matrix of size m plus n.

So I'm asking are all its eigenvalues positive, but I don't want to really confuse eigenvalues. Also, in a lot of cases I would. Finding the eigenvalues of this matrix would be the lead me to the singular value decomposition, absolutely crucial topic in linear algebra that we'll see. But let me just take it as a linear system. If I do elimination, what are the first m pivots? Let me not be abstract here. Let me be quite concrete. Let me put the identity here. Let me put some matrix -- oh, I want the matrix here to be -- I better put a bigger identity just so we see the picture.

Here I'm going to put to an a transpose, which might be 2, 3, 4, 5, 6, 7 -- when I write numbers like that you'll realize that I've just pick them out of the hat. Here is the transpose 2, 3, 4, 5, 6, 7. Here's the zero block. I'd like to know about that matrix. It's symmetric. And it's full rank. Its invertible. How do I know that? It's the invertibility -- of course, the identity part is great. I guess I see that this is inveritable because I've ended up with a transpose a, and here my a has rank two -- those two columns are independent. They're not in the same direction -- 2, 3, 4 is not a multiple of 5, 6, 7. That's an inveritible matrix. This process finds the inverse. Elimination finds the inverse. It's the pivots I want to ask you about.

What are the pivots in this matrix? So what do I do? The first pivot is a 1 -- I use it to clean out that column. The second pivot is a 1 -- I use it to clean out that column.

The third pivot is a 1 -- I use it to clean out that column. What's the next pivot? What do I have -- of course, now some stuff has filled in here. What is actually filled in there? So this is the identity, if I can write it fast. This guy is now the zero. This guy didn't move. What matrix filled in here? Well, just what I was doing there. Elimination is exactly what I'm repeating with numbers, what I did there with letters. What's in here is minus a transpose a. I could figure out what I could do. If I was fast enough I could do 2, 3, 4, 5, 6, 7 times 2, 3, 4, 5, 6, 7 and I'd get this little 2 x 2 matrix that sits there.

With a minus sign, and that's the point. That the pivot -- of course, what's this first number going to be? 4 plus 9 plus 16. 29 is that number. So a minus 29 sits right there. That's the next pivot. The next pivot is a negative number, minus 29, and the fifth pivot is negative. So what I'm seeing is a matrix, this matrix has three positive pivots, and two negative pivots. I sort of say that was a saddle point. Positive pivots describe for me a surface that's going upwards. This surface is going upward. I'm a surface in five dimensions here. It's going upwards in three directions, but it's going downwards in two. The point at the heart of it, the saddle point, is the solution to our system, is the y hat, u hat. Well, one conclusion is that I wouldn't be able to use conjugate gradient methods, for example, which we've just learned how powerful they are, on the big matrix because it's not positive-definite. It's symmetric. I could use some other available methods. I couldn't use conjugate gradient. So if I wanted to use conjugate gradient, I better do this reduction to the definite system.

We're longer than I intended to spend on the simple example. But if you see that example, then we'll be ready to move it to the wide variety of applications. So let me just note that one section already up on the web called saddle point systems solves differential equations that are of this kind. So we'll come to that, and we'll come to matrix problems, too. It's a very, very central question, how to solve linear systems with matrices of that form. In fact, I guess I could say it's almost a fundamental problem of numerical linear algebra is to solve systems that fall into that saddle point description.

I'll try to justify that by the importance I'm assigning to this problem in the next week.

OK. Thanks for today and I'll turn off volume.