

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.

**PROFESSOR
STRANG:**

You can pick up the homeworks after if you didn't get one. And the good grade is five. And the MATLABs are coming in, that's all good. And I'll have some thoughts about the MATLAB Monday. I wanted to say a little more about finite differences in time just because they're so important. And I got the idea of a forward Euler and a backward Euler but I want to give you a couple of more possibilities just so you see how finite difference methods are created. And then I want to get started on least squares, the next major topic, the next case where A transpose A shows up at the core of everything.

So what we did with finite differences was this was our model problem for one spring and this is our model problem for n springs with n different masses. So this is like, scalar. u is just a single unknown. Here u is a vector unknown. If I reduce things to like, even this model problem, if I introduce the velocity, because everybody's also interested in computing velocity, then the velocity is u' and the equation tells me that v' is minus u . So I have a system. Better to think in terms of first order systems because they include everything. So the first order system there, do you see what the matrix is on that right-hand side? I'm always seeing a matrix that's multiplying $[u, v]$. So this would be our example of this general set-up. This general set-up. First order systems, taking them to be linear here. And then I better say a little bit about what happens when they're not linear today and later.

So this is my problem. $du/dt=Au$. And the exact solution would come from the eigenvalues and eigenvectors of A . We would have the $e^{(\lambda*t)}$'s, giving us the time growth or decay or oscillation, times their eigenvectors. And a combination of those would be the exact answer. So that's the general method for using eigenvalues and eigenvectors to get exact solutions. I'm not speaking now about exact solutions. I'm going to talk about finite differences in time because for more general problems I must expect to go to finite differences. I can't expect exact solutions like pure $\cos(t)$ or $\sin(t)$. So here's my problem, model problem. Here is Euler's first idea. So idea one from Euler was replace the derivative by a finite difference taking time steps forward in time and use the equation to tell you the slope at the start of the step. Do you see Euler's equation there? And that is definitely a reasonable thing to start with. It's not very

accurate. It's not perfectly stable. It doesn't preserve energy. We saw the answer spiraling out but if you only want to compute up to a limited time you could use it with a pretty small time step. But you can do better.

Now backward Euler was the other way. So it's just the contrast of the two that you want to see. Neither one is an outstanding method. They're both only first order accurate. So backward Euler, this time step in time is still the difference that replaces the derivative. But now everybody notices the big difference. The slope is being computed at the end of the time step. And that's more stable. That's the one that spiraled in. And I would call this method explicit. What does explicit mean? It means that I can find $u_{(n+1)}$ directly. This method I would call implicit. And what does implicit mean? Implicit means that $u_{(n+1)}$ is appearing on the right-hand side as well as the left so I've got to move it over and I have to solve a system of equations to find the next $u_{(n+1)}$. So you see that basic separation of ideas here. Forward, faster, but less stable. Backward, slower, generally more stable and in fact, in this case somehow too stable. It dissipates more energy. You don't want to lose all your energy. So you look for something better. And also you look for something more accurate.

I want to suggest a more accurate method, the trapezoidal method. Which just follows that basic principle that if I center things at halfway I'm probably going to pick up an order of accuracy. So centering it halfway I took half of backward Euler and half of forward Euler. So is this one explicit or implicit? Implicit. $u_{(n+1)}$ is still appearing here, has to move over there. So if I move it over I could say here I have, let me multiply up by Δt and bring the $u_{(n+1)}$ over here. So I'd have $u_{(n+1)}$. I'll have minus Δt over two A . All that's what's multiplying $u_{(n+1)}$. Right? When I multiply up, let me do that. Let me make it easy for your eyes. I'll multiply up by the Δt . And then I'm bringing this part, the implicit part. So that's my left-hand side. And what's my right-hand side? u_n is going over as the identity plus Δt over two A . All that's what's multiplying u_n . Good. So this is the matrix that has to get inverted at every step. Of course, in this model problem that's not expensive to do. If we have the same matrix at every step, if we have a linear time-invariant system, well we can just invert it once or factor it in into L times U once. We don't have to do it at every step. So that's, of course, very fast. But when the problem becomes non-linear we're going to start paying a price for a more complicated implicit part.

What I'm interested in right now just to sort of see these differences, there's a particular matrix A that is my model here. It's antisymmetric. Notice it's antisymmetric and that sort of goes with

conserving energy. I'm not going to-- In a first order system, that antisymmetric tells me that the eigenvalues of that matrix are pure imaginary. And so I'm going to get e^{it} . I'm going to get things that don't blow up, that don't decay. So I want to see, what's the growth factor? Suppose you want to understand, is this method good, what's its growth factor? So that will tell me about stability. It'll tell me about accuracy, too. So its growth factor, all I want to do is put everything on the right side of the equation. I want to write it in this form. So that G will be the growth matrix. And what is G ? Well, I just have to invert that. That's $I - \Delta t A$. The inverse of that, so that's the implicit part, times the explicit part. I'm not doing anything difficult here. Just seeing how would you approach to understand whether the solution to this is growing, decaying, or staying as we hope, with maybe constant energy. Because that's what the true solution is doing. The true solution, you remember, is just going around in a circle in our model problem and just oscillating forever in our model system.

What about the growth or decay or whatever of that growth factor? Well again, this is the point where I would look for eigenvalues. If I have a matrix G , and everybody recognizes that after n time steps, what matrix am I going to have? What matrix connects u_n to the original u_0 ? That matrix is? G^n . At every step I'm multiplying by a G . So with finite differences, you have powers. That's the rule. With differential equations, you have exponentials. With finite differences, finite steps, you have powers of G . So I'm interested in G^n and what tells me about that is the eigenvalues here. If I had to ask for the eigenvalues of this matrix, they would be the eigenvalues of G . So $\text{eig}(G)$, can I say, for this case.

Actually, can I just say what it would be? What are the eigenvalues of this guy? The eigenvalues of that factor are one plus Δt over two times the eigenvalues of A . That's what we're getting from here. And here, this is the inverse. So its eigenvalues will come in the denominator. This'll be one minus Δt over two times the eigenvalues of A . That's pretty good. That's pretty good. Actually this sort of shows me a lot. Well, in the case of this model example, the eigenvalues were i and $-i$. This is for the special case A equals $[0, 1; -1, 0]$. This will be one plus $i \Delta t$ over two divided by one minus $i \Delta t$ over two. And the complex conjugate for the other eigenvalue. It'd be two eigenvalues. Let me work with the one, let me take the one that's i and then there would be a similar deal with $-i$. That's our eigenvalue of G . Where in the complex plane is that number? If I give you that complex number. We're meeting complex numbers here. We're not meeting complex functions, just we have to be able to deal with complex numbers.

Actually, when I look at those two numbers, what do I see right away about them? How are they related? They're conjugates, right? Conjugates of each other. This one is in the complex plane, I would go along one, the real axis, and up Δt over two. And this one I would go down Δt over two. And that symmetry is what I, the word you used, the complex conjugate. Compare that length for the top with that length for the bottom and what do you get? Same length. So I'm concluding that in this case the magnitude of these eigenvalues of G is what? So magnitude, absolute value, is just the absolute value of that divided by the absolute value of that. It's this distance divided by this distance. And you told me already the answer is one. The eigenvalues are right on the unit circle. In some ways that's wonderful. The solution if you compute exactly will stay on the unit circle. Of course, it will not be exactly the same as the continuous solution. But the energy won't change. We're not going to spiral out. We're not going to spiral in. It's this average and it's more accurate. So trapezoidal method is like, the workhorse for finite element codes. Trapezoidal method is a pretty successful method.

What's the price? The price is this implicit stuff that you have to solve. So I should say it's the workhorse among implicit methods. And it's simple. So just to have a picture, finite element codes usually are not shooting for great accuracy. Many finite element calculations, you're happy with two or three decimal places. So we're not shooting for great accuracy. Second order is very adequate. What we don't want, of course, is to be unstable. We don't want to lose all the energy. So this is really a good method. I have to say it's a good method, but not perfect. For linear problems, for my model problem this is fine. For a real problem, a difficult problem in mechanics you might find that the energy, you might find it goes a little unstable and non-linearity tends to grab onto a little instability and make it worse. So like Professor Bathe, if you take his course on finite elements, this trapezoidal rule-- Many people have reinvented it. There's a Newmark family of methods and with special parameter, you get back this one and everybody makes that choice practically. There's just a host of finite difference methods.

I'm wondering whether you want me to tell you one more. Do you want one more finite difference method? Just to like, see what. You said yes, right? One more method. One more and then, this is a subject on its own. But just to see what else could you do. What else might you do? And let me say why you would. And Professor Bathe had to do it. In some problems he found that with non-linear problems, he found that this method, which for a perfect linear problem stays exactly one, that's great, but you're playing with fire. To be right on the unit circle, if non-linearity pushes you off then you wish you had not tried for that perfection.

So let me describe a backward difference. Let me write down, I'll call this BDF2. All these formulas have names and numbers. So this would be Backward Difference Formula second order accurate. You'll see, it goes two steps back. So here it is. $u_{(n+1)} - u_n$ over Δt . And then there's another term which gets the second order accuracy. It happens to be $u_{(n+1)} - 2u_n + u_{(n-1)}$ over Δt . And then that really is Δt . Equals $Au_{(n+1)}$. It's good practice.

That formula came from somewhere. What if we look at it. What do we see? What's the picture for a formula like that? Is it implicit or explicit first, our first question. Implicit. Because it's using, this right-hand side involves this. And if it was a non-linear equation, whatever that right-hand side is, not linear, would be evaluated at the new step and therefore would have to go back over here. It is second order accurate and I won't go through the checking on that. And is it stable? That's another question. We have to find eigenvalues here. Let me not go through all details there. It is stable. And it's slightly dissipative. It's not as dissipative as backward Euler. There you're losing energy fast. Here the eigenvalues, the thing would stay much closer to the unit circle than backward Euler. I'll just put up there so that you see.

What are other features that you see right away from this method? The fact that it involves not only $u_{(n+1)}$ and u_n , but also $u_{(n-1)}$. What does that mean? How do I get started with that method? Right? This calls to find the new u . I need the previous one and the one before that. No problem, I have them. Except at the start I don't. So it'll need a sort of special start to be able to figure out what should u_n be. It'll need a separate formula to decide. And it could use one step of backward Euler to find u_1 . And then take off, now finding u_2 from u_1 and u_0 . And then onward. So that's quite fast. More stable, good method. And you maybe can see that I could get more formulas by going back further in time. And by doing that I can get the accuracy higher. So that's good to see that particular BDF2. So that's a backward difference formula.

Oh, just to mention what's developed. So this is stable. It actually loses a little energy. So in fact now both Professor Bathe and I have studied the possibility of taking a trapezoidal step, which was a little dangerous in the non-linear case because you were playing with fire, you were right on the circle. And then put in a backward difference step to recover a little stability. And then trapezoidal backward difference. So split step. Split the step into a trapezoidal part and a backward difference part. That's actually discussed later in Chapter 2. Section 2.6 and I might come back to that. I just wanted to say that much this morning. Having got as far as forward and backward Euler. I didn't want to leave you without a better method which is the

trapezoidal method.

I could take any question. I would like to devote the second half, if you're okay to just change gear, to beginning on least squares. So this was our kind of excitement to have time dependence for a little while. But now I'm going back to steady-state problems. So they're not moving. I'm looking at equilibrium. And what I'm going to do now in the next lectures is more to see this framework that we identified once for the masses and springs, to see it again and again. Because it's the basic framework of applied math. So now I'm ready for least squares. Least squares. What's the problem? Well the problem is I'm given a system of equations $Au=f$. These f 's are observations. The u is the unknown vector. You say what's different here? It's just a linear system. What's different is too many equations. This is an m by n matrix with m bigger than n . Maybe much bigger than n .

So what do we do? We've got too many equations and no solution, no exact solution. I would say probably that what we're coming to, what we're starting on today is the most important, the application of linear algebra that I see the most. So it interests everybody. It interests engineers, scientists, statisticians, everybody has to deal with this problem of too many equations. And those equations come from measurements so you don't want to throw them away. I don't want to just throw away. I want to somehow get the best solution. I'm looking for the u that comes closest when I can't find an exact u . So that's the idea. There's no exact solution and the problem is what is the best u . And I'm going to call that \hat{u} . My favorite choice of u will be \hat{u} . So I'm going to get an equation for \hat{u} . That is my goal.

And what I'm starting here just goes all the way to, there will be weighted least squares, there will be Kalman filters. It's a giant world here of estimating the best solution when there's noise in the right-hand side. And what's the model problem? Always good to have a model problem. Let me draw a model problem. Model problem is fit by a straight line. So say $C+Dt$. Shall I use that as the--? $C+Dx$. It's got two unknowns. So n is two. But m is big. What do I have? Let me draw the picture. You've seen this often. This is the t -direction, this is the f . These are the measurements. So I measure at time $t=0$, let's say. I measure my position. That would be f_1 . At time $t=1$ I've moved somewhere, I've measured where I am. I'm tracking a satellite, let's say. So I'm tracking this satellite. Well the times don't have to be equally spaced. I'll take the next time to be three. Let's say the engine is off, this satellite. If the measurements were perfect-- Does that look too perfect to you? It's almost on a straight line, isn't it? Of course my point is that measurements, well I mean, of course in reality measurements would be close to

a straight line. I'm going to have to draw, in order for you to see anything, I'm going to have to draw a really--

Suppose f_1 is one. Suppose it starts at position one and suppose f_2 is two and this guy will be-- Where do you want me to take it? Let's see, if it was linear, what would be the--? It would be four, right? So can I take a different number? Three. Is three okay? Because five I haven't got space for. And you don't want to see pi or some dumb thing or e. So let me take three. I want to fit that data which is, we're saying, close to a straight line, I want to fit it by the best straight line. So the best straight line would go probably, I don't know what your eyes suggest for the best straight line through three points. Do you see I've got three equations, two unknowns? That's the first point to see. Somehow I'm trying to fit three things with only two degrees of freedom and I'm not going to succeed, usually. But I'm going to do my best and probably the best line goes sort of, it won't exactly go through any of them. So I'm doing the best least squares approximation. And what does that mean?

Well, what would the three equations be? What does my linear equations, my unsolvable ones, say that at time zero? So at time zero, at time one and at time three, at each of those times I have an equation $C+Dt$ should agree with-- So that $C + D$ times zero should match f_1 . At $t=1$ my line will be $C+D*1$, it should equal f_2 . And at $t=3$, the height of the line will be $C+3D$, and I would like it to go through that height f_3 . But I'm not going to be able-- If there was noise in the measurements that system, that's my unsolvable system. What's the matrix? I want to write three equations and you're getting good at seeing three equations like so. So I've a 3 by 2 matrix. And my unknown u is $[C, D]$. Those are my unknowns. And my right-hand sides are these heights, well, I decided on particular numbers, one, two and three. One, two, three. And what's the matrix? What's the matrix A that you read off when you see that system of equations? The first column of the matrix is? All ones because that's multiplying the C 's. And the second column of the matrix is the times. Zero, one, three, is that right? That multiply the D . So this is the same as that.

So here I'm in my set-up. I'll erase m equal big because m was only three, not that big. What's the best answer? What's the best u hat? The best u hat will now be C hat and D hat. The best I can do. I need some idea of what does best mean. And there is not a single possible meaning. There are many possible ways I could say the best line. One way would be to make the, well, what could a best line be? I'm going to have three errors here, right? That did not go right through the point. This did not go right through the point. They came pretty close. I've got

three small errors. e_1 , e_2 , e_3 . Those are the errors in my equations. So I will get equality when I add in the e_1 , e_2 , e_3 , the little bits that will bring it onto the line.

One idea. Make the largest error as small as I can. Minimize the maximum of the e's Try to balance them so no e, no error is bigger than the others. Look for that balance. That's a reasonable idea. But it's not the least squares idea. So what's the least squares idea? The least squares idea makes the sum of the squares of the errors as small as possible. So the least squares idea will be to minimize the sum of the squares of the errors. e_1 squared plus... e_m squared. It would be just e_1 squared plus e_2 squared plus e_3 squared. What is this? Let me began to write this in matrix-- I want to bring in the matrix here. This is the error. The error is the difference between the measurements and Au . So that's what I'm trying to make small. I'd love to make it zero but I can't. I've got more equations than unknowns. There's no two unknowns that will make all three errors zero. So I want to make the errors small. And this is the length of e squared. The length in this sum of squares method. It's a pretty good measure of the error.

Gauss was the first to apply least squares. What I'm going to do today is Gauss. Who was, by the way, the greatest mathematician of all time. And here, he was doing astronomy actually. And writing in Latin. The message got out somehow. So his idea was sum of squares. So this e is the distance between f and Au . I have to begin to write. I have to write some things. I can write some things out in detail, but then I also, at the same time, have to carry along the way I would look at it for any matrix A and any right-hand side f . So do you see that this is the error? The meaning of this double bars squared is exactly that. That it's the sum of the squares of the components. So that's where the word least squares come in.

Can I just say what's better about least squares and what's maybe a drawback. So actually this next sentence is pretty important in practice. What's better about least squares, what's really nice about least squares is well, for one thing, the equations we get for the best $[C, D]$ will be linear, will be linear equations. You may say, not surprising, I started out with a linear system and I'm going to end up with a linear system. Actually I prefer to use-- Well, might be too late. Next lecture I'm going to put b in there for the right-hand side. But I'll leave it with that for now. So good point is we'll get linear equations. The not so good point in some applications is when I look at the squares of errors, well big errors, outliers, really bad measurements have a big influence on the answer because of getting squared. So if I have ten readings that are very accurate but then in an eleventh reading that is way off and I don't know it and if I don't

realize that that's way off, then that eleventh error will-- It's like having a whole lot of points close to a line and then another point way off. That will have a significant effect on the best line. So you might say too great an effect. Depends on the application.

I just had to say before starting on least squares, as always, there are advantages and disadvantages but the advantages are very, very great. So it's an important idea here, least squares. I'm ready now to ask for the equation for \hat{u} . So the equation for \hat{u} is the u that minimizes here. So we have touched on minimizing quadratics. This is squares. I could expand that out as $f - A^T u$ times $f - Au$ just to see another way to write it. The length squared of a vector is always the transpose-- Its inner product with itself. And I could split this out into all these different terms. I would have then, some quadratic expression to minimize. In other words, let me jump to the answer. Let me jump to the equation for the best u . And then come back to see why. Because you must see what that equation is. It's the fundamental equation of, this might be called linear regression, fitting data. You're just constantly doing it. So what is the equation for the best \hat{u} ? Can I put it here? This'll be equation that we get to. It'll be $A^T A$. You're not surprised to see $A^T A$ up here. First of all because this is 18.085 and also because this A is rectangular and when you have rectangular matrices, sooner or later $A^T A$ comes up. So that's the matrix. And then the right-hand side is $A^T f$. So that's the key equation for least squares. That's the central equation of least squares.

And let's just see what it looks like. You could say the way I arrived at it, I mean the short way is this is an equation that I can't satisfy. I multiply both sides by A^T and now this is the equation for \hat{u} . And what that did was kind of average out the m equations. Because how many equations do I now have? A as m by n . What's the shape of $A^T A$? Everybody's on top of that, right? The shape of $A^T A$ is square, n by n . Because A^T is n by m . n by m times m by n leaves us an n by n . So we've averaged the m equations that were too many to get n equations. And of course this is what it should be, n by m , m by one. So it's n by one. It's a good right-hand side. That's the equation of least squares.

That's the equation I want to explain, understand and solve. Actually why don't we solve it for this particular problem just to see the whole thing for this example. Just to do it. So there is A . Here is u . And here is f . what shall I call these? They're mostly called the normal equations. That's one possible word for the key equation of least squares, the normal equations. Can you tell me these matrices $A^T A$? And \hat{u} I know. That'll be the best C and the best D .

And over here can you compute $A^T f$? If I write A^T above it, will that help you do these multiplications? Let me just do. So there was the matrix A . Let me write A^T above it. So it has a row of ones and then a row of twos. So what shape is the matrix? The $A^T A$ matrix. It's going to be that A^T times that A . The size will be two by two, right? Two by three times three by two, it's averaging out to get me a two by two matrix. What's the first entry of this? Can you do $A^T A$ just so we see this matrix. Three. And off the diagonal? Four. And here? And you knew it would be symmetric. And here? Ten. And tell me $A^T f$ while we're at it. So that's just a vector. If you multiply that by the right-hand sides, looks like a six. Is that right? 11, maybe. Is that right? Two, nine making 11, yeah. So those are the numbers.

I can't write those numbers, write $A^T A$, without asking you to tell me one more time what kind of a matrix have I got here? It's symmetric positive definite, right? We know that's going to be. And we see it. Our test for positive definite might be the determinant, that one by one determinant is three, that two by two determinant is 30 minus 16, 14, positive. We got a good problem here. And I could solve for \hat{C} and \hat{D} . I see the numbers are not coming out fantastically but they would produce a line that would, I'm pretty sure, it would be, this looks optimal to me. If I rotated any, I'm going to make things worse. I think it would look like that. So that's the system that you end up with. But why? You really have to understand, where did this equation come from. Where did it come from? It's worth understanding, this least squares stuff. So I'm going to try to draw a picture that makes it clear where that equation comes from. So what am I doing here? $Au=f$. Start there. And the particular A was, I'll even copy the A . It was $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \end{bmatrix}$, multiplied u to give me f as $[0, 1, 3]$. But of course, I couldn't solve it because I don't have enough unknowns. What's the picture? Everybody likes to see what's happening by a picture as well as by algebra. So the picture here is I'm in three dimensions and I have a vector $[0, 1, 3]$. So 0 in that direction, one in that, three up. So somewhere there is my f . Now I'll put in C, D here. What is the equation asking me to do? Which actually, I won't be able to do because I can't solve it. But the equation, how do we see a system of linear equations? If I have a system of linear equations I'm looking for numbers C and D so that C times column one plus D times column two gives me that. That's how I think of a system of equations. A combination of the columns.

Tell me, what vectors do I get if I take combinations of the columns. Well, if I took the combination $C=1, D=0$ I would just get the first column. So that's a candidate. $[1, 1, 1]$, I don't know where that is. Wherever $[1, 1, 1]$ might be. I'm not too sure where to draw $[1, 1, 1]$. I want

to go one there, one there and one there. Damn. Let's define that to be the vector $[1, 1, 1]$ right there. Wait. You let me put that up there and I didn't mean to, right? f should be zero, what, sorry? f was $[1, 2, 3]$, yeah. Damn! Don't let me make mistakes. These mistakes are permanent if you let them slide by. That's it, same point. I didn't have the point right in the first place so now it's just perfect. There it is. Before of course, if I had $[0, 1, 3]$, I could've solved the equation. But with $[1, 2, 3]$ I can't. Here's the situation. This vector is not a combination of those two. Because the combinations of two vectors, what's the picture? If I try to draw, if I look at all combinations of two vectors, $[1, 1, 1]$ which is that vector, $[0, 1, 3]$ which is maybe this vector, let's just say. If I take the combinations of these two column vectors, what do I get? Now this is for everybody to know. If I take the combinations of two vectors here in three-dimensional space I get a plane. I get the plane that contains those vectors. So this I could call the column plane or the column space. This is all combinations of the columns. That's the same thing as saying this is all the f 's that have exact solutions.

So let's just see this picture. This particular right-hand side is not in the plane, right? That's my problem. This particular vector f points out of the plane. But if I change it a little, like if I change it to $[1, 2, 4]$. Do you see that that would--? What's different now that I've changed it to $[1, 2, 3]$ for a moment? What's different about $[1, 2, 4]$? Where is $[1, 2, 4]$ in my picture? Do you see what's great about $[1, 2, 4]$? It is a combination. Right? $[1, 2, 4]$ is a combination, with $C=1$, $D=1$. It would satisfy the equation. So where is $[1, 2, 4]$ in my picture? It's in the plane. The plane are the heights that do lie on a straight line. So the plane are all the ones that I can get exactly. But this vector, these observations, $[1, 2, 3]$, I couldn't get exactly. So let me-- In 30 seconds or less, let me tell you the best thing to do. Or let you tell me the best thing to do. I have a right-hand side that's not in the plane. I can get my straight lines correspond to vectors, right-hand sides that are in the plane. So what do I do? I project. I take the nearest point that is the plane as my right-hand side. I project down. And it's that projection that's going to lead us to the equation that I'm shooting for, $A^T \hat{u} = A^T f$. This comes from projecting f down into the plane where straight lines do work exactly. So there's an error here that I can't deal with. And there's a part here, the projection part, that I can deal with. This is important and it's fun and we'll come back to it Monday. Thanks for patience.