PROFESSOR: This is a topic that's sometimes called calculus of variations. That's the classic name for it. And-- Cookie, are you causing trouble? OK. Calculus of variations, but all it really is the same idea of derivatives where-- so we're going to have f of u, a plus du, minus f of u. That's our df. Is going to be-- I'm going to call it f prime of u, du.

It's the same formula, except that now where u is in a vector space of functions. And usually f and-- and usually f is a scalar. F of u is just a real number.

So we're to take a function, map it to a real number. And this is one of the things when I teach 1806, I come back to, periodically, that vectors are not just column vectors. They're even matrices. You can even think of functions as a vector. Because you can take two functions, and you can add them. You can take two-- and get another function.

You can take two functions and multiply them-- a function and multiply by 2 and get another function. And so a lot of the things you can do in linear algebra for column vectors, you can do for functions. You can talk about linear operators on functions. Like a derivative is a linear operator on function.

You can talk about orthogonal functions, and dot products of functions, and Gram-Schmidt on functions, and eigenfunctions. If you do quantum mechanics, you do lots of things with eigenfunctions. So it's the same-- all the same kinds of concepts.

You have to be a little bit careful with functions. They can do crazy things that column vectors can't. Like they can oscillate infinitely fast, or they can blow up, or do things like that. So if you want to do rigorous math with functions, you spend a lot of time nailing down the space of functions that are not terribly behaved. I'm going to do that. I'm just going to be implicitly-- I'm going to write down an integral of functions-- of a function. I'm going to implicitly assume that we're dealing with functions where that integral is defined. So things like that.

So let me just do an example. All right, so let's do u of x is a function from 0 to 1. X is in 0 to 1. Two real numbers. It's just here's x equals 0, x equals 1, and u is some function, u of x. And let's suppose that f of u-- I need to write down something that takes in a function and gives me a number. And there are a few ways to do that. But the one that I want to deal with is just integral, 0 to 1, of say, sine of u of x, dx. Let's just start with.

That takes in a function, gives me a number. Technically, it doesn't work for all functions. You could have functions where this integral doesn't exist or something like that. So implicitly, I'm limiting myself to the space of functions where the integrals work. And what I want to do is find out what is the-- what is the gradient of f. And what is f primed also with respect to u, of course. So now with respect to a function.

And it's going to be exactly the same definition for it to take a derivative. To take a derivative, I just need to linearize. I need to ask how does f change for a small change in the input, linearize it. And then to get-- one step further, to get a gradient I'm going to need a dot product.

Before we do the gradient and the dot product, let's just linearize. So we're going to take df. We take f of u. Remember u is a function. By the way, this is sometimes called a functional. It's something that takes in a function and gives you a number.

And I'm going to-- and here I need to-- f of u plus du, minus f of u. And the key thing to remember is that this is a function u of x. That's a terrible x. And this du here, this is also a function. You can think of it as a small function, du of x.

And as I said last lecture, to get some notion of small, we're going to need a norm of a function. But pretty soon, I'm going to have dot products of functions So. I'll have norms as well. So let me sweep that under the rug. But it's just like a little perturbation. So you're taking u of x-- u of x, and you're adding some arbitrary little change to this. This is a u plus du of x.

So you're going to change that function by a little bit anywhere you want. It has to be small. And so we can drop terms that are like du squared. And so what is this so this? This is integral from 0 to 1. And I said f was sine of ud-- of xdx. So this is looks like sine of u of x, plus du of x, minus sine of x, dx. And I'm just plugging that into my f. My f was integral sign u.

And I'm going to ask, how much did that integral change for a small change in u, and I want to linearize it. So I want to write this as-- drop every term that goes like du squared or higher. And, of course, the point is all only need to do is linearize this, linearize the integrand. That thing, sine-- sorry-- and this is sine of x. This is sine of u of x.

OK, so I want to linearize this. What is sine? So sine-- sine of u plus du, that's equal to sine of u plus the derivative, which is cosine of u times du. This is just-- du is a function, but du of x is just a number. At each x, I have a small change in u of x. So this is just ordinary 18.01, linearization, or from lecture one of this class. Or this could be-- yeah, anyway.

So this is just a number, ordinary linearization. And then, of course, this term here cancels that term. So what this is is integral 0 to 1 of what's left-- cosine of u of x, du, dx. And this is exactly-- this is a linear operator. This is f primed of u adding on du. Sorry, and again, I should put a du of x here.

Is that clear? If I plug in any function here, any change in my u, this is clearly linear in that. If I double this, it doubles the integral. If I add two things here, it adds two things in the integral. So this is a linear operator. We're done. That takes du in, and it gives you out a number. this is my df.

So once you get used to this as being the definition of a derivative, you can always fall back on this. When u is some crazy object, matrix, function, whatever, as long as it's a vector space, or technically a Banach space-- it has to have a norm-- then you can define the derivative as the linear operator that gives you the first-order change.

So, as I said, technically we're dropping terms of little o, du, terms that go to 0 faster than du here. So to make that precise, you need to have a norm. But anyway, this is a function that takes a vector-- by function, I'm thinking of it as a vector space-- a vector, and it gives you out a scalar. So I should be able to write that linear operator, this linear form, as a dot product with something.

So for that, also, I want a dot product, which will give me a norm. So that's just to remind you-- hopefully, you've seen this before in 1806 or some other linear algebra-- dot products are functions. So-- yes?

STUDENT: I have a quick question. One takeaway that I have so far is that a derivative is something that lives in the dual space or the thing, that u. Could the input be a function?

PROFESSOR: Yeah. Well, if the output is a scalar. Yes, so this is exactly something in the dual space. Physicists would call this a bra.

STUDENT: But then, if this is the function, I would expect that the derivative be something like distribution and measure? Or should I think of this as like the cosine of-- could it be like the density of it?

PROFESSOR: Yeah, this is a measure. Because-- exactly, this is the density of that measure. Yeah.

STUDENT: Thank you.

PROFESSOR: Absolutely. This takes the measure of that function. Exactly, yeah. So there's all sorts of-- you can think of this as a linear form, or something in the dual space, or a covector, this vector space. But I want to say a little bit informal in terms of those abstractions here. And as I said, technically speaking, we have to restrict ourselves to not all possible functions here. They have to be functions where this integral is defined.

So they can't be functions that blow up, or are not measurable, or something like that. But that's something that pure mathematicians worry about that physicists and engineers don't. Because the functions-- basically, if this is a physical context, implicitly you're dealing with functions where these things are defined. Like you only really worry about ruling out those crazy functions if you're doing pure math that's unmoored from reality.

So dot products are functions. So if you take a dot product with two column vectors, the standard-- there are many dot products, but the standard one, the Euclidean one, is you just multiply the components and add them up.

So if you think of two functions-- u of x and v of x on x in 0 to 1, then the Euclidean dot product-- the obvious dot product, the Euclidean inner product, should be to multiply the components and add them up. So the components, if you think of these as column vectors-- they're like infinite dimensional column vectors. The components are the values at each point.

So basically, what you want to do is multiply u of x times v of x and add that up for all the x's. And the analog of adding it up is an integral. That's our u dot v. So we can-- and, of course, this is not the only possible dot product, but this is the most obvious one-- the first one that you typically reach for.

So once you have a dot product, then, of course, we have a norm. So if you want the norm of u, that's just the square root of the integral from 0 to 1 of u of x, square dx. I guess I don't need absolute values here because everything is real right now. If they were complex, I'd be putting complex conjugates over one of these. So let me remove my absolute values there, just for-- I don't need them.

So this is-- and once we have the norm, then that's the precise sense in which we're dropping terms that go like du squared. So we're going to drop terms that go like the norm of du squared and higher, or actually, anything higher than linear.

And once you have this, now we can compute-- we can find the gradient. Because what? Was our df? Our df, which was f primed of u acting on due, we said that was integral from 0 to 1 of cosine of u of x, times du of x, dx.

We want to think of this-- just our generic definition of the gradient is when you have something that takes a vector in and scalar out, the first derivative had better be the dot product of the gradient with du. And just by inspection, this is-- here, this is our definition of a dot product. So what is the gradient? What would I take the dot product with du of to get this? Cos of u-- exactly. So it's a function. It's just cosine of u of x.

By the way, in this context, this is called-- this is sometimes also called a Frechet derivative. Frechet derivative technically refers just to this general notion of if you have any Banach space, the Frechet derivative is the linear operator that does this. But most often, you see this coming up in the context of vector spaces of functions. So that's where they use that terminology. For everything else, people just say the derivative. But it really is just the derivative, but that's it. This is a gradient.

So let's do a more complicated case. So let's do another example, a more tricky example.

Suppose that-- suppose that f of u-- so we're still going to be functions from 0 to 1. Suppose it's the integral from 0 to 1 over the square root of 1 plus u primed of x squared, dx. Where this is just the derivative, the ordinary du, dx, the 18.01 derivative.

So this takes a function in, gives you a number out. Clearly, we have to restrict ourselves to differentiable functions now. And what is this? Does anyone recognize this? Can you describe what this quantity is geometrically? So if I have a function u of x, and it goes-- here is u your x-- what is this quantity that I've just computed here?

From first year calculus, there's an important name of this. Yeah? It's the length. It's the arc length. It's the length of the curve. The arc length of u of x from x equals 0 to x equals 1. That's what this quantity is.

If you multiply through by dx, this is basically square root of dx squared plus dy squared. If you take-- each little line segment here it's like the length of that little line segment, and you're adding them up.

So now let's think about the derivative. So now let's take f of u. And we're going to add a little change to u. Subtract f of u. So it's always good to just fall back on the definition, the most basic definition if we're in doubt of what to do. And we're going to just add a little change.

And what's this going to be? So let's just-- it's the integral of from 0 to 1, the square root of 1, plus u plus du, primed, squared, minus the square root of 1 plus u primed and squared. So, again, we just think-- so now we have to Taylor expand this. We have to assume that du is now a differentiable thing, and its derivative is getting smaller and smaller as well. So I can do this kind of expansion-- first-order expansion.

So what is this? Well, it's the-- if I have-- let's see, if I have 1, plus v plus dv, squared, what's the Taylor expansion of that? 1 plus v squared, that's the first term. And the second term is the derivative with respect to v, which ends up being v over square root of 1 plus v squared, times dv. I went through that a little bit. But this is just 18.01-- this is just 18.01, so plus.

This is this function for dv equals 0. This is the derivative of this. So when we take the derivative of the square root, it's 1/2, 1 over the square root. But then there's a derivative of this, which gives you a 2v on top, so the two cancels the 1/2.

So this just-- so just from calculus and expanding everything to first order in du and also du primed-- so dropping anything that goes like-- anything with a du, whether you take a derivative or not, squared, we're going to just drop. This is the integral of that second term, which is u over square root of 1 plus u squared, times-- and du, that's primed. and times du primed dx.

So this is a linear operator, but it's acting on du primed, not on du, per se. The derivative is a linear operator on du, but I'd like to write it directly in terms of du. Because I'd like to write it eventually as-- take a gradient, write it as a dot product of something with du. So what I really want to do is write this as a dot product of a gradient-- dot du. And somehow, I have to get a du and not a du primed.

So we just saw this in the first hour. If you have an integral, and you have a term that has a derivative, and you don't want the derivative, what do you do? What can you do in first year calculus to get rid of a derivative on a term in an integral? Integrate by parts.

So what do we do? So that whole thing, let's see df. There is a boundary term that looks like u primed over square root of 1, plus u prime squared times due at the boundaries, 0, 1. And then you have a minus sign, and an integral from 0 to 1 of-- we take this whole thing primed, du, and then times dx. All the u's and the du's are functions of x.

So to go further, this is it. This is the linear operator. This is now our-- this is our f prime of u acting on du. And this is a linear operator acting on du. And there's no du primes here. This is nice. But it has these-- these boundary terms are kind of annoying.

But usually, in the context where I'm interested in this, I'm actually going to set some boundary conditions. So in particular, suppose I want to fix the endpoints of my u. And I want to find the curve that has the minimum distance from one point to another. Of course, that's going to be a straight line. But suppose I didn't know that. This is a simple problem where you know the solution ahead of time, but the more complicated things, you won't know that.

So I want to-- suppose we want to fix-- to fix-- let's just even say u of 0 equals u of 1, equals 0, just for simplicity. And now we want to find the u that minimizes the arc length f of u. We know it's a straight line from one point to the other, but let's pretend we don't know that.

So what we want-- if you want to minimize something, what you want is for the gradient to be 0. So we're going to take a gradient set it equal to 0 in a minute. But if we fix the endpoints, we're going to want du also at 0 or 1 to be 0. If we fix them at 0-- because if we start with our vector space of u is at our 0 at the endpoints, and we want to look at not all possible perturbations, but perturbations that keep those end fixed. Or if I kept the ends fixed at some other point, still I would want to look at not all possible perturbations, I'd want to look at the perturbations that keep the ends fixed. So the u is 0 at the ends, so this keeps the end points fixed.

And then what happens is that boundary term disappears. What was it-- u primed, over square root of 1 plus u prime squared, du taken at 0, 1. This is now just 0 because we fixed the endpoints.

So now df is just integral of 0 to 1 of u primed, over square root of 1 plus u prime squared, primed, times du, dx. So our gradient is just this thing here. It's the thing we're taking with the dot product of du with to get the solution. So it's just u primed, over square root of 1 plus u primed squared, primed. Does everyone see that?

So this df had better be the dot product of grad f with dot grad u. I integrated by parts. So I turned my u primed into a du. I set boundary conditions to fix the boundary so the boundary terms disappeared. What I have left is just a dot product of this with the u. So I can think of this as the gradient.

And if you're at a minimum, it's the same-- for the same reason as in 18.01, and 18.02, and so forth, at a minimum, the gradient has to be 0. Because if it's not 0, you can move in that direction and make it bigger, or move in the opposite of that direction and make it smaller. So this has to equal 0 at a minimum.

And so now let's just-- I need to work out-- OK, now, it's a little bit of a calculus exercise. What is the derivative of this thing? OK, so now, it's quotient rule. So this is going to be quotient rule.

So on the quotient rule, on the bottom, you get that denominator squared. So you get 1 plus u prime squared, the square root squared. On the top, you get the derivative of the top, which is u double primed times the bottom, square root of 1 plus u primed squared. And then minus the top times the derivative of the bottom. And we just did that. The derivative of the derivative of that square root is, in fact, u primed, over square root of 1 plus u primed squared, I think. Hopefully, I'm not making any mistakes here.

OK, so now, if I want to put this over-- let's simplify this a little bit. This is 1 plus u prime squared. I'm just going to-- I'm going to put a 3/2 on the bottom. So I'm going to pull this term and put it on the bottom. So I'm just going to multiply and divide by that square root. And what you get is a u primed, times 1 plus u prime squared, minus u prime squared. I'm just simplifying this a little bit.

STUDENT: That could be used on the second one.

PROFESSOR: The second one? No, Because this came from u primed times u primed. This one is--

STUDENT: No, no. The u prime that is--

PROFESSOR: There? That came from this square root. I multiplied the top and the bottom by a square root of 1 plus u prime squared.

STUDENT: OK, and then the second term.

PROFESSOR: The second term--

STUDENT: The u prime.

PROFESSOR: What's that?

STUDENT: The u prime actually stays in the form--

PROFESSOR: So wait, this term? Are we talking about the right term or the left term?

STUDENT: We're talking about the term on the right.

PROFESSOR: This one here?

STUDENT: Yeah, we're talking about the u prime squared because if it is 1 plus u prime squared, you would--

PROFESSOR: Sorry, are you saying this term is wrong? Did I make a mistake?

STUDENT: Yeah, I think.

PROFESSOR: So was this term correct?

STUDENT: I don't think it is u prime.

PROFESSOR: Oh, yes. Thank you. Oh good. Good, because that's going to screw things up. OK, so-- yes, OK. So good, sorry. Yes, you're right. The top is not even that. It's u double primed-- good. OK, I knew something was supposed to cancel here, good. It's actually u double primed times u primed. Right?

OK, good. Because it's the derivative of the u primed squared. It gives you a 2 u double primed, u primed. The two cancel the derivative of the square root. Good. Good, I'm missing a u double primed there. Good, I knew it was supposed to simplify. I was a little worried about that.

So now this term cancels this term. And what you get is just u double primed, over 1 plus u primed squared to the 3/2. And this is supposed to equal 0, we said. So the only way it can 0, is if u double primed equals 0. And that's a straight line. So what we just derived is that the shortest distance between two points is a straight line. Of course, we knew that.

But let's do a generalization. OK, so let's say f of u is an integral from a to b of some other function. Let's call it capital F of-- it may depend on u. It may depend on u primed. It may depend on x. This is a functional.

And then what is df? So df is f of u plus du, minus f of u. So now what are we going to have? So there is going to be a term that looks like partial f, partial u-- this an ordinary 18.01 partial derivative-- I'm assuming u is a number-- I guess it could be a vector or something like that, but let's just keep it simple-- times du.

There is also a term that looks like partial f, partial u primed, times du primed, times dx. And this sometimes confuses people. It's like how can I-- if I take derivatives with a u, keeping u primed fixed, because obviously, u and u are related. It really means the derivative of f with respect to the first argument. This is the derivative of f respect to the second argument. And then you plug in u and u primed afterwards.

So it's just-- it's just doing exactly the same. And you're doing much like this in homework, but I'm skipping a couple of steps here. This change-- there's a change that comes from the first term, like we did on the first example, a sine of u. There's a change that comes from the second term, like in the second example. And you have to add them up.

But now you integrate by parts. And on the du primed term, because we don't like having du primes, and so you get a partial f, partial u primed, times du at a, b. And then we have a plus an integral a to b. And there is a term that looks like df, du. And there's another term that looks like, with a minus sign, df, du primed, primed, multiplying du times dx. Again, du is-- my small change is a du of x. I'm just suppressing all of this.

And this term, we're going to set to 0 if the endpoints are fixed-- of u are fixed. So if we keep u fixed at the endpoints, then we're going to only allow changes, du's, that are arbitrary functions, but at the endpoints, they have to be 0. And so what you get is this term here is the gradient.

This is a grad f. This is a dot product with due. And so I can write this out. So this is-- so grad f is df, du, minus df, du primed, primed. And we set this equal to 0 at an extremum, at a minimum and maximum. And the result, this is-- we set this equal to 0, this is some differential equation-- equation in u.

It involves-- and it's actually a second order. Because you started out with u and u primed in your f, but when you take this derivative here, I'm going to get some u double primed terms. Does anyone know the name of this differential equation? Has anyone seen this before? Euler-Lagrange equation-- so these are exactly the Euler-Lagrange equations.

Where did you see Euler Lagrange equations before?

STUDENT: Optimal control.

PROFESSOR: Optimal control. But you also famously see them in advanced mechanics, like 806 or-- I don't know what the number is these days. Where this is-- there's something called the least action principle. And you can write down the equations of motion as minimizing some integral. And you get these Euler-- instead of Newton's laws, you get-- as a different way of getting Newton's laws, you get these Euler-Lagrange equations. They're really nice because they let you write down basically Newton's laws in really weird coordinate systems pretty easily.

Yeah, so there's a famous example of the Euler-Lagrange equation that everyone gives, which is called the brachistochrone problem. Has anyone heard of the brachistochrone problem? We have one person-- a couple people. So this is a famous problem from, I think, the 18th century.

So suppose you have a mass falling down a slope, just sliding down a curve with gravity. So it's going down this curve, like a ramp with no friction. It just slides down. And you're starting here, and you're ending up here. And the question is, what shape should this ramp be to get down as quickly as possible? Should it be a straight ramp? Or should it be like this?

You can pretty easily see that it shouldn't be like this because then it'll start out going very, very slowly and then go-- a straight ramp, or should it go down very steeply and then flatten out? If it goes down very steeply, it'll get going-- get a good head of speed pretty quickly, but then it has a longer distance. The straight line would be the shortest distance, but it's not the highest speed.

And so it was a famous problem to find-- called the brachistochrone problem-- to find the curve of this ramp that minimizes the time. And it's exactly an integral. You can write down this. And you write down a differential equation. You can solve it to get this curve. Unfortunately, the differential equation you get is rather nasty. It is solvable in closed form, but it's nothing I would ever assign as a homework problem in 1803, for example.

So it turns out that the minimum shaped curve is called a cycloid. It's the curve that if you-- if you have a wheel rolling, and you attach a pen to one spot in the wheel, and as it rolls, you mark out the curve that wheel traces out, that's the shape that you get. So it's a famous-- so I'll even write that down, to google example-- google the brachistochrone problem-- brachistochrone problem. That's one example.

Another example would be google the principle of least action. That's how you get the Euler-Lagrange equations in mechanics. So a great example. But it lets you turn these things into differential equations.

But what you're really doing is just taking the gradient and setting it equal to 0. And it just you have to understand functions are a kind of vector. I can still talk about linear operators on functions. I can talk about dot products of functions.

So this linear operator is exactly for the-- the derivative just is the linearized change in the output in this integral. And if I do a little bit of algebra with the integration by parts, I can write that as an integral of something times du. That's a dot product in a generalized sense. So this thing is a generalized gradient that you set equal to 0 if you want a minimum or maximum.

Any questions? And I'll put up some references as well on both what Frank talked about-- we have some-- there's some papers, and other reviews, and things on I-joint methods for ODEs. And there's lots of papers and books on Euler-Lagrange equations. And the traditional topic for deriving these Euler-Lagrange equations is calculus of variations. That's what it's called, or variational calculus.

But it really is just taking that-- now that we've generalized the notion of derivative, it's the same notion. It's not nothing really new. It's just applied to a new vector space. There's a beautiful, beautiful book by Gelfand and Shilov that I'll put up a link to on calculus of variations. I really like. It's one of those things that starts out-- Chapter 1 is accessible to a freshman. And then it gets more and more advanced as it goes along. But it's a very, very slim book, like 150 pages or something like that.