

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](http://ocw.mit.edu).

OK, so anyway, let's get started.

So, the first unit of the class, so basically I'm going to go over the first half of the class today, and the second half of the class on Tuesday just because we have to start somewhere.

So, the first things that we learned about in this class were vectors, and how to do dot-product of vectors.

So, remember the formula that  $A \cdot B$  is the sum of  $a_i$  times  $b_i$ .

And, geometrically, it's length  $A$  times length  $B$  times the cosine of the angle between them.

And, in particular, we can use this to detect when two vectors are perpendicular. That's when their dot product is zero. And, we can use that to measure angles between vectors by solving for cosine in this.

Hopefully, at this point, this looks a lot easier than it used to a few months ago. So, hopefully at this point, everyone has this kind of formula memorized and has some reasonable understanding of that.

But, if you have any questions, now is the time.

No? Good.

Next we learned how to also do cross product of vectors in space -- -- and remember, we saw how to use that to find area of, say, a triangle or a parallelogram in space because the length of the cross product is equal to the area of a parallelogram formed by the vectors  $a$  and  $b$ .

And, we can also use that to find a vector perpendicular to two given vectors,  $A$  and  $B$ .

And so, in particular, that comes in handy when we are looking for the equation of a plane because we've seen --  
So, the next topic would be equations of planes.

And, we've seen that when you put the equation of a plane in the form  $ax + by + cz = d$ , well, in there is actually the normal vector to the plane, or some normal vector to the plane.

So, typically, we use cross product to find plane equations. OK, is that still reasonably familiar to everyone? Yes, very good.

OK, we've also seen how to look at equations of lines, and those were of a slightly different nature because we've

been doing them as parametric equations.

So, typically we had equations of a form, maybe  $x$  equals some constant times  $t$ ,  $y$  equals constant plus constant times  $t$ .  $z$  equals constant plus constant times  $t$  where these terms here correspond to some point on the line. And, these coefficients here correspond to a vector parallel to the line.

That's the velocity of the moving point on the line.

And, well, we've learned in particular how to find where a line intersects a plane by plugging in the parametric equation into the equation of a plane.

We've learned more general things about parametric equations of curves. So, there are these infamous problems in particular where you have these rotating wheels and points on them, and you have to figure out, what's the position of a point? And, the general principle of those is that you want to decompose the position vector into a sum of simpler things. OK, so if you have a point on a wheel that's itself moving and something else, then you might want to first figure out the position of a center of a wheel than find the angle by which the wheel has turned, and then get to the position of a moving point by adding together simpler vectors.

So, the general principle is really to try to find one parameter that will let us understand what has happened, and then decompose the motion into a sum of simpler effect.

So, we want to decompose the position vector into a sum of simpler vectors. OK, so maybe now we are getting a bit out of some people's comfort zone, but hopefully it's not too bad. Do you have any general questions about how one would go about that, or, yes? Sorry? What about it?

Parametric descriptions of a plane, so we haven't really done that because you would need two parameters to parameterize a plane just because it's a two dimensional object. So, we have mostly focused on the use of parametric equations just for one dimensional objects, lines, and curves.

So, you won't need to know about parametric descriptions of planes on a final, but if you really wanted to, you would think of defining a point on a plane as starting from some given point.

Then you have two vectors given on the plane.

And then, you would add a multiple of each of these vectors to your starting point. But see, the difficulty is to convert from that to the usual equation of a plane, you would still have to go back to this cross product method, and so on. So, it is possible to represent a plane, or, in general, a surface in parametric form.

But, very often, that's not so useful.

Yes? How do you parametrize an ellipse in space? Well, that depends on how it's given to you. But, OK, let's just do an example. Say that I give you an ellipse in space as maybe the more, well, one exciting way to parameterize an ellipse in space is maybe the intersection of a cylinder with a slanted plane. That's the kind of situations where you might end up with an ellipse.

OK, so if I tell you that maybe I'm intersecting a cylinder with equation  $x^2 + y^2 = a^2$  with a slanted plane to get, I messed up my picture, to get this ellipse of intersection, so, of course you'd need the equation of a plane.

And, let's say that this plane is maybe given to you.

Or, you can switch it to form where you can get  $z$  as a function of  $x$  and  $y$ . So, maybe it would be  $z =$  I've already used  $a$ ; I need to use a new letter.

Let's say  $c_1x + c_2y + d$ , whatever, something like that.

So, what I would do is first I would look at what my ellipse does in the directions in which I understand it the best.

And, those directions would be probably the  $xy$  plane.

So, I would look at the  $xy$  coordinates.

Well, if I look at it from above  $xy$ , my ellipse looks like just a circle of radius  $a$ . So, if I'm only concerned with  $x$  and  $y$ , presumably I can just do it the usual way for a circle.  $x = a \cos t$ .

$y = a \sin t$ , OK? And then,  $z$  would end up being just, well, whatever the value of  $z$  is to be on the slanted plane above a given  $xy$  position. So, in fact, it would end up being  $ac_1 \cos t + ac_2 \sin t + d$ , I guess. OK, that's not a particularly elegant parameterization, but that's the kind of thing you might end up with. Now, in general, when you have a curve in space, it would rarely be the case that you have to get a parameterization from scratch unless you are already being told information about how it looks in one of the coordinate planes, this kind of method. Or, at least you'd have a lot of information that would quickly reduce to a plane problem somehow. Of course, I could also just give you some formulas and let you figure out what's going on.

But, in general, we've done more stuff with plane curves. With plane curves, certainly there's interesting things with all sorts of mechanical gadgets that we can study.

OK, any other questions on that? No?

OK, so let me move on a bit and point out that with parametric equations, we've looked also at things like velocity

and acceleration. So, the velocity vector is the derivative of a position vector with respect to time.

And, it's not to be confused with speed, which is the magnitude of  $v$ . So, the velocity vector is going to be always tangent to the curve.

And, its length will be the speed.

That's the geometric interpretation.

So, just to provoke you, I'm going to write, again, that formula that was that  $v$  equals  $\hat{T} ds dt$ .

What do I mean by that? If I have a curve, and I'm moving on the curve, well, I have the unit tangent vector which I think at the time I used to draw in blue.

But, blue has been abolished since then.

So, I'm going to draw it in red. OK, so that's a unit vector that goes along the curve, and then the actual velocity is going to be proportional to that.

And, what's the length? Well, it's the speed.

And, the speed is how much arc length on the curve I go per unit time, which is why I'm writing  $ds dt$ .

That's another guy. That's another of these guys for the speed, OK?

And, we've also learned about acceleration, which is the derivative of velocity.

So, it's the second derivative of a position vector.

And, as an example of the kinds of manipulations we can do, in class we've seen Kepler's second law, which explains how if the acceleration is parallel to the position vector, then  $r \times v$  is going to be constant, which means that the motion will be in a plane, and you will sweep area at a constant rate. So now, that is not in itself a topic for the exam, but the kinds of methods of differentiating vector quantities, applying the product rule to take the derivative of a dot or cross product and so on are definitely fair game.

I mean, we've seen those on the first exam.

They were there, and most likely they will be on the final. OK, so I mean that's the extent to which Kepler's law comes up, only just knowing the general type of manipulations and proving things with vector quantities, but not again the actual Kepler's law itself. I skipped something.

I skipped matrices, determinants, and linear systems. OK, so we've seen how to multiply matrices, and how to write linear systems in matrix form. So, remember, if you have a  $3 \times 3$  linear system in the usual sense, so, you can write this in a matrix form where you have a  $3 \times 3$  matrix and you have an unknown column vector. And, their matrix product should be some given column vector.

OK, so if you don't remember how to multiply matrices, please look at the notes on that again.

And, also you should remember how to invert a matrix.

So, how did we invert matrices? Let me just remind you very quickly. So, I should say  $2 \times 2$  or  $3 \times 3$  matrices. Well, you need to have a square matrix to be able to find an inverse.

The method doesn't work, doesn't make sense.

Otherwise, then the concept of inverse doesn't work.

And, if it's larger than  $3 \times 3$ , then we haven't seen that.

So, let's say that I have a  $3 \times 3$  matrix.

What I will do is I will start by forming the matrix of minors.

So, remember that minors, so, each entry is a  $2 \times 2$  determinant in the case of a  $3 \times 3$  matrix formed by deleting one row and one column. OK, so for example, to get the first minor, especially in the upper left corner, I would delete the first row, the first column.

And, I would be left with this  $2 \times 2$  determinant.

I take this times that minus this times that.

I get a number that gives my first minor.

And then, same with the others. Then, I flip signs according to this checkerboard pattern, and that gives me the matrix of cofactors. OK, so all it means is I'm just changing the signs of these four entries and leaving the others alone. And then, I take the transpose of that. So, that means I read it horizontally and write it down vertically.

I swept the rows and the columns.

And then, I divide by the inverse.

Well, I divide by the determinant of the initial matrix. OK, so, of course, this is kind of very theoretical, and I write it

like this. Probably it makes more sense to do it on an example. I will let you work out examples, or bug your recitation instructors so that they do one on Monday if you want to see that.

It's a fairly straightforward method.

You just have to remember the steps.

But, of course, there's one condition, which is that the determinant of a matrix has to be nonzero.

So, in fact, we've seen that, oh, there is still one board left.

We've seen that a matrix is invertible -- -- exactly when its determinant is not zero. And, if that's the case, then we can solve the linear system,  $AX = B$  by just setting  $X = A^{-1}B$ . That's going to be the only solution to our linear system. Otherwise, well,  $AX = B$  has either no solution, or infinitely many solutions. Yes?

The determinant of a matrix real quick?

Well, I can do it that quickly unless I start waving my hands very quickly, but remember we've seen that you have a matrix, a  $3 \times 3$  matrix.

Its determinant will be obtained by doing an expansion with respect to, well, your favorite.

But usually, we are doing it with respect to the first row. So, we take this entry and multiply it by that determinant. Then, we take that entry, multiply it by that determinant but put a minus sign.

And then, we take that entry and multiply it by this determinant here, and we put a plus sign for that. OK, so maybe I should write it down. That's actually the same formula that we are using for cross products.

Right, when we do cross products, we are doing an expansion with respect to the first row.

That's a special case. OK, I mean, do you still want to see it in more details, or is that OK?

Yes? That's correct.

So, if you do an expansion with respect to any row or column, then you would use the same signs that are in this checkerboard pattern there. So, if you did an expansion, actually, so indeed, maybe I should say, the more general way to determine it is you take your favorite row or column, and you just multiply the corresponding entries by the corresponding cofactors.

So, the signs are plus or minus depending on what's in that diagram there. Now, in practice, in this class, again, all

we need is to do it with respect to the first row. So, don't worry about it too much. OK, so, again, the way that we've officially seen it in this class is just if you have  $a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3$ , so if the determinant is  $a_1$  times  $b_2 b_3, c_2 c_3$ , minus  $a_2 b_1 b_3 c_1 c_3$  plus  $a_3 b_1 b_2 c_1 c_2$ .

And, this minus is here basically because of the minus in the diagram up there. But, that's all we need to know.

Yes? How do you tell the difference between infinitely many solutions or no solutions?

That's a very good question. So, in full generality, the answer is we haven't quite seen a systematic method.

So, you just have to try solving and see if you can find a solution or not. So, let me actually explain that more carefully. So, what happens to these two situations when  $A$  is invertible or not?

So, remember, in the linear system, you can think of a linear system as asking you to find the intersection between three planes because each equation is the equation of a plane. So,  $Ax = B$  for a  $3 \times 3$  system means that  $x$  should be in the intersection of three planes.

And then, we have two cases. So, the case where the system is invertible corresponds to the general situation where your three planes somehow all just intersect in one point.

And then, the situation where the determinant, that's when the determinant is not zero, you get just one point. However, sometimes it will happen that all the planes are parallel to the same direction.

So, determinant  $A$  equals zero means the three planes are parallel to a same vector. And, in fact, you can find that vector explicitly because that vector has to be perpendicular to all the normals.

So, at some point we saw other subtle things about how to find the direction of this line that's parallel to all the planes. So, now, this can happen either with all three planes containing the same line.

You know, they can all pass through the same axis.

Or it could be that they have somehow shifted with respect to each other. And so, it might look like this.

Then, the last one is actually in front of that.

So, see, the lines of intersections between two of the planes, so, here they all pass through the same line, and here, instead, they intersect in one line here, one line here, and one line there.

And, there's no triple intersection.

So, in general, we haven't really seen how to decide between these two cases. There's one important situation

where we have seen we must be in the first case that when we have a homogeneous system, so that means if the right hand side is zero, then, well,  $x$  equals zero is always a solution.

It's called the trivial solution.

It's the obvious one, if you want.

So, you know that, and why is that?

Well, that's because all of your planes have to pass through the origin. So, you must be in this case if you have a noninvertible system where the right hand side is zero. So, in that case, if the right hand side is zero, there's two cases.

Either the matrix is invertible. Then, the only solution is the trivial one. Or, if a matrix is not invertible, then you have infinitely many solutions.

If  $B$  is not zero, then we haven't really seen how to decide. We've just seen how to decide between one solution or zero, infinitely many, but not how to decide between these last two cases.

Yes? I think in principle, you would be able to, but that's, well, I mean, that's a slightly counterintuitive way of doing it. I think it would probably work.

Well, I'll let you figure it out.

OK, let me move on to the second unit, maybe, because we've seen a lot of stuff, or was there a quick question before that? OK.

OK, so what was the second part of the class about?

Well, hopefully you kind of vaguely remember that it was about functions of several variables and their partial derivatives. OK, so the first thing that we've seen is how to actually view a function of two variables in terms of its graph and its contour plot.

So, just to remind you very quickly, if I have a function of two variables,  $x$  and  $y$ , then the graph will be just the surface given by the equation  $z$  equals  $f$  of  $xy$ .

So, for each  $x$  and  $y$ , I plot a point at height given with the value of the a function.

And then, the contour plot will be the topographical map for this graph. It will tell us, what are the various levels in there?

So, what it amounts to is we slice the graph by horizontal planes, and we get a bunch of curves which are the points at given height on the plot. And, so we get all of these curves, and then we look at them from above, and that gives us this map with a bunch of curves on it.

And, each of them has a number next to it which tells us the value of a function there. And, from that map, we can, of course, tell things about where we might be able to find minima or maxima of our function, and how it varies with respect to  $x$  or  $y$  or actually in any direction at a given point. So, now, the next thing that we've learned about is partial derivatives.

So, for a function of two variables, there would be two of them. There's  $f_x$  which is  $\partial f / \partial x$ , and  $f_y$  which is  $\partial f / \partial y$ . And, in terms of a graph, they correspond to slicing by a plane that's parallel to one of the coordinate planes, so that we either keep  $x$  constant, or keep  $y$  constant.

And, we look at the slope of a graph to see the rate of change of  $f$  with respect to one variable only when we hold the other one constant. And so, we've seen in particular how to use that in various places, but, for example, for linear approximation we've seen that the change in  $f$  is approximately equal to  $f_x$  times the change in  $x$  plus  $f_y$  times the change in  $y$ .

So, you can think of  $f_x$  and  $f_y$  as telling you how sensitive the value of  $f$  is to changes in  $x$  and  $y$ .

So, this linear approximation also tells us about the tangent plane to the graph of  $f$ . In fact, when we turn this into an equality, that would mean that we replace  $f$  by the tangent plane. We've also learned various ways of, before I go on, I should say, of course, we've seen these also for functions of three variables, right? So, we haven't seen how to plot them, and we don't really worry about that too much.

But, if you have a function of three variables, you can do the same kinds of manipulations.

So, we've learned about differentials and chain rules, which are a way of repackaging these partial derivatives.

So, the differential is just, by definition, this thing called  $df$  which is  $f_x dx + f_y dy$ . And, what we can do with it is just either plug values for changes in  $x$  and  $y$ , and get approximation formulas, or we can look at this in a situation where  $x$  and  $y$  will depend on something else, and we get a chain rule. So, for example, if  $f$  is a function of  $t$  time, for example, and so is  $y$ , then we can find the rate of change of  $f$  with respect to  $t$  just by dividing this by  $dt$ . So, we get  $df/dt = f_x dx/dt + f_y dy/dt$ . We can also get other chain rules, say, if  $x$  and  $y$  depend on more than one variable, if you have a change of variables, for example,  $x$  and  $y$  are functions of two other guys that you call  $u$  and  $v$ , then you can express  $dx$  and  $dy$  in terms of  $du$  and  $dv$ , and plugging into  $df$  you will get the manner in which  $f$  depends on  $u$  and  $v$ .

So, that will give you formulas for partial  $f$  partial  $u$ , and partial  $f$  partial  $v$ . They look just like these guys except there's a lot of curly  $d$ 's instead of straight ones, and  $u$ 's and  $v$ 's in the denominators.

OK, so that lets us understand rates of change.

We've also seen yet another way to package partial derivatives into not a differential, but instead, a vector. That's the gradient vector, and I'm sure it was quite mysterious when we first saw it, but hopefully by now, well, it should be less mysterious.

OK, so we've learned about the gradient vector which is  $\nabla f$  is a vector whose components are just the partial derivatives.

So, if I have a function of just two variables, then it's just this. And, so one observation that we've made is that if you look at a contour plot of your function, so maybe your function is zero, one, and two, then the gradient vector is always perpendicular to the contour plot, and always points towards higher ground.

OK, so the reason for that was that if you take any direction, you can measure the directional derivative, which means the rate of change of  $f$  in that direction.

So, given a unit vector,  $u$ , which represents some direction, so for example let's say I decide that I want to go in this direction, and I ask myself, how quickly will  $f$  change if I start from here and I start moving towards that direction?

Well, the answer seems to be, it will start to increase a bit, and maybe at some point later on something else will happen. But at first, it will increase.

So, the directional derivative is what we've called  $f$  by  $ds$  in the direction of this unit vector, and basically the only thing we know to be able to compute it, the only thing we need is that it's the dot product between the gradient and this vector  $u$  hat. In particular, the directional derivatives in the direction of  $i$  hat or  $j$  hat are just the usual partial derivatives.

That's what you would expect. OK, and so now you see in particular if you try to go in a direction that's perpendicular to the gradient, then the directional derivative will be zero because you are moving on the level curve.

So, the value doesn't change, OK?

Questions about that? Yes?

Yeah, so let's see, so indeed to look at more recent things, if you are taking the flux through something given by an equation, so, if you have a surface given by an equation, say,  $f$  equals one. So, say that you have a surface here or a curve given by an equation,  $f$  equals constant, then the normal vector to the surface is given by taking the gradient of  $f$ .

And that is, in general, not a unit normal vector. Now, if you wanted the unit normal vector to compute flux, then you would just scale this guy down to unit length, OK?

So, if you wanted a unit normal, that would be the gradient divided by its length. However, for flux, that's still of limited usefulness because you would still need to know about  $ds$ . But, remember, we've seen a formula for flux in terms of a non-unit normal vector, and  $\mathbf{n}$  over  $|\mathbf{n}|$  dot  $\mathbf{k} dx dy$ . So, indeed, this is how you could actually handle calculations of flux through pretty much anything. Any other questions about that?

OK, so let me continue with a couple more things we need to, so, we've seen how to do min/max problems, in particular, by looking at critical points.

So, critical points, remember, are the points where all the partial derivatives are zero.

So, if you prefer, that's where the gradient vector is zero. And, we know how to decide using the second derivative test whether a critical point is going to be a local min, a local max, or a saddle point. Actually, we can't always quite decide because, remember, we look at the second partials, and we compute this quantity  $ac - b^2$  squared.

And, if it happens to be zero, then actually we can't conclude. But, most of the time we can conclude. However, that's not all we need to look for an absolute global maximum or minimum.

For that, we also need to check the boundary points, or look at the behavior of a function, at infinity.

So, we also need to check the values of  $f$  at the boundary of its domain of definition or at infinity.

Just to give you an example from single variable calculus, if you are trying to find the minimum and the maximum of  $f$  of  $x$  equals  $x$  squared, well, you'll find quickly that the minimum is at zero where  $x$  squared is zero.

If you are looking for the maximum, you better not just look at the derivative because you won't find it that way.

However, if you think for a second, you'll see that if  $x$  becomes very large, then the function increases to infinity. And, similarly, if you try to find the minimum and the maximum of  $x$  squared when  $x$  varies only between one and two, well, you won't find the critical point, but you'll still find that the smallest value of  $x$  squared is when  $x$  is at one, and the largest is at  $x$  equals two. And, all this business about boundaries and infinity is exactly the same stuff, but

with more than one variable.

It's just the story that maybe the minimum and the maximum are not quite visible, but they are at the edges of a domain we are looking at. Well, in the last three minutes, I will just write down a couple more things we've seen there. So, how to do max/min problems with non-independent variables -- So, if your variables are related by some condition,  $g$  equals some constant.

So, then we've seen the method of Lagrange multipliers.

OK, and what this method says is that we should solve the equation  $\text{gradient } f = \text{some unknown scalar } \lambda \text{ times the gradient, } g$ . So, that means each partial,  $f_{\text{sub } x} = \lambda g_{\text{sub } x}$  and so on, and of course we have to keep in mind the constraint equation so that we have the same number of equations as the number of unknowns because you have a new unknown here.

And, the thing to remember is that you have to be careful that the second derivative test does not apply in this situation.

I mean, this is only in the case of independent variables.

So, if you want to know if something is a maximum or a minimum, you just have to use common sense or compare the values of a function at the various points you found. Yes?

Will we actually have to calculate?

Well, that depends on what the problem asks you.

It might ask you to just set up the equations, or it might ask you to solve them.

So, in general, solving might be difficult, but if it asks you to do it, then it means it shouldn't be too hard. I haven't written the final yet, so I don't know what it will be, but it might be an easy one. And, the last thing we've seen is constrained partial derivatives.

So, for example, if you have a relation between  $x$ ,  $y$ , and  $z$ , which are constrained to be a constant, then the notion of partial  $f$  partial  $x$  takes several meanings.

So, just to remind you very quickly, there's the formal partial, partial  $f$ , partial  $x$ , which means  $x$  varies.  $Y$  and  $z$  are held constant.

And, we forget the constraint. This is not compatible with a constraint, but we don't care. So, that's the guy that we

compute just from the formula for  $f$  ignoring the constraints.

And then, we have the partial  $f$ , partial  $x$  with  $y$  held constant, which means  $y$  held constant.

$x$  varies, and now we treat  $z$  as a dependent variable.

It varies with  $x$  and  $y$  according to whatever is needed so that this constraint keeps holding.

And, similarly, there's partial  $f$  partial  $x$  with  $z$  held constant, which means that, now,  $y$  is the dependent variable.

And, the way in which we compute these, we've seen two methods which I'm not going to tell you now because otherwise we'll be even more over time.

But, we've seen two methods for computing these based on either the chain rule or on differentials, solving and substituting into differentials.