IIT Madras

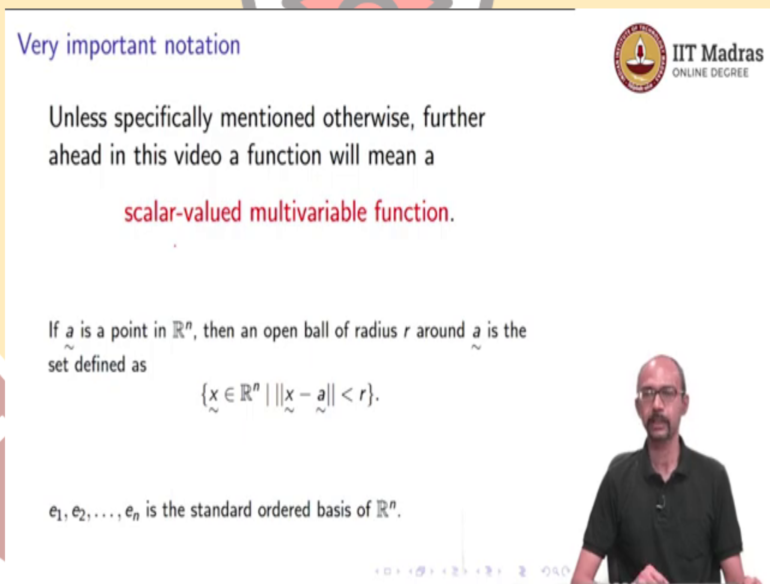
ONLINE DEGREE

Mathematics for Data Science 2
Professor Sarang Sane
Department of Mathematics
Indian Institute of Technology, Madras
Directional derivatives in terms of the gradient

Hello and welcome to the Maths 2 component of the online BSc program on Data Sconce and Programming. This video is about direction derivatives in terms of the gradients. So, first we will define what is the gradient and then we will talk about its relation with the directional derivatives. So, if you may remember that few videos ago when we defined directional derivatives, we found them hard to compute, not very hard, but nevertheless they involved some limits which we had to compute.

And towards the end I said there is often an easier way of computing them and the way to compute them involves some hypothesis call continuity which is why we studies continuity. So, now that we have, we know what is a continuous function, we will talk about directional derivatives again, once we have talked about the gradient. So, the gradient is a very important idea and it will be used extensively in the machine learning course that you will see later on. We will also have use for it in the material that comes later.

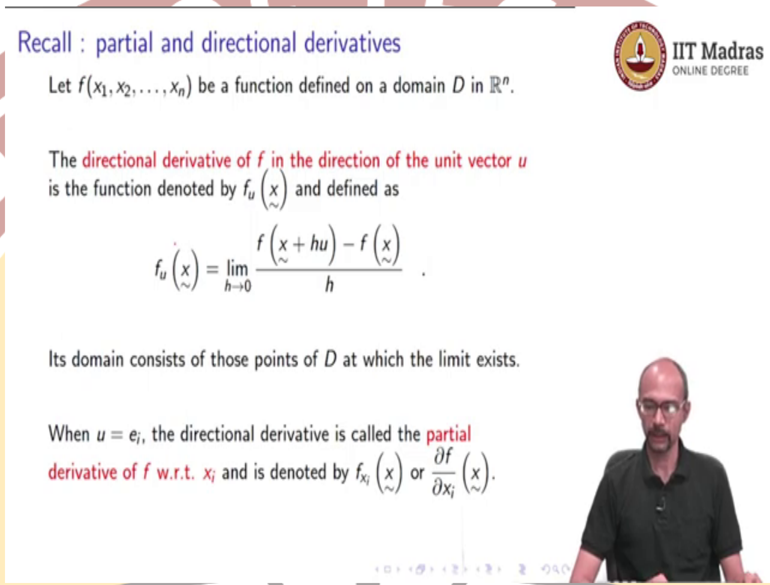
(Refer Slide Time: 1:27)



So, let us recall first, so unless specifically mentioned otherwise in this video, function will mean a scalar-valued multivariable function, so when I say function, by default that means a scalar-valued multivariable function. We will have occasions to use vector-valued functions in this videos but on those occasions I will specifically mention that.

If ã is a point in R^n, then an open ball of radius r around ã is the set defined as those points x̃, such that the distance between x̃ and ã is less than r. So, that is an open ball of radius r. And finally let us recall that e1, e2,.....en is the standard ordered basis of R^n. So, these are all notations we have using for a while now and we will continue to use them in this video as well.

(Refer Slide Time: 2:14)

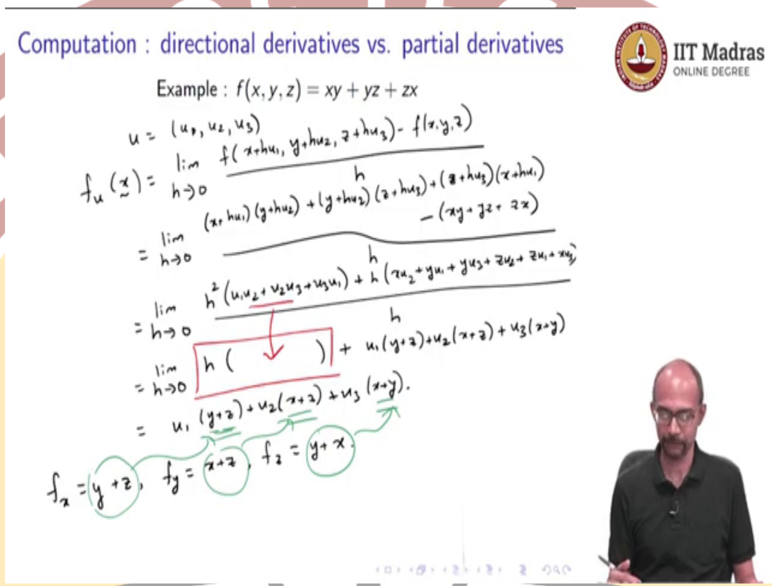


Let us recall partial and directional derivatives. So, if f(x1, x2,...., xn) is a function defined on a domain D in R^n, the directional derivatives of f in the direction of the unit vector u, remember u is, we always think of unit vectors when we talk of directional derivatives, at least when you want to actually to find them, so this is the function denoted by $f_u(\tilde{x})$ and it is defined as $f_u(\tilde{x}) = \lim_{h \to 0} \frac{f(\tilde{x} + hu) - f(\tilde{x})}{h}$.

So, the expression in the bracket that is x̃ + hu, that is addition in R^n, which we know because both x̃ and your vectors and hu is scalar multiplication, so we are very familiar with what that means. So, its domain of course, may not consists of all points in D because this limit may not exist at all points, we have seen examples like that, so it is domain consists of those points at which the limit exists. So, at other points we do not talk about these directional derivatives, we will say it does not exist.

In particular you can take the unit vector, u is e_i, so these are the standard basis vectors and when u is e_i then the directional derivatives is exactly the partial derivative of f with respect to the ith variable x_i and we have a special notation for it, which is $f_{x_i}(\tilde{x})$ or $\frac{\partial f}{\partial xi}(\tilde{x})$ and both are useful in different context. Fine, so this is a recall of partial and directional derivatives. We have studied this in previous videos.

(Refer Slide Time: 4:10)



Let us recall a computation of these. So, if you have f(x,y,z) = xy+yz+zx, let us compute the directional derivatives with respect to vector (u1, u2, u3), so this is a unit vector, so let us compute what is f_u.

So,

$$f_u(\tilde{x}) = \lim_{h\to 0} \frac{f\left(x+hu_1, y+hu_2, z+hu_3\right) - f(x,y,z)}{h}$$

$$= \lim_{h\to 0} \frac{\left(x+hu_1\right)\left(y+hu_2\right) + \left(y+hu_2\right)\left(z+hu_3\right) + \left(z+hu_3\right)\left(x+hu_1\right) - (xy+yz+zx)}{h}$$

We have actually done this computation in previous video, but I want to do it again just to recall how difficult it is to compute a directional derivative as of now. So, if you look at what happens to the numerator, several terms gets cancelled, so xy, yz and zx get cancelled but all the h terms remain, so we have some h terms with h^2, so for h^2 we get $u_1u_2+u_2u_3+u_3u_1$ and then we have plus $h(xu_2 + yu_1+ yu_3+zu_2+zu_1+xu_3)$.

I will suggest you check this, there is some symmetry in this that is why I could write it so easily, which is limit, if you take the limit as h tends to 0, so h and h cancels, and what you will get is h times this expression here plus this entire expression which I will rewrite, x times u_2, I actually want to rewrite this in a different way, so let us take the coordinates. So, $u_1(y+z)+u_2(x+z)+u_3(x+y)$.

So, if you write this on yourself, you will see that this is exactly what you get, so as limit h tends to 0, this term here become 0 because everything inside is a constant with respect to h and h goes to 0, so this is 0 and the other term does not have an h so it is just whatever it is, so the net result is this is $u_1(y+z)+u_2(x+z)+u_3(x+y)$, so we have computed this directional derivative, but notice this that this was a fairly non-trivial computation and this is a fairly easy function.

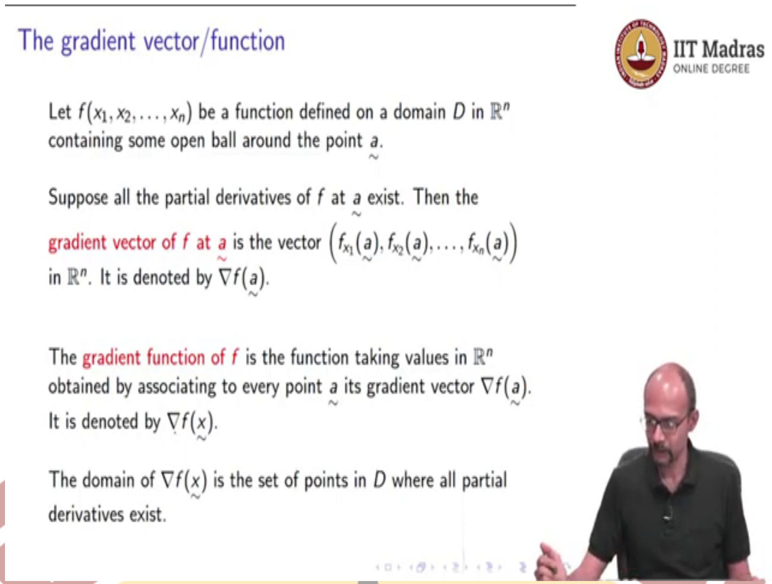
It is a polynomial function, in fact, it has degree just 2, so even for such an easy function it took us some work to compute what is the directional derivatives, contrast this with the partial derivatives which we can write down in a gify, so f_x is, treat x as a variable and y and z as constant, so if you go that you get y+z, f_y is, treat x and z as constant, and y as a variable, so you get z+x, and f_z is treat z as a variable and x and y as constants, so you get y+x.

So, the point here is which is the title of the slide directional derivatives versus partial derivatives, partial derivatives are fairly easy to compute because we have this very nice way of interpreting them in terms of usual derivatives meaning one variable derivatives by treating all the other variables as a constant, whereas for the directional derivative I have to really sit and compute the limit.

Now, I will ask you to observe one important fact here, so notice these coefficients and so this y+z is exactly what you get as the coefficient of u_1, this x+z is exactly what you get as the coefficient of u_2, and this x+y is exactly what you get as the coefficient of u_3. So, maybe there is a relation here between the directional derivatives and the partial derivatives, which we can exploit and it will make computing partial derivatives easy and that is what this video is about.

So, the main upshot is partial derivatives are easy to compute, directional derivatives may be not so easy to compute because you have to actually sit and compute the limit and that, maybe there is a relation we want to exploit. Fine, so to do that we have to define something called the gradient vector or the gradient function.

(Refer Slide Time: 10:04)



This is an exceedingly important function or vector and this is an idea you must understand because you need to know this well to carry forward not just in the next videos in this course but in the next courses as well. So, let us define what the gradient vector or the gradient function is, so suppose $f(x_1, x_2, \ldots x_n)$ is a function defined on a domain D in R^n containing some open ball around the point ã.

Suppose all the partial derivatives of f at ã exist. Then the gradient vector of f at ã is the vector, where in the first coordinate you put the partial derivative with respect to x_1 at ã, in the second coordinate you put the partial derivative with respect to x_2 at ã and in the ith coordinate in general you put the partial derivative with respect to x_i evaluated at ã and form the corresponding vector, so this is a vector in R^n.

It has n coordinates so this is a vector in R^n. So, when you evaluate, since we are evaluating at this point, these are all numbers and this gives you a vector in R^n, so these are all, this is an n tuple of real numbers and so it gives you an element of R^n or a vector in R^n. So, depending on the context you write it as a column vector or a row vector. Again, I have been using this fairly, liberally which way do I interpret vectors as columns or rows.

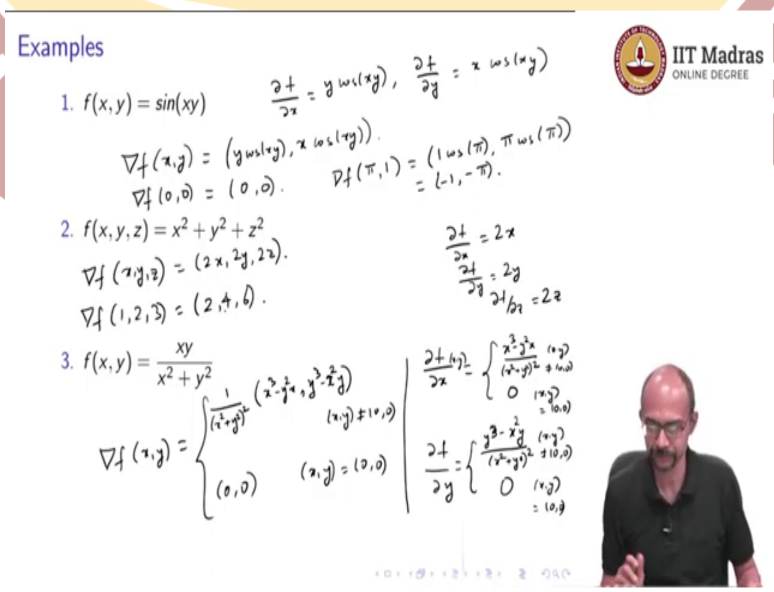
Often, we did it as columns but sometimes I have also done it as rows and the same thing will apply to this vector as well, so depending on whether you write in the context R^n has row vectors or column vectors, you will write the gradient vector as a row vector or a column vector. Important point is this is a vector in R^n, so given a point ã, we had this function f, which was a scalar-valued multivariable function defined on maybe some subset of R^n, some domain in R^n.

And out of this by evaluating at the point ã each partial derivative we have created a vector. So, now first of all notation, this is denoted by sometime it is called del, sometime it is called nabla, so it is denoted by del ∇f evaluated at ã. And then out of this we can form a function, namely what is called a gradient function, this is similar to how we form the derivative function or the partial derivative functions.

So, for each point ã you can associate a corresponding vector so that association we make into a function, so the gradient function of f is the function taking values in R^n obtained by associating to every point ã its gradient vector gradient f at ã, $\nabla f(\tilde{a})$. Of course, this gradient function may not be defined everywhere, because may be some partial derivative is not defined at some point.

So, the domain of definition for this function is exactly those points where all partial derivatives exist, so all partial derivatives must exist and only at such points this function is defined. So, just to emphasize this, this is now a vector-valued function. So, we started with a scalar-valued function f and out of that we have created a vector-valued function gradient of f or del of f or nabla of f, typically we can del f or gradient f, nabla f is not often used, although the symbol is actually called nabla.

(Refer Slide Time: 14:15)



So, this slide if it is a bit abstract, let us do a couple of examples. This is really easy, all we have done is taking all the partial derivatives and put them into a vector, that is really all we have done, but let us do some examples just to set these ideas straight. So, f(x,y) = sin(xy), well we know what is $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, we have computed this earlier. So, $\frac{\partial f}{\partial x} = y\cos(xy)$, $\frac{\partial f}{\partial y} = x\cos(xy)$.

so the gradient function at (x,y) is the vector (ycos(xy),xcos(xy)). And what is the gradient vector at some point? Let us say if I want to compute the gradient vector at (0,0), then this is, you evaluate this at (0,0), so this is a gradient vector at (0,0) and you can see if you evaluate it, you will get (0,0).

If you evaluate it at let us say (π,π), then we have, so well, I get a cosine which I do not really know what to do with, so let us say this is (π,1). So, if I do that then this is, the gradient vector is (1.cos(π),πcos(π)). So, cos(π)= -1, so this is (-1,-π), so I hope it is clear what the gradient function is. So, it is a function from R^2 to R^2 in this case and the gradient vector is just the evaluation of that function at whatever point you want to evaluate it at.

Let us do this other example. So, the gradient of f here is, so I have to write down what are the partial derivatives, so in this case the partial derivatives they are very easy, this is 2x, 2y and 2z, so the gradient function is you just take these and put them into a vector. So, it is a function from R^3 to R^3, so the gradient function is a function from R^3 to R^3. Now, if I want to evaluate this at some point, so what is the gradient vector at the point (1, 2, 3)?

Well, you put those values in and see what you get, so this is (2, 4, 6). Let us do this last example because this is a fairly important example. So, now the gradient function will depend, so we have actually computed the partial derivatives of this function, this was from a previous video, I think the one where we computed partial derivatives. So, let us write down what those are.

So, the values, so this depended on whether we are evaluating in that (0, 0) or not, so it is not at (0, 0), then $\frac{\partial f}{\partial x} = \frac{y^3 - x^2 y}{(x^2 + y^2)^2}$ and if it is at (0, 0), it is 0. And similarly by symmetry we know what this is, $\frac{\partial f}{\partial y} = \frac{x^3 - y^2 x}{(x^2 + y^2)^2}$.

So if you do not remember this or feel uncomfortable with how we got this, please compute this again, so this is if it is not (0, 0), and this is if it is (0, 0). So, that is what the partial derivatives are, which means that now the gradient function is, I want to write this in a particular way, so $\frac{1}{\left(x^2+y^2\right)^2}\left(y^3-x^2y, x^3-y^2x\right)$.

So, this is if (x, y) is not (0, 0) and it is the vector (0, 0) if (x, y) is (0, 0). So, there is two things to note here. First of all, since we know now what is scalar multiplication, I have pulled out this common thing $\frac{1}{\left(x^2+y^2\right)^2}$, so we can do this even for functions, and you have to understand what this means is that it is multiplied by each of the components.

So, we have done this idea extensively when we did linear algebra, so that still continues to hold here, in spite of the fact that these are the functions. And second thing is that the gradient function, it could have this, it is like any other function, it could be piecewise, it could be defined in a different way, it may not be defined everywhere, so all of those things are possible and it takes values again in R^2.

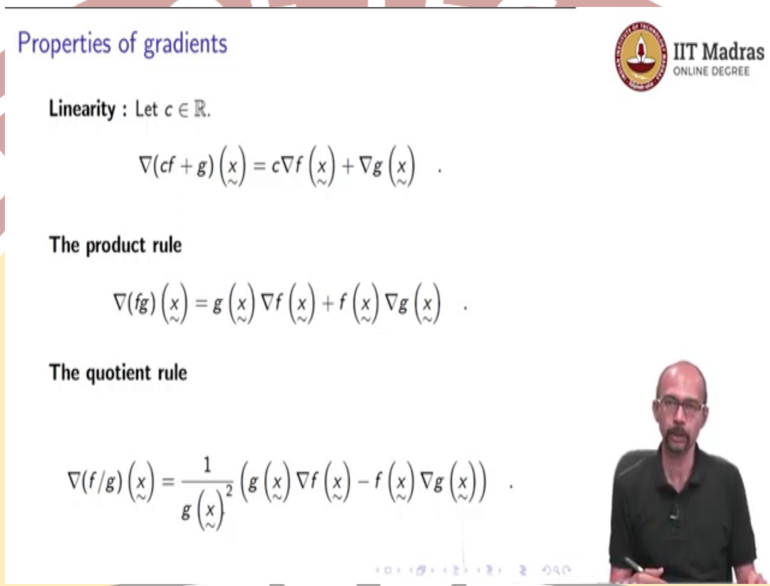
So, this is a function from R^2 to R^2, so I hope these examples are illustrative of what is the gradient function and what is the gradient vector. So, before going ahead maybe we should also ask, what does the gradient vector do for us or what does the gradient function do for us?

So, what are these $\frac{\partial f}{\partial x}$, and $\frac{\partial f}{\partial y}$? They are the rates of changes of the function on the x axis or on the y axis, so with respect to x and with respect to y.

And when you write this two together as function, they are saying that at that point you can compare these two, these two numbers, suppose the, let us say at this element (1, 2, 3), the gradient vector is (2, 4, 6), so what this is saying is the rate of change with respect to x is 2, with respect to y is 4 and with respect to z is 6, so the rate of change is three times that for z as that for x or two times that for y as that x.

So, basically the gradient vector is trying to tell you, it is keeping track, it is one of the things it is doing, it is keeping track of the relative rates of change between the different variables, that is why we are bunching them together, so that we can compare them. So, this is an idea that we have seen in linear algebra earlier which goes back to why do we create vectors in the first place. So, this is one of the reasons we have created the gradient vector. It keeps track of the rate of change for all the variables together. Fine.

(Refer Slide Time: 21:54)



So, now that we have studied the gradients, let us quickly study some properties. These are the same properties that we have studied for directional derivatives, so I am going to summarize them very fast. So, if you have two functions f and g and a scalar c. If you take the gradient of cf+g, then that is $c\nabla f + \nabla g$. So, now what are we doing? We are doing vector addition and scalar multiplication.

So, all these three terms here $\nabla(cf+g)$, ∇f, ∇g are vectors consisting of functions, so they are basically vector-valued functions and $c\nabla f$ is a scalar multiplication of that vector-valued function and we are saying that these vector-valued functions satisfy this equation.

So, this is actually saying that you compare coordinates, and in terms of coordinates we have already seen this when we do directional derivatives. So, this is nothing very deep, this is just rehashing of what we saw for directional derivatives, apply it specifically to partial derivatives. Similarly, we can do the product rule and if we think of the product rule we can summarize it as follows. So, $\nabla(fg) = g\nabla f + f\nabla g$.

And similarly, we can do the quotient rule. Of course, here we have to, we need some hypothesis which I have ignored for now, so $\nabla\left(\frac{f}{g}\right)=\frac{1}{g^2}(g\nabla f-f\nabla g)$. Now, I want to emphasize, of course, all these 3 come with some hypothesis.

You have to know that all these gradients of f and gradient of g are defined and only then all this makes sense and g(x) is nonzero, etc, so with those hypothesis in place. So, the main point I want to make here is that, yes, we need some hypothesis but the main point is this looks very very similar to the rules we had for derivatives in one variable calculus.

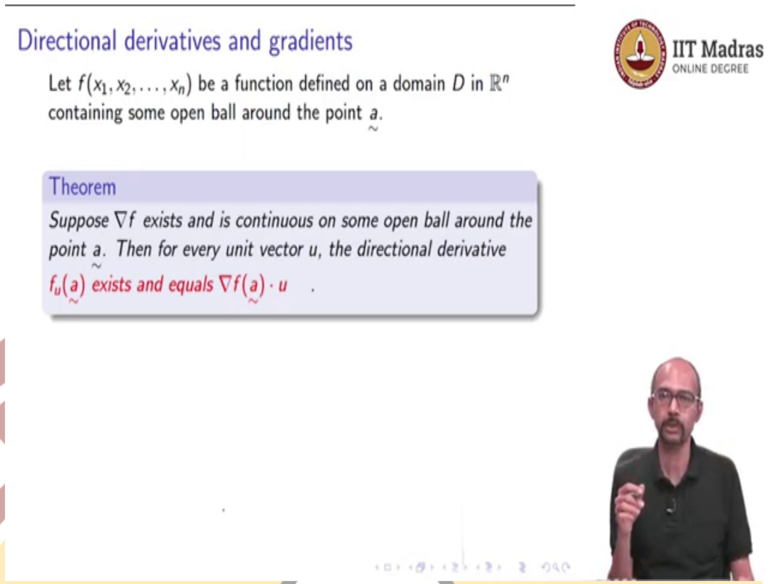
So, when we had one variable calculus we said derivative of cf+g is c times derivative of f plus derivative of g. Derivative of fg is g times derivative of f plus f times derivative of g, and derivative of f by g is 1 by g squared g times f prime minus f times g prime. And of course, when we did one variable calculus I wrote the f prime first and the g next.

So, the reason I have written it like this is because here we have vectors, so I always write the scalar first, this is just convention, there is nothing deep in this. So, the point is the gradient really plays the role of assimilating all the information about the rate of change of this function and can be thought of as the derivative. So, note that for multivariable functions we have not talked about the derivative at all, we only talked about partial derivatives.

So, we only said, how does, what is the rate of change in a particular direction. So, essentially, we talked only about one variable calculus, we reduced it to checking along lines, which is like one variable calculus, but we can ask globally is there some kind of a rate of change for this function, can we define something like the derivative for one variable calculus.

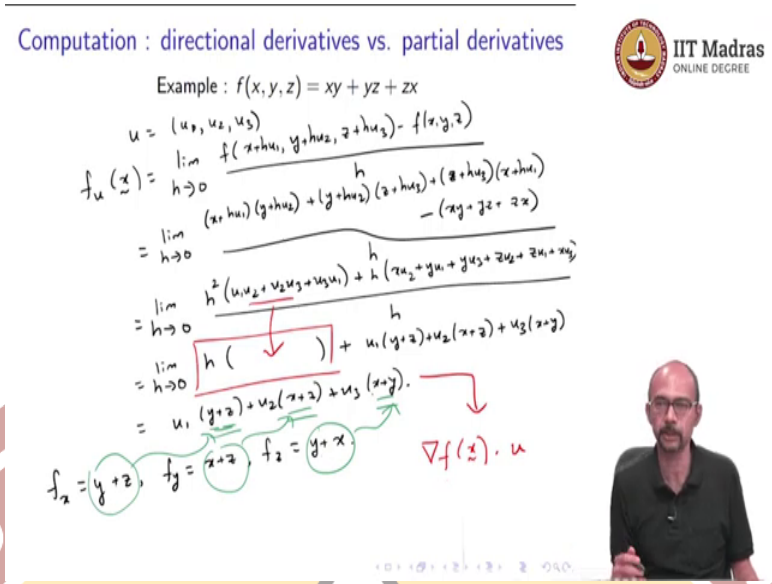
And the answer is, if we want to do it, so that the property is that we have for one variable calculus hold, then this gradient seems like a candidate for the derivative and I will expound more on this later on. So, these are properties of gradients, fine.

(Refer Slide Time: 25:43)



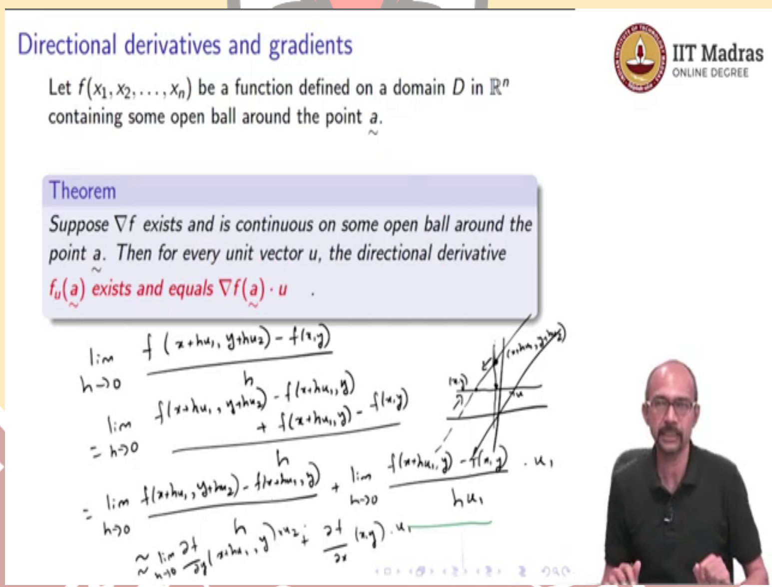
So, now comes the all-important point of this video and this is what we want to take home from this video. Suppose there is a function f defined on a domain D in R^n containing an open ball around the point ã. Theorem: Suppose gradient f exists and is continuous, so this is really why we had to talk about limits and continuity and is continuous on some open ball around the point ã. So, it is not just continuous at that point but it is continuous around that point as well, so at all points close by that point. Then for every unit vector u the directional derivative f_u(ã) exists and equals $\nabla f(\tilde{a}) . u$. This is exactly the formula that we saw for the example that we did earlier, if you remember the example: xy+yz+zx, when we did f_u what we got was $u_1(y+z)+u_2(x+z)+u_3(x+y)$.

(Refer Slide Time: 26:47)



So, this expression here can be rewritten as $\nabla f(\tilde{x}).u$. So, the vector u is (u_1, u_2, u_3), $\nabla f = (y+z, z+x, y+x)$ and if you dot this is the expression you get.

(Refer Slide Time: 27:16)



And that is exactly what this theorem is saying, that you can compute the directional derivatives from the gradient and that vector u. So, the main upshot is that difficult computation that we did with limits for the directional derivative is no longer needed provided we have this hypothesis. So, if the gradient f is continuous, which often happens for many functions, like polynomials and so on, then we need not do the difficult limit, we can just compute it directly by this formula, $\nabla f(\tilde{x}).u$.

So, I want to just make a quick remark about why this holds. So, what is the directional derivative? That is $\lim_{h\to 0}\frac{f\left(x+hu_1,y+hu_2\right)-f(x,y)}{h}$.

So, in terms of my axis, this is like, suppose this is u, badly drawn line, yeah, so this is the line on which u is, this is your u and (x,y) is let us say a point here, so move your vector and your line, so you draw this line here passing through.

And then you take the derivative along this line. So, now what we can do is, we can break this into two parts. We can see this is

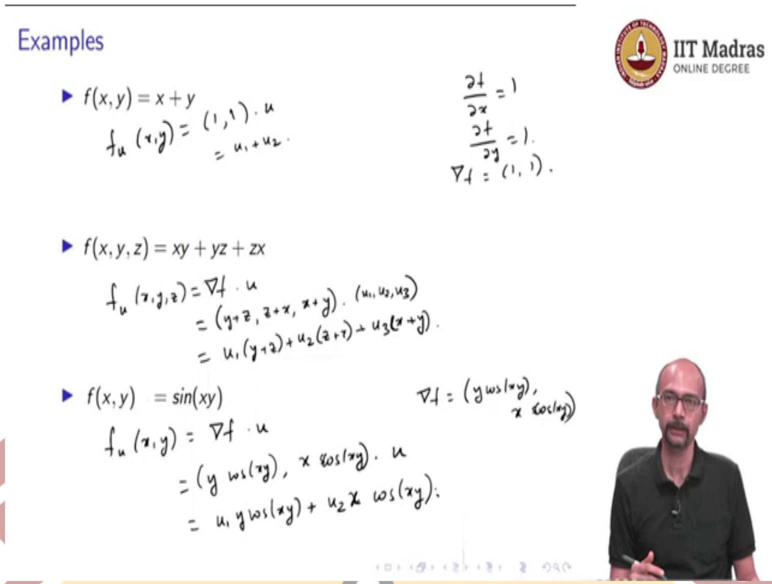
$$\lim_{h\to 0}\frac{f\left(x+hu_1,y+hu_2\right)-f\left(x+hu_1,y\right)+f\left(x+hu_1,y\right)-f(x,y)}{h}$$

$$=\lim_{h\to 0}\frac{f\left(x+hu_1,y+hu_2\right)-f\left(x+hu_1,y\right)}{h}+\lim_{h\to 0}\frac{f\left(x+hu_1,y\right)-f(x,y)}{h}.$$

Now, let us interpret what these two things are. Let us look at the first one. So, this is saying, suppose this is my point $(x+hu_1,y+hu_2)$. So, this is saying let us keep the x coordinate fixed, so that means look at this line, but let us vary y. So, this difference as h tends to 0, this is like y is over here and you are looking along this line.

So, this looks now like a translate of the y axis, so it is like you are computing the partial derivatives with respect to y for this point here. And then the second one, this second expression is like you are computing it for this line here. So, with some little bit of approximation and mumbo-jumbo, this is more or less like saying this is $\frac{\partial f}{\partial y}$ at $(x+hu_1,y)$. Of course, when you take the limit $x+hu_1$ also has to.. there is a limit here so with some mumbo-jumbo you can put in that limit as well and this is actually $\frac{\partial f}{\partial x}$ at (x,y). And because there is a limit here and we know continuity we can say this is same as $\frac{\partial f}{\partial y}$ $(x+hu_1)y$ but now note here that when we did this, here the as $h \to 0$ the h thing change, your parametrization change, so here you have to divide and multiply by u_1 and now you do this. So, now it is like $hu_1\to 0$ and so I get an u_1 extra, and the same thing happens here where you get u_2 extra and that is exactly how we get this formula.

(Refer Slide Time: 32:07)



Let us do some examples and clear out what is going on. So, let us say I want to compute f_u at (x, y). Well, I know I need my gradient, so I have to first compute the partials, so this is $\frac{\partial f}{\partial x} = 1$, $\frac{\partial f}{\partial y} = 1$ that means the gradient is (1,1). So, then this is (1,1).u, so this is just $u_1 + u_2$. So, u = (u1,u2).

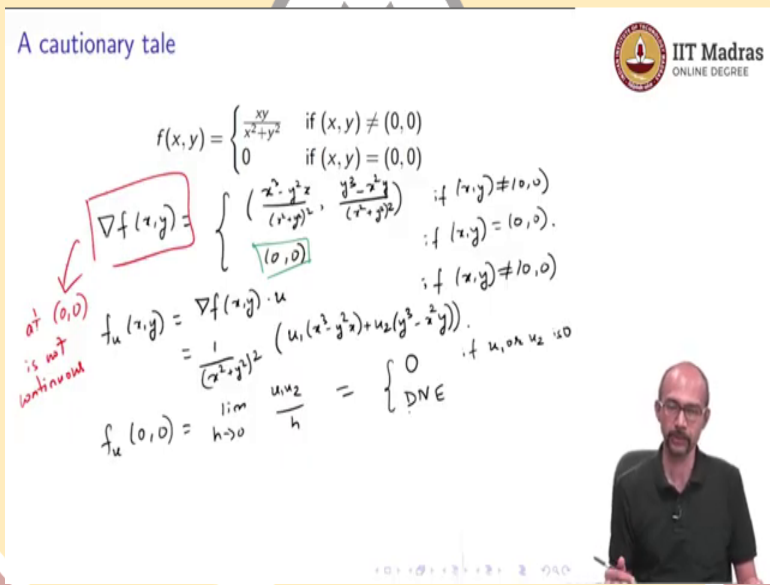
And I will suggest you go back to your video on directional derivatives where we have computed this where it was a fairly, relatively longer computation and we got the same answer. This is something that we did just a few minutes ago, but I am going to do just emphasize that if we use this formula, this is just a one-step procedure, so this is (y+z, z + x, x+y).(u_1, u_2, u_3)=u_1(y+z)+u_2(z+x)+u_3(x+y).

So, you can see now also why this vector notation is useful and the dot products are also useful. We finally have scope to use them in calculus and finally we have f(x,y) = sin(xy), so here f_u(x,y) is, so we need the gradient, so the gradient if you remember was (ycos(xy),xcos(xy)). So, if we put this in, this is (ycos(xy),xcos(xy)).u = $u_1$ycos(xy)+ $u_2$xcos(xy), very easy to compute.

Now, again I will ask you to go back and try to do this from first principle to compute the directional derivative of sin(xy), in fact, we started doing this in our directional derivatives video and then I stopped at some point and suggested that you do it because it became fairly tedious and also I wanted you to check the answer for yourself and now you see if it matches with what we have here.

So, I hope the tediousness of the computation for the directional derivatives is no longer an issue if are bothered in when we first computed directional derivatives that this is a really difficult thing, I hope that now you will feel that this is no longer difficult and that this is actually something very-very easy. Of course, the main idea is to use the gradient vector and, so there is one small warning and the warning is that the hypothesis over there has to be satisfied.

(Refer Slide Time: 35:39)



And so, I want to do this final example:

$$f(x,y)=\begin{cases}\dfrac{xy}{x^2+y^2} & if\ (x,y)\neq(0,0)\\ 0 & if\ (x,y)=(0,0)\end{cases}$$

We have computed already what the gradient is for this. So, the gradient for this is,

$$\nabla f(x,y)=\begin{cases}\left(\dfrac{y^3-x^2y}{\left(x^2+y^2\right)^2},\dfrac{x^3-y^2x}{\left(x^2+y^2\right)^2}\right) & if\ (x,y)\neq(0,0)\\ (0,0) & if\ (x,y)=(0,0)\end{cases}.$$

So, this was a computation we did, the gradient computation we did a few minutes ago and that was dependent on the partial derivatives which we have computed in the partial derivative's video. So, in the directional derivative's video we computed the directional derivatives at (0, 0), let us see what happens for the directional derivative at any other point.

So, if your point is not (0, 0), both these functions $\frac{x^3 - y^2x}{\left(x^2 + y^2\right)^2}$ and $\frac{y^3 - x^2y}{\left(x^2 + y^2\right)^2}$ are continuous. Why is that because they are rational functions and the denominator is non-zero, so once that happens we are assured they are continuous and so we can apply the result that we saw on the, the theorem that we have seen in this video.

So, $f_u(x, y) = \nabla f(x, y) . u$; if (x, y) is not (0, 0), which tells us that this is

$f_u(x, y) = \frac{1}{\left(x^2 + y^2\right)^2}\left(u_1\left(y^3 - x^2y\right) + u_2\left(x^3 - y^2x\right)\right)$. This is what the directional derivative is if the point is not (0, 0). So, notice how that theorem allows us to avoid all the difficult computation with limits.

What happens at (0, 0)? And this is the real cautionary tale here. So, at (0, 0) we actually computed what happens and what we got if you remember was that

$f_u(0, 0) = \lim_{h \to 0} \frac{u_1 u_2}{h} = \begin{cases} 0 & if\ u_1\ or\ u_2 = 0 \\ DNE & \end{cases}$, which means it is in the direction of the x axis or y axis, so those are the partial derivatives, partial derivatives are 0 which we have in this, from here as well, but if it is not, that is not the case then this does not exist, because this is a limit of 1 by h or some constant by h and so this does not exist.

And so, this formula for the gradient, so the gradient dot u will give you the directional derivative does not hold in this case and what went wrong? What went wrong is that at (0, 0) the gradient function here, so this gradient function here at (0, 0) is not continuous and this is actually something we checked in our video on continuity. So, I will again suggest you go back and check that, we have actually seen both of these functions are not continuous at (0, 0).

And we have seen in continuity that for a multivariable function, meaning a vector-valued multivariable to be continuous each of its component functions have to be continuous, in this case neither of them are continuous at (0, 0), so this is not continuous at (0, 0) and immediately the theorem fails to, so we cannot apply that theorem as a result and so we have this situation where the partial derivatives do indeed exist and 0, but directional derivatives does not exist, so it is not given by gradient dot you, which would otherwise have been 0.

So, you have to also careful when you apply that theorem and there is a hypothesis lurking in the background, if you are lucky and your functions are nice you do not have to check that and for this course that is most of the time that is what going to happen, but if you do math in particular, you do have to worry about such details.

So, let us recall quickly what we have seen in this video, very importantly we have seen what is the gradient function or the gradient vector, it is a very easy function, you just put together all the partial derivatives into a vector, that is all it is, but it is a very important quantity and it is going to show up later on in this course as well as other courses.

And the main punch line was that we can use that to compute the directional derivatives with some hypothesis as the gradient vector dot with the vector u, the unit vector u in whose direction we are taking the directional derivative. Thank you.