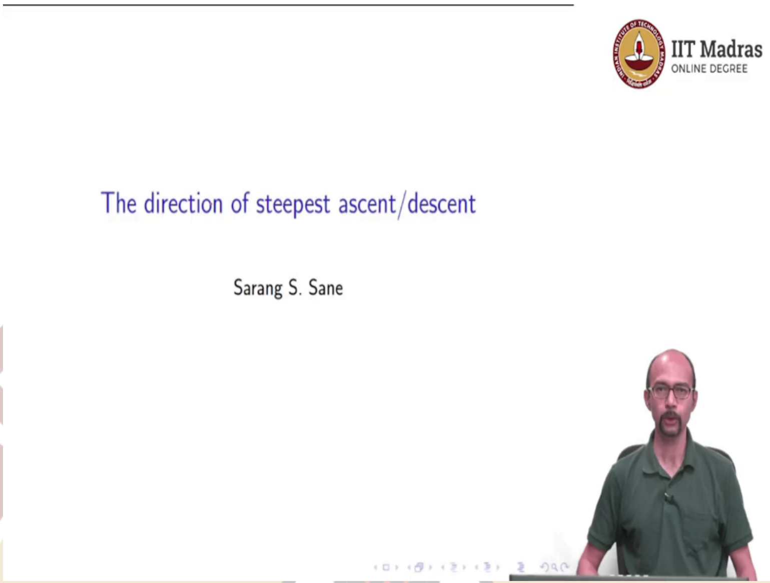
IIT Madras

ONLINE DEGREE

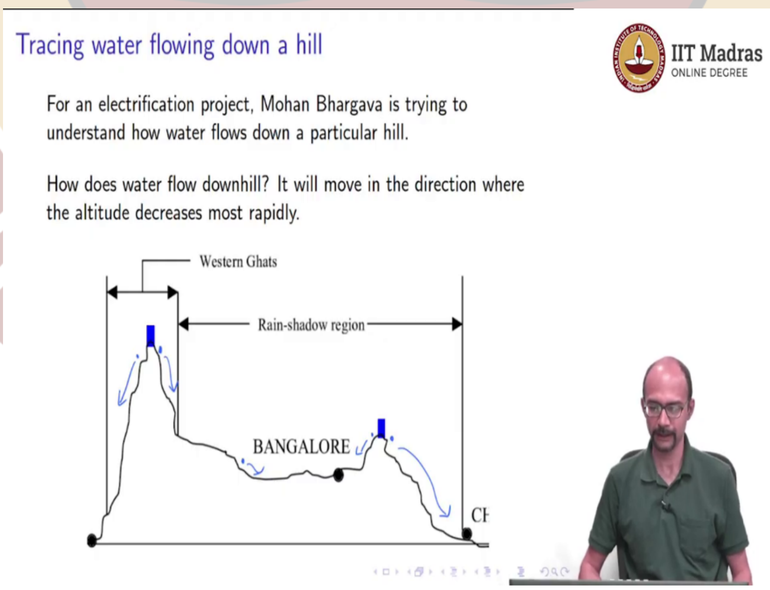
Mathematics for Data Science - 2
Professor Sarang Sane
Department of Mathematics
Indian Institute of Technology, Madras
The Directional of Steepest Ascent/Descent

(Refer Slide Time: 00:14)



Hello, and welcome to the Maths 2 component of the online BSc Program on Data Science and Programming. This video is about the direction of steepest ascent and descent. So, the ideas in this video will be based on what we have studied about gradients and directional derivatives.

(Refer Slide Time: 00:30)



So, let us start with this question about how to trace water flowing down the hill. So, maybe we will do this with an example. So, for an electrification project, someone named Mohan

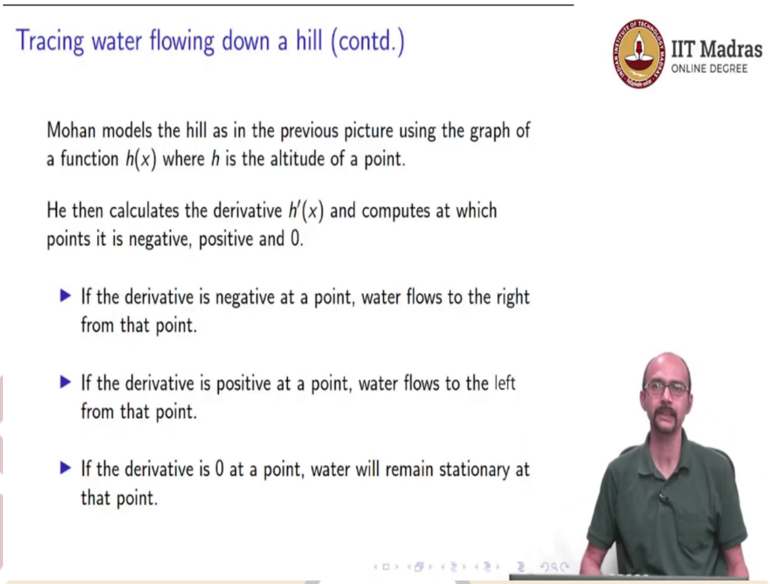
Bhargava is trying to understand how water flows down a particular hill. The name may be somewhat familiar. So how does water flow downhill? So, let us ask this question, how does water flow downhill?

So, if you think about how it flows downhill it always moves in the direction where the altitude decreases most rapidly. So, if you have a rock face, if you have something like this, it is going to flow down here because this is most rapid. So, just as an example here is the, here is a cross section of the Deccan Plateau around Bangalore and Chennai, and so you can see Bangalore here in the middle, you can see Chennai here at very close to sea level on the right and there is the western ghats on the left, and a little bit of inclination between Bangalore and then a steep downfall towards Chennai.

So, just as an example, if you have water, which is somewhere over here it is going to flow down here. If you have water on this side, it will flow down on this side. And let us say if you have water somewhere here it will flow down here, if you have water here it will flow down here, if you are here it flows towards Chennai and so on. So, I think it is clear what we mean by it will move in the direction where the altitude decreases most rapidly.

Now, then, of course, one question. So, what happens if you have water somewhere over here? Somewhere exactly at the top. And there it is actually kind of stable. So, if you have gone to mountain tops you might often find that on the top of the mountain it is often like a plateau and there is, often pools of water over there. So, that does not happen. When you have, when you are on the edges. So, which say something about the rate of change, if you think of it carefully about what is happening to the altitude. This picture is just to give a general idea of what will happen to water. Let us come back to the example of our friend Mohan Bhargava.

(Refer Slide Time: 03:03)

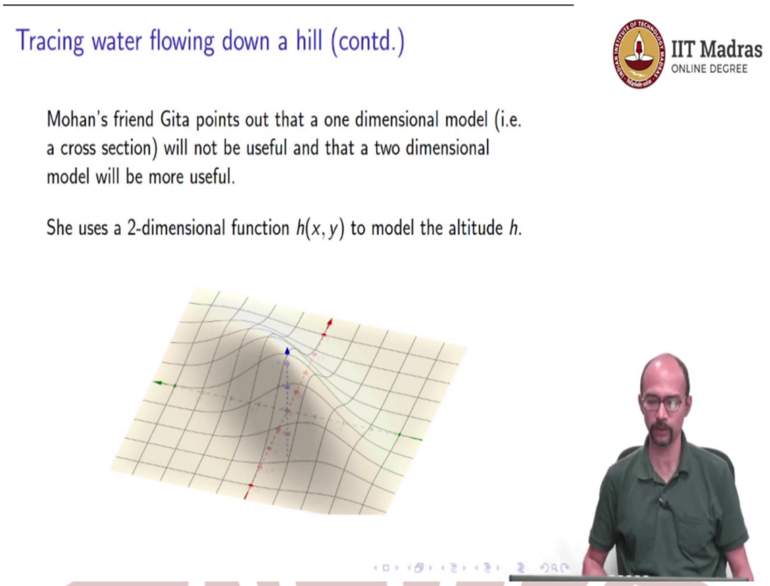


So, what Mohan Bhargava does is, he models the hill similar to the previous picture using a function h(x), where h is the altitude. And then he calculates the derivative $h'(x)$ and computes at which points it is -ve, +ve and 0 and that is exactly what we meant when we said that it moves in the direction in which the altitude decreases most rapidly.

So, the decrease in the altitude will be captured by the rate of change of the altitude. And so, that means $h'(x)$ is going to tell you what is happening at the point x or to the altitude. And so, if the derivative is negative that means the water flows to the right from that point. So, if we say it is negative that means the tangent line is like this. And if the tangent line is like this, it flows down here, which is to the right.

And similarly, if the derivative is positive then the tangent line is like this. So, it flows down here, that is to the left, and if the derivative is 0, then that means the tangent line is like this, and that means the water will remain stationary. And that is exactly what the phenomenon that I was talking about, right at the tip of the mountain at the top. So, this is what Mohan Bhargava tries to do, he tries to model the altitude by this function h(x) draws a picture similar to what we had before, although, not for the entire cross section of India, but for that specific hill, and then then based on that he tries to understand how water is flowing down.

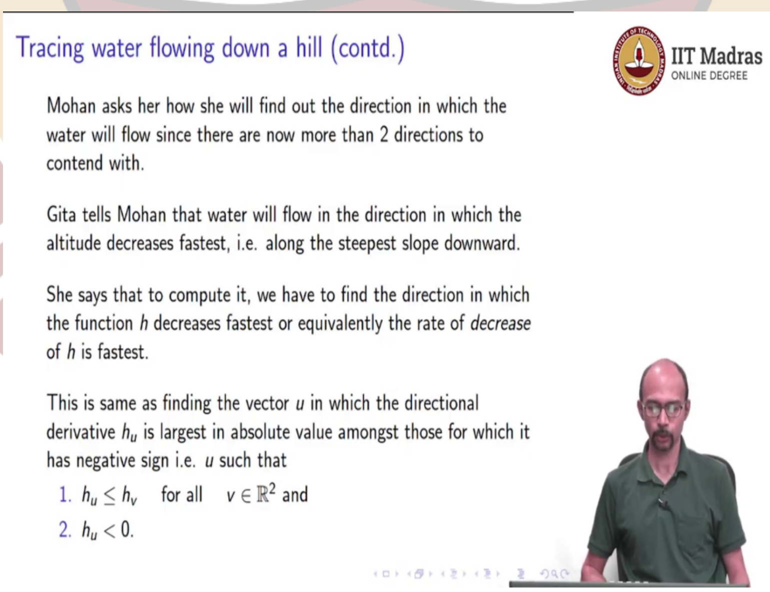
(Refer Slide Time: 04:50)



So, Mohan's friend Gita points out that a one-dimensional model is probably not going to help because the hill is two dimensional, so it has, it is like this and so a one-dimensional picture is not going to help us because there are more than right and left, in real life there are on the ground. That is there are many directions, not just right and left.

So, she uses a two-dimensional function h(x,y) to model the altitude h. And here is an example of such a thing. So, you can see something like a hill, and this is the graph of a function and suitable exponential. And so, what she comes up with is a function of this type h(x, y) the actual hills have more contours and she models it using that function h(x, y). And based on that she tries to understand how the water is going to flow down.

(Refer Slide Time: 06:05)



So, Mohan asked her, how she will find out the direction in which the water flows, because now there is more than two directions. So, if you had only a function of one variable h(x) then computing the derivative told you in what direction the water is going to move, but now you have several directions. And how do I know now based on this function how the water is going to flow down?

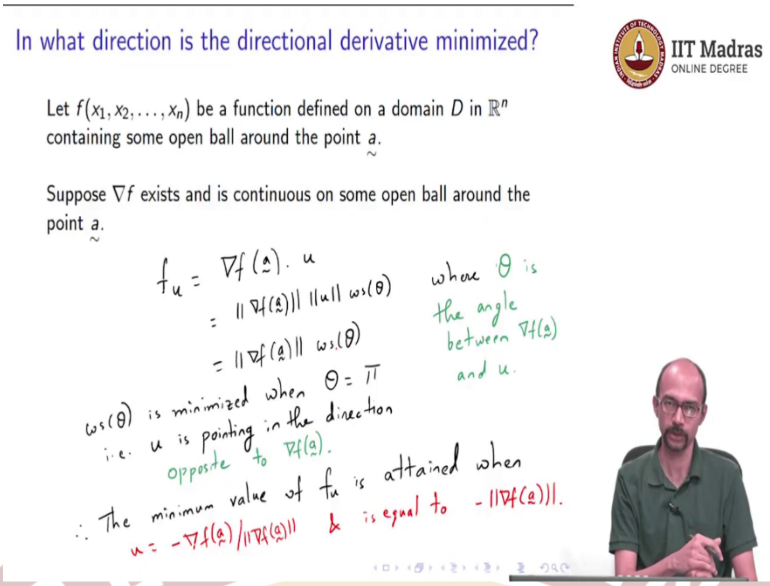
So, Gita says the same principle that the water will flow down along the steepest slope, which is when the altitude decreases fastest. So again, the direction in which the altitude decreases fastest. And now, how do we compute this? So now, rather than saying that we should take h', we know that if we want to know the rate of change in a particular direction, we have something called the directional derivative.

So, she says that, to compute it, we have to find the direction in which the function h decreases fastest, or equivalently, the rate of decrease of h of rate of decrease of h, is fastest. So, this is the same as finding the vector u in which the directional derivative h_u is largest in absolute value amongst those for which it has -ve sign.

So, hu should have negative sign, because we want to decrease, we wanted to decrease h has to decrease fastest, so h_u should have negative sign, but amongst those with negative sign, those directions with negative sign, if you take the absolute value, it is the largest.

So, in terms of the real line, this will be the value at which h_u is smallest in which we include the negative side, so h_u must be smallest, that is what we are saying. So that is u such that $h_u \leq h_v$ for all v in R^2, so that is the smallest value and we also demand that $h_u < 0$, because if it is smallest and positive then that is not going to be of much help. So, this is what we want. So, Gita has now given a different way of doing the same thing, and a more realistic way of doing the same thing.

(Refer Slide Time: 08:32)



So, now we come to the main point of this video, which is, how do we know that direction? So, now we will generalize this. So, suppose we have a function of n variables and it is defined on a domain D in R^n, containing some open ball around the point ã, so we will give a general answer and then we will come back and answer it in this specific question of what direction will the water flow in.

So, we want to know with this hypothesis in what direction is the directional derivative minimized? So, suppose ∇f exists and is continuous on some open ball around the point ã. So once this happens, we have this wonderful theorem that tells us that if you want to compute f_u, then the way to do that is that $f_u = \nabla f(ã).u = ||\nabla f(ã)|| \, ||u|| \cos(\theta)$, where θ is the angle between $\nabla f(ã)$ and u . We have two vectors, so we have seen in the linear algebra part that if you have the dot product then that is the norm of the first vector times norm of the second vector times cos(θ), where θ is the angle between those two vectors. Since it is a unit vector we know that the norm of u is 1, so $f_u = ||\nabla f(ã)||\cos(\theta)$. Now, this function f is fixed, so the gradient is fixed and the gradient at ã in particular is fixed and so its norm is fixed.

So, as you vary across the different directions what changes is θ. So, this is going to change depending on what is θ. So, as θ changes f_u is going to change. So as u changes θ changes, and based on that f_u is going to change. So, where is this going to be maximized and where is this going to be minimized?

So, now that depends on the values of cos(θ), and cos(θ) we understand very, very, very well. So, cos(θ) is smallest, so cos(θ) is minimized when $\theta=\pi$. So, in other words, so if you think of what that means, that means the angle between u and the gradient vector is π that means they are in opposite directions. If gradient is pointing here then you must be pointing here. So that is u is pointing in the direction opposite to $\nabla f(\tilde{a})$.

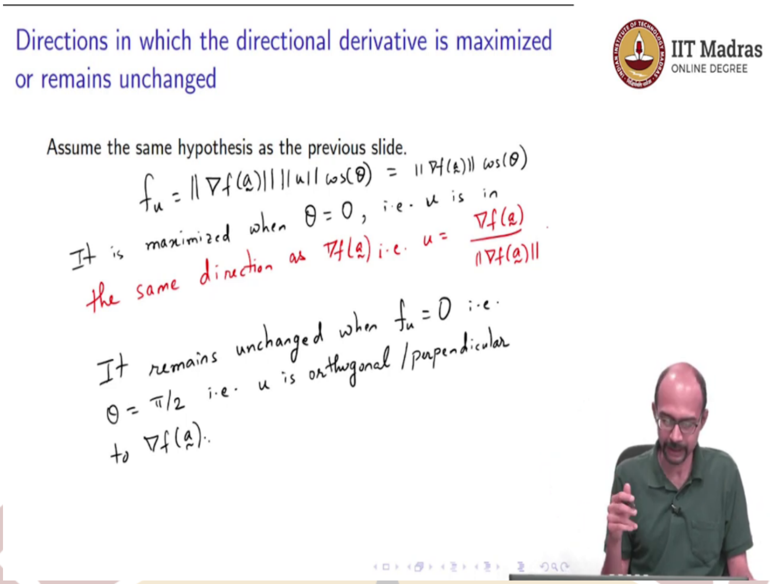
So, this minimum value therefore, the minimum value of f_u is attained when $u = \frac{\nabla f(\tilde{a})}{||\nabla f(\tilde{a})||}$, because we always want a factor of u that has to be of norm 1. So, u is in the opposite direction means u is the minus of gradient it is in that direction, and then to scale it you have to divide by the norm to make it norm 1.

So, the minimum value of f_u is attained when u is this and is equal to. So, now you can substitute u to be this. So, if you do that or you can just say that $\cos(\theta) = -1$,so it equals $-||\nabla f(\tilde{a})||$. So, the upshot of this is that in general if you have the directional derivative, the direction in which it is minimized when your function happens to have this property that gradient f exists and is continuous on some ball around $\tilde{a}$ is in the opposite direction to $\nabla f(\tilde{a})$.

And so, the unit vector you have to that appears there is minus gradient f because that is the opposite direction divided by its norm, because it has to be norm 1. And when you substitute u to be that, which is the same as taking $\cos(\theta) = -1$ in this expression we will get $-||\nabla f(\tilde{a})||$.

So, now, this answers the question that was originally posed. So, if you had the water flow in the steepest direction downwards, it is going to happen at a gradient of h at the point (x, y) minus of that, in the minus of that direction. So, at each point you should take the minus gradient of h at (x, y) and that will be the direction in which the water moves.

(Refer Slide Time: 14:35)



So, one can ask the same question about when is it maximized and when is it, when does it remain unchanged. So, based on our previous discussion, we are going to have the same thing happening provided of course, that the hypothesis holds. So, assume the same hypothesis as in the previous slide which means that the gradient exists and is continuous on an open ball around ã. So, in that case, we will get there h, sorry $f_u = \|\nabla f(\tilde{a})\| \, \|u\| \cos(\theta)$.

Now, if you want to maximize this, so it is maximized and $\cos(\theta)$ is maximized. Well, first of all, I can get rid $\|u\|$ because it is always $\|1\|$, times $\cos(\theta)$ it is maximized when $\theta = 0$, which means u is in the same direction as $\nabla f(\tilde{a})$. So, in other words, $u = \frac{\nabla f(\tilde{a})}{\|\nabla f(\tilde{a})\|}$.

And this also justifies why we call it the gradient. So, the gradient tells you, the word gradient tells you the slope at a point that is what the gradient means, in English in colloquial language. And so, what we are getting is that the gradient actually tells you that at a point what is the steepest slope. So, if you are standing at a particular point on a mountain side the slopes could vary in different directions. So, the gradient tells you the direction in which you have the steepest slope.

And we can also answer the second question, which is when does it remain unchanged? So, it remains unchanged when the rate of change is 0, which means when f_u is 0, so that is when $\cos(\theta) = 0$. And when is $\cos(\theta) = 0$? So $\cos(\theta) = 0$ when $\theta = \Pi/2$, which is the same as saying that u is orthogonal to or perpendicular because here we are using Euclidean norm and so on, so that is perpendicular or at 90 degrees to the gradient vector.

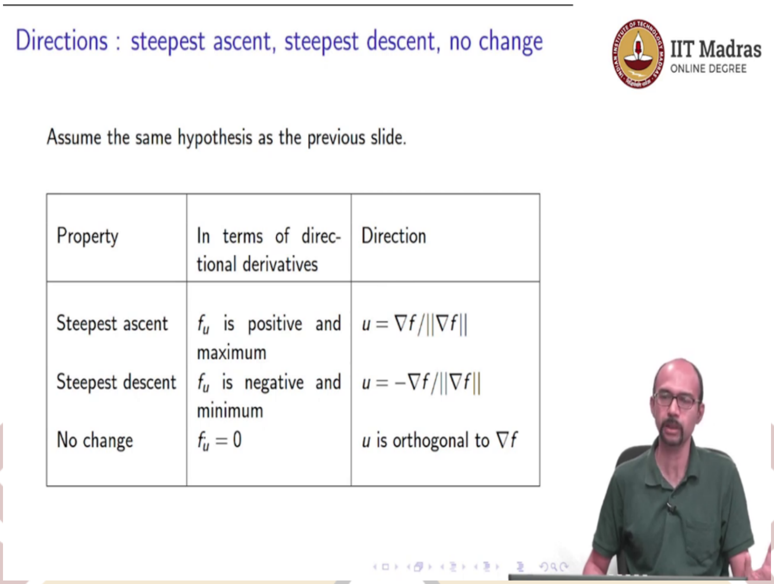
So, we have answered these three things. When is it minimized in the opposite direction to the gradient vector? When is it maximized when it is in the same direction as the gradient vector? And when is the directional derivative constant when you move in a direction perpendicular to the gradient vector.

Now, I want to warn you that in two dimensions, of course, saying perpendicular to the gradient vector means that you go either, so it is a line, you go along this way or this way, but we are in n dimensions. And in n dimension that means there is an equation that you have to solve. And if you do your linear algebra, that means you have a subspace of dimension n-1. And we will not talk a lot about that in this video, but this may come up later.

So, this is this the, these are the answers to these questions. So, we have completely answered the original question, which was in what direction does water flow? On the other hand, if you want to ask, you are an adventure seeker and you want to go in the direction where the slope is steepest, you want to do rock climbing and you want to climb up fastest, so what direction should you go in, the direction of the gradient vector, that is what we have understood from here.

And, on the other hand, if you have, if you do not want to change your altitude at all. You want to keep going at the same altitude and just hope that you do not have to climb down or climb up. And maybe if you go around then you will hit a plateau and you can go you up, you do not want to do any climbing up or down. Then you go in the direction in which the gradient is perpendicular, so you go in that direction. Fine.

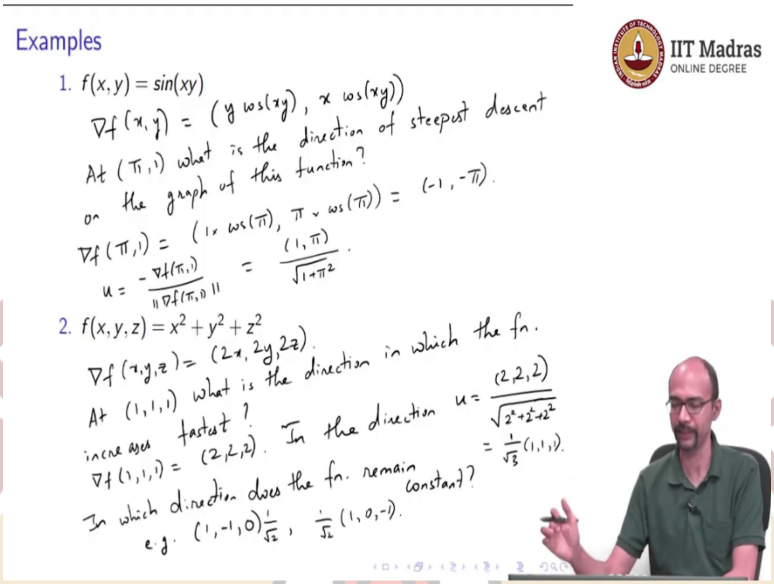
(Refer Slide Time: 19:38)



So, just to summarize what we have said. So, assuming the hypothesis as in the previous slide the property of steepest ascent is in terms of directional derivative that is saying that f_u is positive and maximum and that takes place when $u = \dfrac{\nabla f(\tilde{a})}{||\nabla f(\tilde{a})||}$.

The property of steepest descent occurs in terms of directional derivative when f_u is negative and minimum and the direction is $u = \dfrac{-\nabla f}{||\nabla f||}$. And the property of no change in the height in the altitude or which means in terms of functions no change in the function. So that occurs in terms of directional derivatives when f_u is 0, and that is a direction orthogonal to the ∇ f. So, this summarizes what we have done so far.

(Refer Slide Time: 20:33)



Let us do a couple of examples to put this in perspective. So, suppose I have the function

$f(x, y) = \sin(xy)$, let us compute the gradient. We have actually done this computation before, so the $\nabla f(x,y) = (y\cos(xy), x\cos(xy))$. So, let us choose a particular point and evaluate over there what happens.

So, suppose I choose the point, let us say (Π,1), so at (Π,1) what is the direction of steepest descent on the graph of this function? So, let us first evaluate the gradient vector at this point. So, if you take (Π,1), so, $\nabla f(\Pi,1) = (1\cos(\Pi), \Pi\cos(\Pi)) = (-1,-\Pi)$, so the direction in which you should move is opposite to this. So, $u = \frac{-\nabla f(\pi,1)}{\|\nabla f(\pi,1)\|} = \frac{(1,\pi)}{\sqrt{1+\pi^2}}$. So, this is the direction in which you should move.

If you want to, for example, move in the direction of steepest ascent, where the function increases the fastest, the direction in which the function increases the fastest, then you should move along (-1,-Π), the unit vector in that direction. And if you want to move along the direction where it does not change, so then you should move perpendicular to this vector, which is the vector (Π,-1) or (-Π,1), so in either of those directions.

Let us similarly do this example. So, $f(x,y,z) = x^2+y^2+z^2$ here we compute the gradient, so that is (2x, 2y, 2z). And so, suppose I want to now ask the same thing about let us say at (1,1,1) what is the direction in which the function increases fastest? So, the answer is in the direction of the gradient vector.

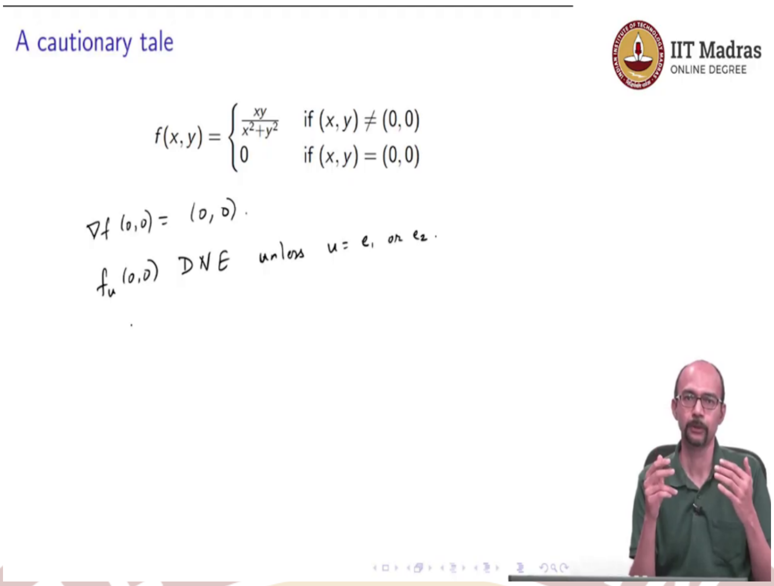
So, the gradient at (1, 1, 1) is (2,2,2). And so, it should be in the direction, so in the direction of u = (2,2,2). And because we typically make this a unit vector we should divide by its norm. So, 2^2 is 4 x 3 is 12, so this is $\frac{1}{\sqrt{3}}(1, 1, 1)$. So, instead if you wanted the direction of no change, so in which direction does the function remain constant?

So, for this, we will have to move perpendicular to the vector (2, 2, 2). Now there is many choices for the perpendicular direction. So, this will be any vector which is on the plane XYZ such that 2x+ 2y + 2z = 0. So, I will encourage you to work out some vectors. So, examples of such would be (1,-1,0), of course, we should take the unit vectors, so, I will divide by root 2 or you could have $\frac{1}{\sqrt{2}}(1, -1, 0)$.

And, in fact, this is a basis for those directions. So, any linear combination of these two would be perpendicular to this direction, and that will be a direction in which the function remains constant. So, I hope the idea here is clear. I want to emphasize that this video is extremely important because this is exactly what you are going to use when you do something called gradient descent in machine learning.

So, in machine learning, you will have a phenomenon called gradient descent where you try to keep decreasing something or increasing something decreasing usually, and so, at a when you are at a particular solution you try to minimize some quantity so as to get what we will call a better solution. So, please, yeah, please understand this video very well.

(Refer Slide Time: 09:05)



I want to end on this cautionary note, since we have been doing this example for a while, now, I want to again emphasize this example to say that we need to know the continuity of the gradient function. So, recall that for this function, the $\nabla f(0, 0) = (0,0)$, but the directional derivatives at (0, 0) other than the x and the y axis meaning the i and j these did not exist unless use e_1 or e_2 or $-e_1$ or $-e_2$, which will be just negative of these.

So, this thing that we used earlier that $f_U = \nabla f.u$, and which we use in order to compute the direction in which it is maximum or minimum that here it is none of those things are applicable. So, you need to know that your gradient is in fact continuous in a small neighborhood. And so, it may happen that the function actually has edges.

So, it is jagged over there. And if that happens then you cannot use any of what we are doing here. So, in such places, the gradient is not going to be, the partial derivatives will not be continuous, and then we cannot apply this theory. So, it may still happen that there is a direction. For example, if it is like this and inclination is more over here you do not want to go in this direction, but we cannot say that from the gradient.

So, that condition is important. So, let us summarize quickly what we have done in this video. So, we have seen this very important idea that, if you have a function f for which the gradient functions continuous, then in order to compute the direction in which the function increases most at a point that is the gradient at that point, the direction in which the function decreases the most rapidly at that point, that is the negative of the gradient direction, and the direction in which the function stays constant that is orthogonal or perpendicular to the direction of the gradient at that point. Thank you.