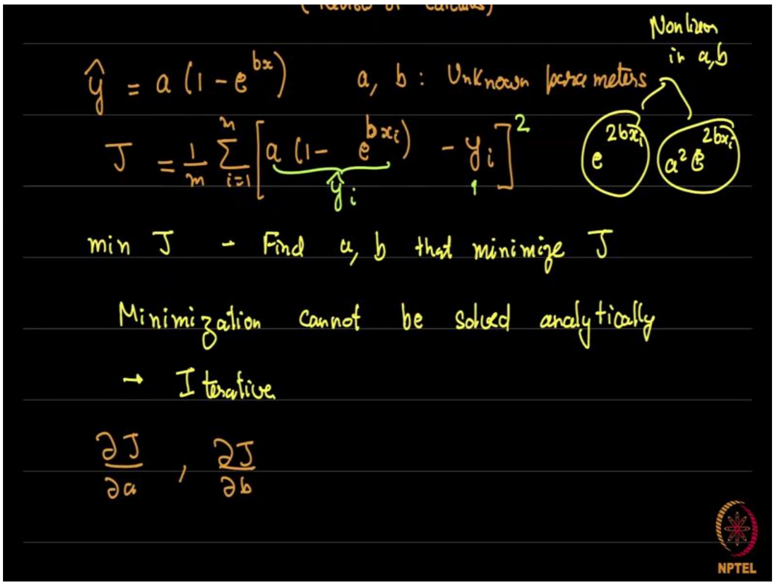
Inverse Methods in Heat Transfer
Prof. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology – Madras

Lecture – 20
Review of required calculus results

(Refer Slide Time: 00:19)



Welcome back. We are in week four of inverse methods in heat transfer. What I will be doing in this video is giving you a very brief review of the calculus that you require for the rest of the course, especially these couple of weeks. In the last video we saw that there are problems within heat transfer which require non-linear regression and when you have such problems. So, let us take the example that we saw in the last video we saw an example of something like $\hat{y} = a(1 - e^{bx_i})$.

And we had a and b as unknown parameters. So, when we have such a case, we take the cost function which we want to minimize for all the points that we have collected. So, we would have something like $\frac{1}{m}\sum_{i=1}^{m} a(1 - e^{bx_i})$, very similar to linear regression except now we use this as $\hat{y}$ and this as y_i. So, remember this is $\hat{y_i}$.

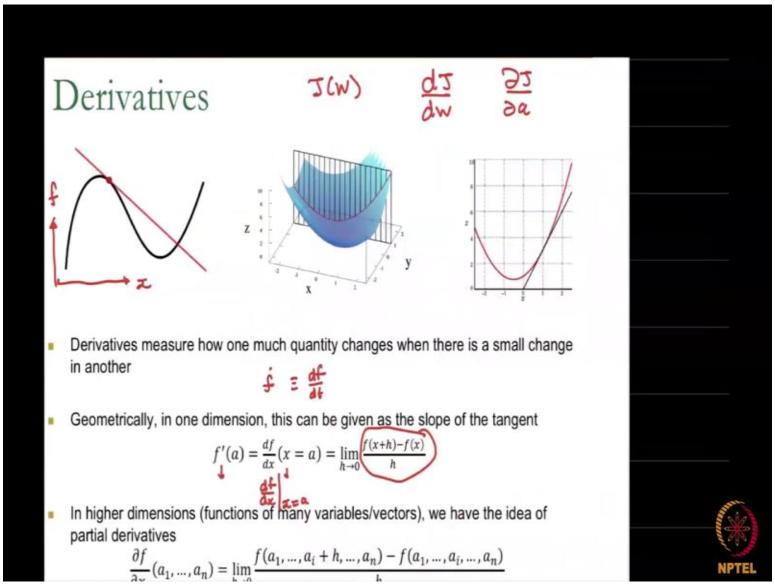
So, this is the prediction, this is the truth and this whole Square and we sum this up and what we want to do is of course find out minimum of J. So, we want to find out find a and b that minimize J. now when you open this bracket up here and see what it looks like you will have terms like e^{x_i} etcetera. So, you would also have terms like $a^2\ e^{x_i}$ all these are non-linear terms in a and b.

So, these are not terms that are linear in a and b linear terms in a and b would look like a multiplied by some constant or b multiplied by some constant whereas if you see you have an exponential here and you have a multiplied by a b term. So, all these are already non-linear in parameters. So, you cannot solve them using the usual techniques. So, this minimization requires a different algorithm. Usually this cannot be solved analytically unlike what we saw in the case of linear regression.

So, this is why we need iteration techniques but in all of them. you will remember that you will have terms like when you want to minimize J with respect to a and b you will have terms like $\frac{\partial J}{\partial a}$ and $\frac{\partial J}{\partial b}$. So, just how to do this in multiple variables is a very quick review which will be sitting within this video. In the future videos of course, I am assuming that all of you have undergone a course in multivariable calculus within college.

In case you have not I would request you to review this and these few minutes that I will spend within this video are just meant for a refresher of what you would have already studied in college. Obviously, it cannot be a full-fledged course within these few minutes that we have.

(Refer Slide Time: 03:54)



So, all these are of course gradients and the simplest version of the gradient that we would have seen in school is what we call the derivative. So, the 1D version one-dimensional, one variable version is what we call the simple derivative now all of us know what the derivative mean all the derivative says is how much does one quantity change when there is a small change in another quantity.

So, the standard derivative we will look at within inverse heat transfer is stuff like $\frac{dJ}{dw}$ or $\frac{\partial J}{\partial a}$, all this means as all of you know is J depends on W, if I change w a little bit how much does J change that is the simple question that the derivative answers. All of you know that physically it can be visualized as the limit of the tangent. So, the limit of the secant which comes to the tangents for example you make a small change in W it makes some change in J.

And you keep on reducing this small amount of w basically you take the limit of the change going to zero. So, notation that you remember from school is if I want f prime, prime denoting the derivative with respect to x at the value x equal to a. So, this is the way we write things there are a few equivalent notations again all this should be familiar to you from school. So, we could write this also as $\left.\frac{df}{dx}\right|_{x=a}$.

So, this is another possibility that you would have seen from your school days another possibility is to write it this way f'(a), in case you take time then we use things like f dot it just denotes derivative with respect to right time. So, what it is, is physically sorry mathematically it is just this,

$$f'(a) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

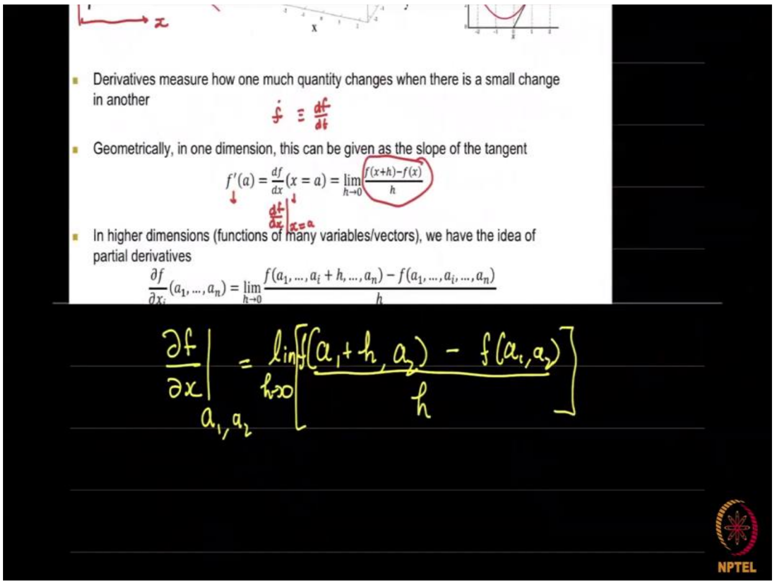
Now within higher Dimensions which is what we are dealing with we have the idea of a partial derivative. So, if this is the one-dimensional derivative if you have x and you have f of x here.

Here the only axis in which you can change is this whereas in higher dimensions when you are at a point you can go in any direction you want you in particular you can go in 2 directions which are coordinate directions. You can go along the horizontal and as you go along the horizontal, there will be a change in the value of the function z let us say is the value of the function. If you travel along this direction and keep on looking at what height you are at you will see that the height changes.

Similarly, you could travel along this direction and keep on saying what height your function is at. So, as is shown here. So, this for example this is a fixed y and x varies and you can see that along that x varying the function changes. Now these are not the only 2 directions along which you can change. you can change along any particular direction at a given point you can go along a 45-degree line in that case, you will have a partial derivative along the 45 degree line we use x and y because they are convenient.

You can also use a 45-degree line, a 60-degree line, 90-degree line any line that you want as long as you move along that line that is still a partial derivative.

(Refer Slide Time: 07:23)



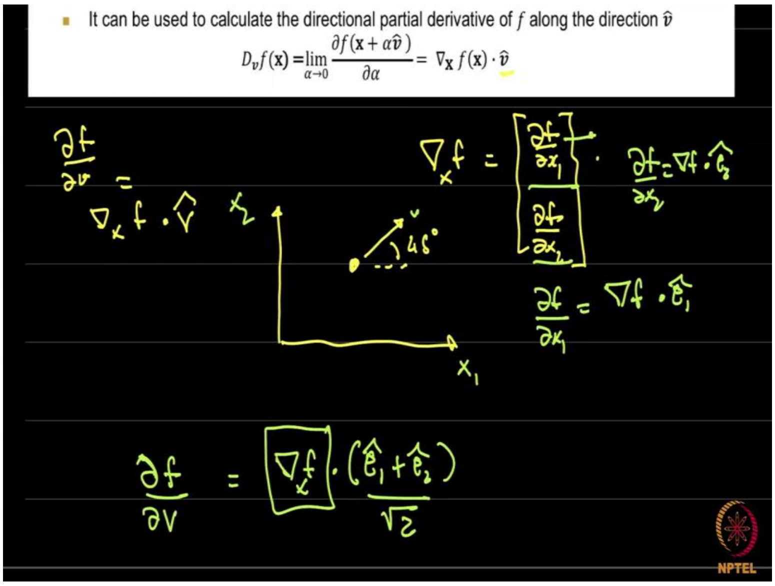
Now a simple expression that we often choose is this $\left.\frac{df}{dx}\right|_{a_1,a_2}$ in case it is a 2 variable thing would be limit at as is shown here if you want this at a_1, a_2 let us say this is 2 variables. So,

$$\left.\frac{\partial f}{\partial x}\right|_{a_1,a_2} = \lim_{h\to 0}\left[\frac{f(a_1+h,\ a_2)-f(a_1,a_2)}{h}\right]$$

So, you stick at the original Point put up the x by a little bit and take limit x tends to zero this will be $\frac{\partial f}{\partial x}$.

Similarly, you can do for $\frac{\partial f}{\partial y}$and that is generalized in the expression here whichever equation whichever independent variable you are taking a partial derivative with respect to just put up that look at the original position and take the limit as h tends to zero. Anyway, like I said this is just refresh it. You can do a lot of examples of this we have asked some simple examples for gradients within the assignment just to serve as a refresh it for you.

(Refer Slide Time: 08:35)



So, now we come to the generalization of a derivative. So, this is the generalization of a derivative, it might seem like it should be the partial derivative, but in higher dimensions you have other things going on. because as we saw just develop del x does not contain the complete information about the change of the variable at that point. So, we visualize it in multiple ways and those ways are shown here again this is only supposed to be a refresher.

So, again now imagine that your variables are here right, again let us say the variables we are considering right now are x_1, x_2 for x and y usually we will be dealing with W's but $x_1$and x_2 something that is familiar to you from school. So, once you come there what the gradient is, it is actually a vector. So, notice this, now why is it a vector. So, this requires a sort of a long explanation. So, I am not going to go into that. it involves something very subtle.

But the point is this when I have a velocity let us imagine some vector like a velocity a 2-dimensional vector. So, let us say this is a 2-dimensional plane and a point starts and it is moving this way and at a particular point you know that there is a direction that you are moving in. now usually the way we represent a vector is by its 2 components. So, I will represent in its x component and its y component.

So, if you have a,

$$\vec{v} = v_1\hat{e_1} + v_2\hat{e_2}$$

the magnitude of V is,

$$|\vec{v}| = \sqrt{{v_1}^2 + {v_2}^2}$$

and of course, the direction is along whatever direction This is let us call it et. So, this is the direction along which this is you can find out, $\tan\theta = \frac{v_2}{v_1}$ etcetera. So, you can find all that now notice this if I ask you not only what is the velocity along x or y but if I ask you the what is the velocity along some other line you can still take a dot product and find out.

What the component of these velocities along any particular line. So, please remember this. The gradient is a similar vector if this is not clear please just use this as I am showing it to you all of us know how to use it. So, gradient we would simply say if you have 2 variables gradient of f with respect to x would be they,

$$\nabla_x f = \frac{\partial f}{\partial x_1}\widehat{e_1} + \frac{\partial f}{\partial x_2}\widehat{e_2}$$

this is one way of writing it or another way of writing it is to say simply gradient of f with respect to x is,

$$\nabla_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$$

But though this is the actual wave in we implement it just like the velocity we represent it via the components. The more important property is the one that I have shown here, which is that when you want the partial derivative. So, when I want the partial derivative along any direction V. So, notice this if I want the partial derivative along any direction just like we have the partial derivative along x and the partial derivative along y we can actually find out the partial derivative along any direction.

By simply saying del f by del V that direction is gradient of f dotted with the unit vector along that direction.

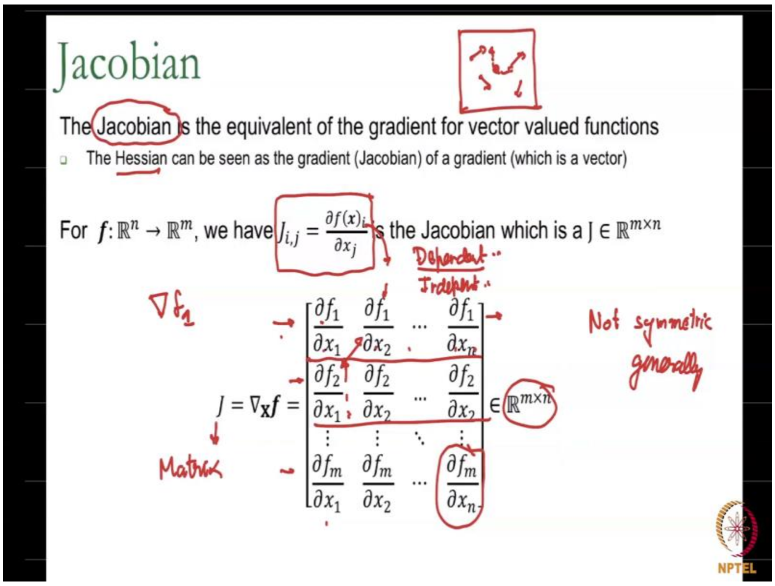
$$\frac{\partial f}{\partial v} = \nabla_x f . \vec{v}$$

So, if I have a point again, I am showing this in 2D if I want del f along one. So, let us call this x_1, x_2 if I want $\frac{\partial f}{\partial x_1}$ all I have to do is take $\nabla_x f \, . \widehat{e_1}$. Now that is obvious when I dot this with $\widehat{e_1}$ all I am going to get is $\frac{\partial f}{\partial x_1}$. Similarly, if I want $\frac{\partial f}{\partial x_2}$ all I need to do is dot this with $\widehat{e_2}$.

But suppose I want the partial derivative of this function along this line and let us say this is the 45-degree line, then all I need to do is to say $\nabla_x f$ or $\frac{\partial f}{\partial v}$ where V is this direction the 45-degree direction is $\nabla_x f.\vec{v}$ at 45 degrees which is $\frac{(\widehat{e_1}+\widehat{e_2})}{\sqrt{2}}$. So, this is what makes the gradient a very powerful quantity not only do you have just 2 pieces of information about what the derivative is along x_1 and what the derivative is along x_2.

You actually can infinitely find out along any direction, what the partial derivative is and we will use this to our benefit in the next video when we talk about what is known as gradient descent.

(Refer Slide Time: 14:05)



Let us come to the next idea which is now going to the next level. So, the next level is like this we looked at the gradient. the gradient was changed of some scalar value f, f here is a scalar with respect to x and x was a vector. x is a vector because x is position and it has components x_1, x_2. Now what about if one vector changes with respect to another Vector, what's an example of this let us say you have a velocity field.

So, you have velocity field you have a flow field. So, you imagine once again for now let us just imagine a 2d space and there is some fluid here. So, this fluid is going to have some velocity at each point. Now if I look at the velocity field you can draw the contours or whichever way you wish to visualize it at each point the vector itself has changes from. So, if the vector moves along y direction it will have a change, if the vector moves along x direction it will have a change. now in any direction that you move your change will also be a vector.

So, just to make this clear, so, let us look at this quantity called a Jacobian. again, this is something that should be familiar to you from your multivariable calculus early in college. ignore the term Hessian written here I will come to that shortly. There is a very simple definition again as far as the matrix is concerned it looks very simple in the sense that you just have to have the element the ij-th element of J is $\frac{\partial f(x)_i}{\partial x_j}$ for example this is element 1, 2.

So, you look at the first component of f and take a derivative with respect to the second component of x. So, notice this is not symmetric in general. These 2 terms have different meanings. So, for example if f is the velocity, then this is the derivative of the first component of the Velocity or x component of the Velocity with respect to the direction y.

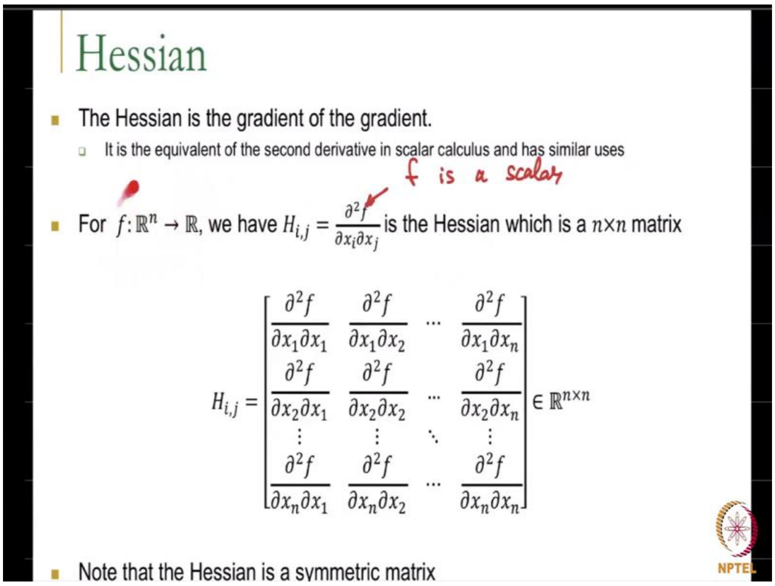
So, how much as we call it in fluid mechanics how much does u change with Y. So, this would look like $\frac{\partial U}{\partial y}$ whereas this term will look like $\frac{\partial V}{\partial x}$. So, those 2 terms in general are not the same they can be different. Now what is the meaning of each one of these. So, remember let us look at split this row wise, what this rows say it says is you can now think of this as the gradient of just the term f1.

So, again if we look at velocity. this would be how much does the U velocity change with each one of the spatial variables. So, how much does the V velocity change with each one of the spatial variables. if you have a three by three that is f itself is a three-dimensional Vector then how much does the W velocity change with all the variables. So, on the top you have the variable that is changing or the dependent variable and, on the bottom, you have the independent variable.

And since each one of them has different components, you are going to in general get del fm by del xn. So, these 2 dimensions need not be the same you are just going to create a m cross n Matrix. So, J is a matrix, now please do not confuse this with the cost function that I showed earlier clear J is the standard name for the Jacobian. So, the Jacobian again you would have seen this in your multivariable calculus classes at college.

So, this is a matrix and it varies, it measures the variation of the dependent variable with respect to the independent variable.

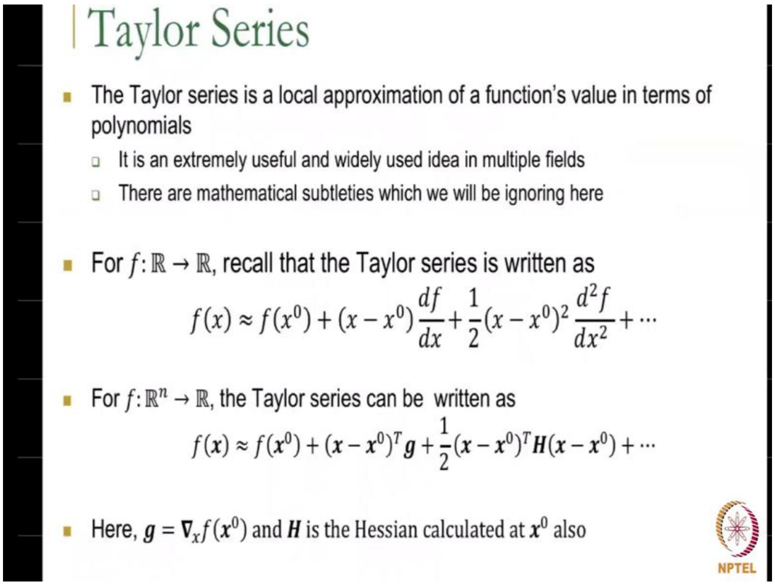
(Refer Slide Time: 18:14)



Even further we have this quantity called the Hessian, which is just like the second derivative equivalent of the Jacobian. So, nothing much changes here but you have cross derivatives you have $\frac{\partial^2 f}{\partial x_1 \partial x_1}, \frac{\partial^2 f}{\partial x_1 \partial x_2},$ etcetera, in this case f is a scalar. So, notice this you can make up a really horrendous function which where that is also a vector but that is no longer a 2-dimensional matrix.

So, this is if f goes from $\mathbb{R}^n \to \mathbb{R}$ means if f is a scalar. So, do not worry about that. So, right now we are not going to look at Hessians in the entire course I just want to use it for one expression in the very next slide. but for the most part you can forget what the Hessian is as long as you understand what the Jacobian is you will be fine.

(Refer Slide Time: 19:08)



So, the important quantity or the important property we want to talk about right now is this local approximation property, which is called the Taylor series. again, something that I am hoping all of you are familiar with from your heat transfer classes earlier calculus classes. the idea is very simple. Whenever you want to approximate a function and why this approximation of functions becomes important should be clear to you by now because most of the times we only have some piecemeal information about the function.

We only have some point information about the function, that at this point my temperature, my thermocouple measured this, at this point some it measured something else what you want to do is to know what it varies like in the middle because you want to take derivatives you want to apply physical loss. So, for that we use Taylor series. So, Taylor series says that if you know the value of a function at a point let us say x0.

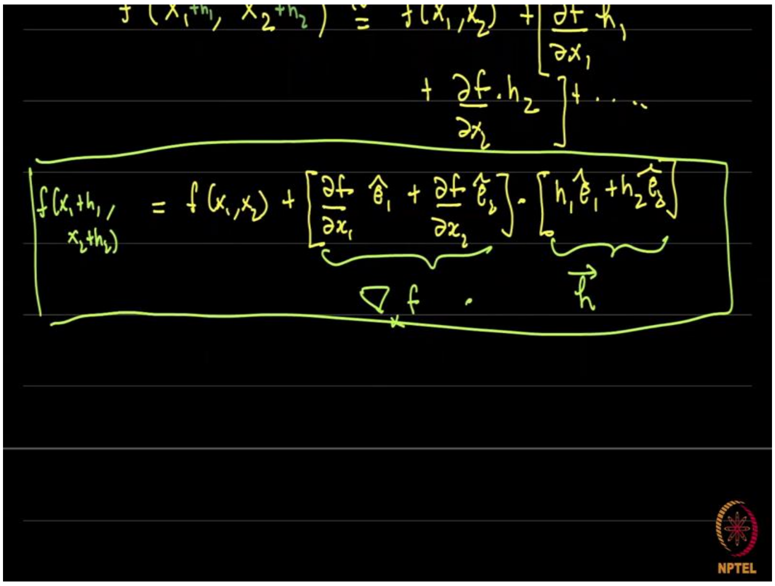
And you also know its derivative then you can kind of reconstruct the function a little bit. So, let us imagine this, you have a point, you know its value and now you know its slope. So, somewhere locally you can say, it will look like, this if you really zoom in. So, if you zoom in you can say that the function looks like this let us say the function actually looks something like this. So, and if you know it is slope here if I go close enough it is going to look like a straight line.

Now we can go further and say I know the point, I know it is slope, I also know its curvature a little bit again locally. So, locally based on the second derivative you can now reconstruct a little bit more. So, this is just a function reconstruction exercise, you know the value, you know its slope, you also know its curvature, then you can say something like this. In fact, you can show that the famous formula S equal to ut plus half at square is literally just the Taylor series assuming that the acceleration is constant.

And the higher derivatives are 0 you will recover s equal to ut plus half at Square from T the mechanics formula. So, you see function first derivative second derivative. Now when you have a multi-dimensional function like this, so, for a multivariable case, this was a single variable case, multivariable case what you need is the first dimensional equivalent is the gradient and this is what is known as the Hessian that I just showed.

So, for now ignore this, for the rest of the course you can happily ignore this. I will briefly refer to it once we come to the machine learning portions, but you can ignore this. Let us see if this first part makes sense. This is the part we are most interested in.

(Refer Slide Time: 21:58)



So, I am going to write this again. So, if I have f the function at a point this will be f at the original function plus gradient of f dotted this is a dot product remember this is a vector, this is also a vector. So, a dot product with h is what the approximation is. So, this is approximated plus you will have some other terms. what does this mean? let us let us look at this. So, let us say this function is of 2 variables x_1 and x_2 and you have the function value at some point in space some x_1, x_2. Now you want to know what its value will be very close to it somewhere slightly different and you know its gradient at that point.

So, this is where all our utility of gradient comes. suppose I want to move a little bit. I am moving a small distance in x, this is the h_1 distance and the small distance in y this is the h_2 distance, then I can say that whatever my old value was plus let me take the derivative in x_1, and multiply it by this distance h_1 then I take. So, I will say approximately the derivative in the y direction and approximate by and multiplied by h_2 and this should be a decent approximation if that is the only information that I have.

Now if you think about this these 2 terms you can write as,

$$f(x_1 + h_1, x_2 + h_2) \cong f(x_1, x_2) + \frac{\partial f}{\partial x_1} h_1 + \frac{\partial f}{\partial x_2} h_2 + \cdots$$

$$f(x_1 + h_1, x_2 + h_2) = f(x_1, x_2) + \left[\frac{\partial f}{\partial x_1} \hat{e_1} + \frac{\partial f}{\partial x_2} \hat{e_2}\right] - [h_1 \hat{e_1} + h_2 \hat{e_2}]$$

Now this is nothing but the gradient of f with respect to x and this is nothing with but h Vector which is what we wrote here. So, you can see that it makes sense what we have written makes sense.

So, this is the primary expression that will be using multiple times throughout this course, including multiple times this week itself. So, please do remember this, the function value at a slightly different point is function value at a given point, plus the gradient dotted with the distance that we are moving. So, in this video we just very quickly covered what should essentially be around a few months of calculus.

But this is just a review in the next video onwards we will start using this idea to derive a new algorithm for optimization that should work for non-linear problems of the sort that I started with at the beginning of this video. Thank you I will see you in the next video.