Multivariable Calculus – Lecture #5 Notes

This lecture explores some of the more technical aspects of limits, continuity, and differentiability of functions of two (or more) variables. In addition, for differentiable functions we'll explore a variety of results growing from the idea of linear approximation in the vicinity of a given point, including error estimation, increments and differentials, rate of change along a parameterized curve, gradients, and the directional derivative.

Tangent plane and linear approximation

We define the partial derivatives of $f(x,y)$ as follows:

$\frac{\partial f}{\partial x} = \lim_{\Delta x \to 0}\left(\frac{f(x+\Delta x, y) - f(x,y)}{\Delta x}\right) = f_x(x,y)$ is the **partial derivative of *f* with respect to *x***

$\frac{\partial f}{\partial y} = \lim_{\Delta x \to 0}\left(\frac{f(x, y+\Delta y) - f(x,y)}{\Delta y}\right) = f_y(x,y)$ is the **partial derivative of *f* with respect to *y***

In order for these derivatives to exist at a given point (x_0, y_0), it's necessary that the cross-sectional curves corresponding to varying just one variable at a time have well-defined slopes at this point. We will rarely need to use this formal definition to calculate partial derivatives. By simply understanding what the definitions are really saying, we can just use familiar rules for differentiation by literally "treating the other variables as though constant".

Taking a vector approach with the parameterization $\mathbf{r}(x,y) = \langle x, y, f(x,y)\rangle$ to produce $\frac{\partial \mathbf{r}}{\partial x} = \langle 1, 0, f_x\rangle$ as a tangent vector to the cross-section where only *x* is varied; and $\frac{\partial \mathbf{r}}{\partial y} = \langle 0, 1, f_y\rangle$ as a tangent vector to the cross-section where only *y* is varied, then if the graph has a well-defined tangent plane the cross product of these two tangent vectors to the graph surface will give a normal vector to the graph at any point (x_0, y_0) on the graph:

$$\frac{\partial \mathbf{r}}{\partial x} \times \frac{\partial \mathbf{r}}{\partial y} = \langle 1, 0, f_x(x_0, y_0)\rangle \times \langle 0, 1, f_y(x_0, y_0)\rangle = \langle -f_x(x_0, y_0), -f_y(x_0, y_0), 1\rangle = \mathbf{n}\,.$$

We can then use this normal vector and the point $(x_0, y_0, f(x_0, y_0))$ on the graph to get an equation for a tangent plane to the graph: $\langle -f_x(x_0, y_0), -f_y(x_0, y_0), 1\rangle \cdot \langle x - x_0, y - y_0, z - f(x_0, y_0)\rangle = 0$ or, after solving for *z*:

$$\boxed{z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)}$$ (equation for tangent plane to graph)

Intuitively, a tangent plane to a graph is that uniquely determined plane that best approximates the graph at a given point. That is:

$$\boxed{f(x,y) \cong f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)}$$ (linear approximation for (x,y) near (x_0, y_0))

Heuristic definition: A function $f(x,y)$ is called differentiable at a point (x_0, y_0) if the graph of $f(x,y)$ has a well-defined tangent plane at $(x_0, y_0, f(x_0, y_0))$ that well-approximates the graph for (x,y) near (x_0, y_0).

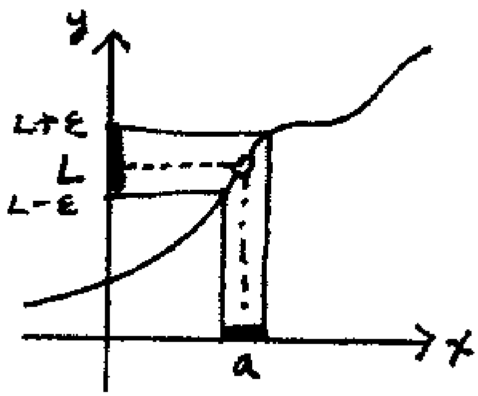
This definition raises a lot of questions. For example, what constitutes a good approximation? What does it mean for (x,y) to be *near* (x_0, y_0), i.e. how close is close enough? How shall distance be measured? In order to better address these questions, we need to review the idea of the limit of a function.

Limit of a function at a point

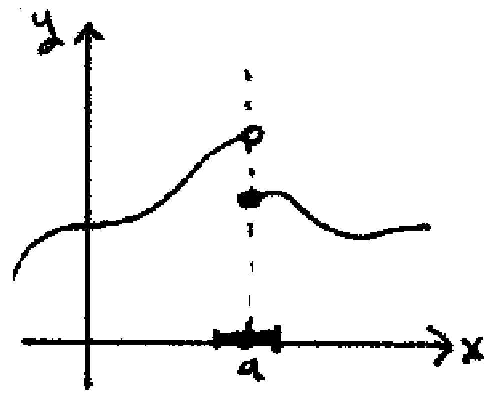
The idea of limits is something that should be familiar to anyone who has experience a course in single variable Calculus. It can be understood either heuristically or using a technical definition. The basic idea is that $\lim_{x \to a} f(x) = L$ if the values $f(x)$ unambiguously approach the value *L* as *x* approaches *a*. What exactly does this mean? This can alternatively be expressed by saying that when x is *close* to a, then necessarily $f(x)$ is close to L. In the case of a function of one variable, the notion of "close" straightforward, namely that $|x-a|$ is small. So $\lim_{x\to a} f(x)=L$ means that if $|x-a|$ is small, then necessarily $|f(x)-L|$ should be small. This continues to evade the question of what is meant by "near" and "small", but this is because these are relative measures. What is "near" to one person might still be "far" to someone with more exacting standards, so how do we resolve this dilemma? To address this, we say phrase the idea in the following more technical definition:

Definition: $\lim_{x\to a} f(x)=L$ if given an allowable error $\varepsilon>0$ (level of accuracy), there is a precision $\delta>0$ such that whenever $0<|x-a|<\delta$, then necessarily $|f(x)-L|<\varepsilon$.

It should be noted that the required precision δ depends very much of the allowable accuracy ε.



Function with a limit as $x\to a$



Function without a limit as $x\to a$

It's worth noting that the specific value $f(a)$ is irrelevant and it doesn't even have to be defined. All that matters is that when x is near a, the values $f(x)$ have to be unambiguous near something. In the right figure, this is clearly not the case and the function has no limit as $x\to a$ (though it does have left-hand and right-hand limits in this example). When the limits and values coincide, we say that a function is **continuous**.

Definition: A function $f(x)$ is called **continuous** at a if (1) $\lim_{x\to a} f(x)$ exists, (2) $f(a)$ is defined, and (3) $\lim_{x\to a} f(x)=f(a)$. We say that a function is continuous on an interval $[a,b]$ if it is continuous at every point in this interval.

How might this work for a function of two or more variables? The complication here is that ***proximity*** is a little more complicated in $\mathbf{R}^2$ and higher dimensions. That said, the idea is still the same.

Heuristically, we say that $\lim_{(x,y)\to(x_0,y_0)} f(x,y)=L$ if whenever (x,y) is *close* to (x_0,y_0), then $f(x,y)$ should be *close* to L. If we choose a consistent way to measure of distance d between points (called a metric), then we might say that whenever $d((x,y),(x_0,y_0))$ is small then the difference $|f(x,y)-L|$ should be small. There are, in fact, different ways to measure the distance between point. We usually default to the **Euclidean distance**, i.e. $d((x,y),(x_0,y_0))=\sqrt{(x-x_0)^2+(y-y_0)^2}$, but we could also use the "**Manhattan metric**" $d_M((x,y),(x_0,y_0))=|x-x_0|+|y-y_0|$, i.e. the distance determined by counting "how many avenues over plus how many streets up" as more appropriate than "as the crow flies". It doesn't fundamentally matter as long as there is some consistent way of capturing the idea of proximity. With this terminology, we offer the following definition:

Definition: $\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L$ if given an allowable error $\varepsilon > 0$ (level of accuracy), there is a precision $\delta > 0$ such that whenever $0 < d((x,y),(x_0,y_0)) < \delta$, then necessarily $\left|f(x,y) - L\right| < \varepsilon$.

Though it's good to have a technical definition, it's also important to know how to use it. More often than not, we just talk our way through a limit without getting ensconced in overly technical details.

Example: What is $\lim_{(x,y)\to(2,1)} \left(\frac{xy}{x^2+y^2}\right)$?

Solution: When x is near 2 and y is near 1, the numerator xy will be approximately 2, and the denominator x^2+y^2 will be approximately $2^2+1^2=5$, so the values of the function $\frac{xy}{x^2+y^2}$ will be approximately $\frac{2}{5}$.

Therefore $\lim_{(x,y)\to(2,1)} \left(\frac{xy}{x^2+y^2}\right) = \frac{2}{5}$. In an example like this there's no need to get more technical than this. On the other hand, consider the following:

Example: What is $\lim_{(x,y)\to(0,0)} \left(\frac{xy}{x^2+y^2}\right)$?

Solution: This time, when $(x,y)\approx(0,0)$ both the numerator and denominator are approximately zero, so we cannot directly determine the values of $\frac{xy}{x^2+y^2}$. One idea is to approach $(0,0)$ along specific curves. If we consider only lines through the origin, we find the following:

Along the x-axis, we have $y=0$, so $\frac{xy}{x^2+y^2} = \frac{0}{x^2} = 0$ at all points, so the limit along this line will be 0.

Along the y-axis, we have $x=0$, so $\frac{xy}{x^2+y^2} = \frac{0}{y^2} = 0$ at all points, so the limit along this line will also be 0.

At this point you might begin to think that $\lim_{(x,y)\to(0,0)} \left(\frac{xy}{x^2+y^2}\right) = 0$, but if we instead approach the origin along the line $y=x$, we'll have that $\lim_{(x,y)\to(0,0)} \left(\frac{xy}{x^2+y^2}\right) = \lim_{x\to 0}\left(\frac{x^2}{2x^2}\right) = \lim_{x\to 0}\left(\frac{1}{2}\right) = \frac{1}{2}$. Therefore, in any small neighborhood of $(0,0)$ this function will take on values that are far apart and **there can be no limit**. In fact, if we approach $(0,0)$ along the line $y=mx$ the function will have the value $\frac{m}{1+m^2}$, i.e. it takes on different (constant) values on different lines all passing through the origin.

Example: Show that $\lim_{(x,y)\to(0,0)} \left(\frac{x^2y}{x^4+y^2}\right)$ does not exist.

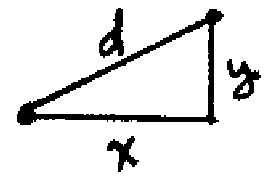
Solution: In this example, along any line $y=mx$ through the origin, the limit becomes $\lim_{x\to 0}\left(\frac{x^2mx}{x^4+m^2x^2}\right) = \lim_{x\to 0}\left(\frac{mx}{x^2+m^2}\right) = 0$, so we might be tempted to conclude that $\lim_{(x,y)\to(0,0)} \left(\frac{x^2y}{x^4+y^2}\right) = 0$. However, if we were to instead approach the origin along the parabolic path where $y=x^2$, the limit would be

$\lim_{(x,y)\to(0,0)}\left(\frac{x^2x^2}{x^4+x^4}\right)=\lim_{(x,y)\to(0,0)}\left(\frac{1}{2}\right)=\frac{1}{2}$. Even if the limit is equal to 0 along every line through the origin, this does not guarantee that the limit is actually 0.

Perhaps the best way to actually prove that a limit exists is to relate as much of the function to distance as possible. The following example illustrates this.

Example: Prove that $\lim_{(x,y)\to(0,0)}\left(\frac{x^2y}{x^2+y^2}\right)=0$.

Solution: As the previous example illustrated, it's not sufficient to show that this limit is equal to 0 along every line through the origin, though we might first try this to convince ourselves that the limit *might* be equal to 0. Using the technical definition of limit, we need to show that if the distance $d((x,y),(0,0))=\sqrt{x^2+y^2}$ is small then this will guarantee that the difference $\left|\frac{x^2y}{x^2+y^2}-0\right|=\left|\frac{x^2y}{x^2+y^2}\right|$ will be small. Referring to the figure, we observe that



$\left|\frac{x^2y}{x^2+y^2}\right|=\frac{x^2|y|}{x^2+y^2}\le\frac{(x^2+y^2)|y|}{x^2+y^2}\le\frac{(x^2+y^2)\sqrt{x^2+y^2}}{x^2+y^2}=\sqrt{x^2+y^2}=d$, so if $\sqrt{x^2+y^2}=d$ is small, then $\left|\frac{x^2y}{x^2+y^2}-0\right|\le d$ will also be small, so $\lim_{(x,y)\to(0,0)}\left(\frac{x^2y}{x^2+y^2}\right)=0$. [$\delta=\varepsilon$ in this case.]

Differentiability of $f(x,y)$ at a point

We have already established that if the graph of $f(x,y)$ has a well-defined tangent plane at $(x_0,y_0,f(x_0,y_0))$, then that tangent plane will have the equation $z=f(x_0,y_0)+f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)$ and for points (x,y) near (x_0,y_0) we will have the linear approximation:

$$f(x,y)\cong L(x,y)=f(x_0,y_0)+f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0).$$

The function $L(x,y)$ is called the **linearization of f at the point** (x_0,y_0). Differentiability can be understood as saying that for points (x,y) near (x_0,y_0), the difference between the actual value $f(x,y)$ and the approximate value $L(x,y)$ will be very small relative to the distance between (x,y) and (x_0,y_0). As the distance grows we would expect this gap to grow, but for nearby points we would like to capture the idea that it should be negligible. Technically, we might capture this idea as follows:

Definition: A function $f(x,y)$ is **differentiable** at a point (x_0,y_0) if $\lim_{(x,y)\to(x_0,y_0)}\left(\frac{|f(x,y)-L(x,y)|}{d((x,y),(x_0,y_0))}\right)=0$.

All this really means is that the tangent plane approximation is a good approximation for the (smooth) graph surface. It should be pointed out that there are examples of functions $f(x,y)$ where both partial derivatives $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$ exist but the function still fails to be differentiable, i.e. the existence of partial derivatives is not sufficient to ensure differentiability.

However, if a function is differentiable, then the linear approximation $f(x,y)\cong f(x_0,y_0)+f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)$ will be valid for (x,y) near (x_0,y_0).

This can be expressed in several alternative forms. For example, if we write this as $f(x,y)-f(x_0,y_0)\cong f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)$ and let $\Delta f=f(x,y)-f(x_0,y_0)$ and $\Delta x=x-x_0$ and $\Delta y=y-y_0$, then using the alternate notation $\frac{\partial f}{\partial x}=f_x(x_0,y_0)$ and $\frac{\partial f}{\partial y}=f_y(x_0,y_0)$ we can write:

$$\Delta f\cong\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y \quad \text{(increment form of linear approximation)}$$

For functions of n variables $f(x_1,\cdots,x_n)$, this construction extends to the more general statement:

$$\Delta f\cong\frac{\partial f}{\partial x_1}\Delta x_1+\cdots+\frac{\partial f}{\partial x_n}\Delta x_n$$

Where the partial derivative $\frac{\partial f}{\partial x_i}$ now presumes that all other variables are treated as though constant when taking this derivative. In particular, for a function of three variables $f(x,y,z)$ we would have that:

$$\Delta f\cong\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y+\frac{\partial f}{\partial z}\Delta z$$

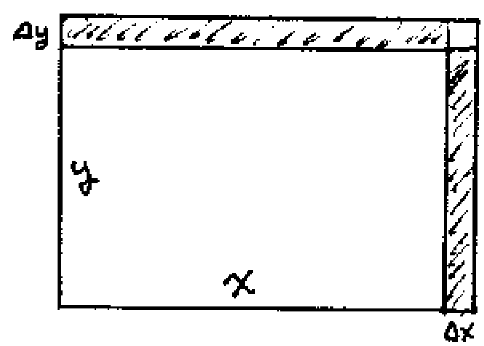
This increment notation can be idealized by considering only the difference in the values of the function as we remain on the approximating tangent plane (or higher order analogue for functions of more than two variables. In this case we identify $dx=\Delta x$ and $dy=\Delta y$ (and $dz=\Delta z$ for a function of three variables) and we express the differential as:

$\boxed{df=\frac{\partial f}{\partial x}dx+\frac{\partial f}{\partial y}dy}$ for a function $f(x,y)$ or $\boxed{df=\frac{\partial f}{\partial x}dx+\frac{\partial f}{\partial y}dy+\frac{\partial f}{\partial z}dz}$ for a function $f(x,y,z)$

The differential notation and the increment notation are basically capturing the same idea with the differential giving the linear approximation for the actual incremental change in the values of the given function as the independent variable in the domain are tweaked.

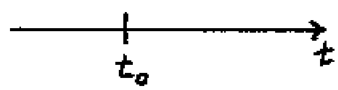
Increments (or differentials) and error estimation

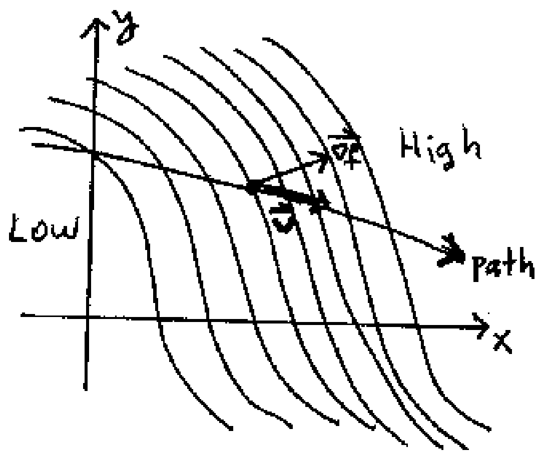
If we have some derived quantity with values expressed in terms of other quantities that are directly measured within some known degree of error, we can use differentials to estimate the error in the derived quantity. The simplest example is in the calculation of the area of a rectangle from the lengths of its sides. In this case, we write $A=xy$, but it must be noted that $A(x,y)=xy$ is, in fact, a function of two independent variables. Therefore if the measured value of x is only known within an error Δx and the measured value of y is only known within an error of Δy, then the incremental change in area associated with these changes will be $\Delta A\cong\frac{\partial A}{\partial x}\Delta x+\frac{\partial A}{\partial y}\Delta y=y\Delta x+x\Delta y$. If the measure error is expressed in terms of relative error (also called percent error when expressed as a percentage), we can divide to get $\frac{\Delta A}{A}\cong\frac{y\Delta x+x\Delta y}{xy}=\frac{\Delta x}{x}+\frac{\Delta y}{y}$. That is, for a quantity derived from two directly measured quantities by multiplication, we add the relative errors to estimate the relative error in the product.



This kind of analysis can be applied to any derived quantity expressed as a differentiable function of directly measured quantities.

Rate of change of a function along a parameterized curve
Suppose a function $f(x, y)$ is given and we think about how its values are distributed by drawing level curves of this function in $\mathbf{R}^2$ (see diagram). Let's also consider a parameterized curve described by $\mathbf{r}(t) = \langle x(t), y(t) \rangle$. We can then ask the question: What is the rate of change of the function $f(x, y)$ as we travel along this parameterized curve?





If we think of this as a composition, we have:

$$t \to (x(t), y(t)) \to f(x(t), y(t))$$

This is a function from $\mathbf{R}$ to $\mathbf{R}^2$ to $\mathbf{R}$ and we wish to calculate $\frac{d}{dt}\left[f(x(t), y(t))\right]$. For brevity we express this more simply as $\frac{df}{dt}$, but it must be understood that this rate is being calculated along a particular parameterized curve. We can relate incremental changes in the values of this function as $\Delta f \cong \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$, so if we simply divide through by Δt we see that $\frac{\Delta f}{\Delta t} \cong \frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t}$ and if we let $\Delta t \to 0$ we get that $\frac{df}{dt} = \lim_{\Delta t \to 0}\left(\frac{\Delta f}{\Delta t}\right) = \frac{\partial f}{\partial x}\lim_{\Delta t \to 0}\left(\frac{\Delta x}{\Delta t}\right) + \frac{\partial f}{\partial y}\lim_{\Delta t \to 0}\left(\frac{\Delta y}{\Delta t}\right) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$. That is $\boxed{\frac{d}{dt}\left[f(x(t), y(t))\right] = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}}$. This is the **Basic Chain Rule**.

Note that this expression for this rate of change is given in the form of a dot product. Specifically, we can write $\frac{df}{dt} = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \cdot \left\langle \frac{dx}{dt}, \frac{dy}{dt} \right\rangle$. The 2nd factor should be familiar as the velocity vector $\mathbf{v}$ to this parameterized curve (at whatever point we are passing through at time *t*). The 1st factor is something new known as the **gradient vector** of the function $f(x, y)$. It gives a vector at every point (x, y). [This is known as a vector field.] We denote the gradient by $\overrightarrow{\nabla f} = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle$. With this we can say that $\boxed{\frac{df}{dt} = \overrightarrow{\nabla f} \cdot \mathbf{v}}$.

The same construction can be done with a differentiable function $f(x, y, z)$ and a parameterized curve $\mathbf{r}(t) = \langle x(t), y(t), z(t) \rangle$ to give the rate of change $\boxed{\frac{d}{dt}\left[f(x(t), y(t), z(t))\right] = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}}$ as the Basic Chain Rule in this context. If we define $\overrightarrow{\nabla f} = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle$, we can then again write that $\boxed{\frac{df}{dt} = \overrightarrow{\nabla f} \cdot \mathbf{v}}$.

Geometry of the gradient vector

We can best understand the direction of the gradient vector as well as its magnitude by considering particular paths. With this we'll see that its **direction is always in the direction of steepest increase** in the values of the function. Furthermore, it must therefore always be **perpendicular to the level sets** of the given function. This observation will give us a simple way of producing normal vectors to curves and surfaces. Here's how we determine these facts:

(1) Choose any parametrized path that lies entirely within a level set. By the Basic Chain Rule (in vector form) we see that $\frac{df}{dt} = \overrightarrow{\nabla f} \cdot \mathbf{v} = 0$ along any such path because the values of the function are constant. But $\overrightarrow{\nabla f} \cdot \mathbf{v} = 0$ then means that either $\overrightarrow{\nabla f} \perp \mathbf{v}$ (and therefore $\overrightarrow{\nabla f}$ must be perpendicular the level set) or possibly that $\overrightarrow{\nabla f}$ vanishes completely (which we will soon identify as indicative of a *critical point*). So as long as the gradient is nonzero it will be perpendicular to the level sets of the function.

(2) Choose any parametrized path along which the values of the given function are increasing. Then the rate of change will be positive, i.e. $\frac{df}{dt} = \overrightarrow{\nabla f} \cdot \mathbf{v} > 0$. This means that the angle between $\overrightarrow{\nabla f}$ and $\mathbf{v}$ must be an acute angle. Referring to the previous diagram we see that $\overrightarrow{\nabla f}$ must, in fact, be pointing in the direction of (steepest) increasing values.

(3) What about the magnitude $\left\|\overrightarrow{\nabla f}\right\|$? For this, let's introduce one more definition. Note that the rate $\frac{df}{dt} = \overrightarrow{\nabla f} \cdot \mathbf{v}$ depends very much on the speed with which the parameterized curve is traversed. We can eliminate this bias by instead tracking the rate of change <u>per distance traveled</u>. To do this, we need only assume that we travel at unit speed, but this also means that time elapsed and distance traveled will be the same. If we express the (unit) velocity vector as **u** and note that the rate can now be expressed as $\frac{df}{ds} = \overrightarrow{\nabla f} \cdot \mathbf{u}$, we get the following definition:

> **Definition**: The **directional derivative** of a function f at a point (position) $\mathbf{x}_0$ in the direction of the unit vector $\mathbf{u}$ is the scalar quantity $\boxed{\frac{df}{ds} = (D_{\mathbf{u}} f)(\mathbf{x}_0) = \overrightarrow{\nabla f}(\mathbf{x}_0) \cdot \mathbf{u}}$.

Note that the directional derivative depends not only on where you are but also the direction in which you go. Indeed, **it's just the scalar projection of the gradient in any given direction**. In fact, it's easy to see that for any function the partial derivative $\frac{\partial f}{\partial x}$ coincides with the direction derivative in the *x*-direction, and the partial derivative $\frac{\partial f}{\partial y}$ coincides with the direction derivative in the *y*-direction, etc. Indeed, the directional derivative can be viewed as a generalization of the idea of the partial derivative in any direction you please.

Getting back to the meaning of $\left\|\overrightarrow{\nabla f}\right\|$, note that if we choose a unit vector in the gradient direction (the direction of steepest increase), the directional derivative in that direction will then be $\left\|\overrightarrow{\nabla f}\right\|$. That is, $\left\|\overrightarrow{\nabla f}\right\|$ can be interpreted as the maximum rate of change of the given function per distance traveled. In the real world, this is what we might generally refer to as the grade. A steep hill, for example, would be indicated by a large gradient vector.

Notes by Robert Winters