A Catalog of Real Differentiation

John McCuan

September 1, 2020

This is probably a suitable and useful review for most of you. Technically, most of this material should be prerequisite for the course. Also, I didn't want to spend a lot of time typing up notes and creating graphics. I was thinking to just say what I was going to say about the material in the lecture, but it probably would be useful for you to have a printed reference. As a compromise, I'll type, but you'll need to provide your own pictures. These are mostly important in the first section, but that is also the most important section, so make sure you produce good pictures with all the details and that you understand them.

1 Calculus I

Differentiation from 1-D (one dimensional) calculus or Calculus I is the basic starting point for all other differentiation. It's now time for you to understand it completely. and it's important for you to understand it. You should understand it well enough to teach someone else about it, so I shouldn't need to write the following. You should be able to write it. But the fact of the matter is that you probably can't. At least you should be able to make the pictures.

This kind of differentiation applies to a function $f : (a, b) \to \mathbb{R}$ where (a, b) is an open interval of real numbers:

$$(a, b) = \{x \in \mathbb{R} : a < x < b\}.$$

One or both of the numbers a and b might be ∞. More precisely, maybe $a = -\infty$ or maybe $b = +\infty$. Thus, a and b are (fixed) in the set of **extended real numbers**.

We take a point $x_0 \in (a, b)$ and form the **difference quotient**

$$\frac{f(x_0 + h) - f(x_0)}{h}.$$

This is where you need to draw your first picture. I'll describe it. You should draw the graph of a function f over an open interval with endpoints labeled a and b. If you're dreadfully uncreative you can use $f(x) = x^2$ with $a = 1/2$ and $b = 2$. Next, pick out and label a point x_0 in the interval (a, b).

The number h is called an **increment**. The number h can be positive or negative, but the number h cannot be zero. This is sort of important to remember and understand. So you should draw two pictures now. One should have $h < 0$ and the point $x_0 + h$ labeled. The other should have $h > 0$.

Now you can plot two points on the graph: $(x_0, f(x_0))$ and $(x_0 + h, f(x_0 + h))$. You'll note that these are points in $\mathbb{R}^2$ which is the set containing the **graph**:

$$\mathcal{G} = \{(x, f(x)) : x \in (a, b)\} \subset \mathbb{R}^2.$$

The graph is also a subset of $\mathbb{R}^2$. Actually, you should plot the two points on each graph, the graph for $h < 0$ and the one for $h > 0$.

Finally, you should draw a right triangle on each graph and label the **increment of the values** $f(x_0 + h) - f(x_0)$. This increment might be positive, and it might be negative, and it might be zero. If you draw pictures for all the possibilities, you'll then need six pictures containing six graphs. You should convince yourself (in each of the six cases) that the difference quotient is a real number representing the slope of a certain **secant line**. You can also extend the hypotenuse of the right triangle in your pictures to show this secant line. (Of course, if you don't remember what is the slope of a line from second grade algebra, you should go back and remind yourself.)

Finally, it's time to try to take a **limit**. You should draw yourself a few pictures in which the limit does not exist (as a real number) for various reasons. The functions

$f : (0, \infty) \to \mathbb{R}$	by $f(x) = \sqrt{x}$,
$g : \mathbb{R} \to \mathbb{R}$	by $g(x) = \lvert x - 2\rvert + 1$, and
$h : \mathbb{R} \to \mathbb{R}$	by $h(x) = x \sin(1/x)$ and when $x \neq 0$ and $h(0) = 0$

might be helpful. Of course, we are primarily interested in situations where the limit does exist, so be sure to draw your six pictures for this case as well.

When the limit does exist, we call the limiting value the **derivative**, or to be more pedantic, **the derivative of the function f at the point** x_0, and we write

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \tag{1}$$

The expression on the left is **Leibniz' notation** for the derivative. Newton would have written

$$\frac{df}{dx} = \frac{df}{dx}(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

but actually, I think he mostly just used the first "fraction" expression, leaving off the evaluation point x_0 as "understood." These days we use both notations (or all three if you like) of course, and each has its advantages.

1.1 Interpretations

After the definition, the most important thing is to understand the interpretations.

1. $f'(x_0)$ is the **slope of the tangent line to the graph of f at the point $(x_0, f(x_0))$.** This is the **geometric interpretation**.

2. $f'(x_0)$ is the **instantaneous rate of change of the value of f at the domain value x_0.** This is the **physical interpretation**. If the domain is measured in units of u which we can express symbolically as

$$[x] = u$$

read "the units of x are u," and the values of f are measured in units of v, then the units of $f'(x_0)$ are u/v. For example, if $[x] =$ seconds and $[f] =$ meters, then $[f'(x_0)] =$ meters per second. But also if if $[x] =$ miles and $[f] =$ degrees Celsius, then $[f'(x_0)] =$ degrees Celsius per mile. Any units will do. Technically, I'm being a little sloppy here and the bracket notation is not properly used for specific units like seconds and meters. What I should say in the first case is $[x] =$ Time $=$ T and $[f] =$ Length $=$ L. In the second example, $[f] =$ Temperature $=$ Temp and $[f'] =$ Temp/L, so that when we indicate units, we are really indicating **type of units** rather than a specific measurement system for those units.

1.2 Differentials and Linear Approximation

Let's start here by rewriting the difference quotient in a little different, but pretty obviously equivalent, form:

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}. \tag{2}$$

Notice that we've replaced the increment h with $h = x - x_0$ where x is the point which is "moving" in the limit. The definition of the limit, which I won't get into in any detail here, says that when x gets close to x_0 (or equivalently when h gets close to zero) the difference quotient gets close to $f'(x_0)$. This means, in particular, that $f(x)$ **must** get close to $f(x_0)$.

The condition

$f(x)$ gets close to $f(x_0)$ when x gets close to x_0

is the condition for f to be **continuous** at x_0. I won't get into the details of that either, but you would do well to think about what it means, and understand the following fundamental assertion:

Theorem 1 *If f is differentiable at x_0, then f is continuous at x_0.*

Differentiability, however, is much stronger than continuity. Remember $g(x) = |x|$ is continuous but not differentiable. The assertion of (2) is not just that $f(x)$ gets close to $f(x_0)$ but that it gets close in a specific way, namely,

$$f(x) \sim f(x_0) + f'(x_0)(x - x_0) \qquad \text{as } x \text{ tends to } x_0. \tag{3}$$

This relation should be read "the value of $f(x)$ is **linearly approximated** by $f(x_0) + f'(x_0)(x - x_0)$ when x is close to x_0."

At this point terminology takes a little bit of a turn toward being confusing, so let me try to spell some things out a little more clearly. As a function of x (with x_0 fixed)

$$\ell(x) = f(x_0) + f'(x_0)(x - x_0) \tag{4}$$

is not technically **linear**. Remember, or pay attention to, this important definition:

Definition 1 *A function $L : V \to W$ from a vector space V to a vector space W is* **linear** *if*

1. *$L(av) = aL(v)$ for every $v \in V$, and*
2. *$L(v + w) = L(v) + L(w)$ for every $v, w \in V$.*

Exercise 1 *The vector space domain of ℓ is $V = \mathbb{R}$ and the vector space co-domain of ℓ is also $W = \mathbb{R}$. Check the two properties required for linearity on ℓ. Distinguish the circumstances under which ℓ is linear or not (linear).*

Nevertheless, whenever $f : (a, b) \to \mathbb{R}$ is differentiable at $x_0 \in (a, b)$ the function $\ell : \mathbb{R} \to \mathbb{R}$ is called the **linear approximation** of f at x_0, even though ℓ is not always linear. Clearly ℓ is closely related to a linear function, and functions like ℓ do have a special name. That name is **affine**.

Definition 2 *A function $\ell : V \to W$ from a vector space V to a vector space W is* **affine** *if there is a linear function $L : V \to W$ such that*

$$\ell(v) = L(v) + w_0.$$

The value w_0 is called an **affine shift** *or an* **affine translation***. The linear function L is called the* **linear part** *of the affine function ℓ.*

So, properly, we should say the function ℓ staring in (3) gives and **affine approximation** of f and (3) is an **affine approximation formula**. But nobody ever seems to say that. They only talk about linear approximation.

Exercise 2 **(a)** *Identify the linear part of the affine function ℓ in the affine approximation formula. Identify the affine shift.*

(b) *Given any linear function $L : V \to W$ and any fixed vectors $v_0 \in V$ and $w_0 \in W$, show that $\ell : V \to W$ by*

$$\ell(v) = L(v - v_0) + w_0$$

is an affine function with linear part L.

Here is something else important:

> *Whenever a function is differentiable, the linear function used to approximate the values near the point where the derivative is calculated, i.e., the linear part of the affine approximation, is called the* **differential***.*

That's a little bit of a mouthful, but it's worth thinking about. It still applies in other cases below which are not just 1-D calculus. Let me repeat it in this special case: The function L which is the linear part of ℓ is called the differential of f at x_0, or the **differential mapping**. The differential is always a linear function. It is also denoted (in the 1-D calculus case)

$$df : \mathbb{R} \to \mathbb{R}.$$

Consequently, another way to express the affine approximation formula is

$$f(x) \sim f(x_0) + df(x - x_0).$$

The following formula should be one you have written down on your own already by now, but I'm going to write it down to make sure you haven't missed it.

$$df(v) = f'(x_0)v.$$

If there is a desire to emphasize the point at which the linear approximation is taking place, a subscript is used and the differential becomes df_{x_0}. Thus, one also writes

$$f(x) \sim f(x_0) + df_{x_0}(x - x_0).$$

Here is a final subtlety probably worthy of note: When we write

$$df(x - x_0) = f'(x_0)(x - x_0)$$

the parentheses around $x - x_0$ on the left and those on the right are representing quite different things. The parentheses in $df(x - x_0)$ are indicating **evaluation of the differential**, that is, function evaluation. The same parentheses on the right are just grouping for multiplication. Function evaluation should *never* be confused with multiplication. Sometimes one emphasizes this distinction by using square brackets for the evaluation of linear functions and writes

$$df[x - x_0] = f'(x_0)(x - x_0).$$

This use of square brackets has nothing to do with the identification of **units** mentioned above. (There are only so many symbols in the world.)

Exercise 3 *Go back through the previous section and identify each appearance of the evaluation of a linear function using round brackets, and rewrite the expression using square brackets to represent this evaluation.*

I have stated in this section just about everything that comes to mind in relation to the definition of the derivative in 1-D calculus. There are probably more things to say, but they escape me at the moment. In any case, this is the prototypical definition of differentiation. All other notions of differentiation are generalizations of this one, in one way or another. Every time you encounter a different kind of differentiation, you should go back through the observations above and try to see how they apply or generalize to the new differentiation. Can you draw pictures? What are the interpretations? Is there an associated differential approximation?

Exercise 4 *Recall that a linear function $L : V \to W$ from a vector space V to a vector space W, of the type defined above, is the primary object of study in* **linear algebra**. *Show that such a linear function always has the following properties:*

(a) $L(0_V) = 0_W$. *(Here* 0_V *denotes the* **zero vector** *in* V *and* 0_W *denotes the zero vector in* W*.)*

(b) $L(av + bw) = aL(v) + bL(w)$ *for any* **scalars** a, b *and vectors* $v, w \in V$*. The expression* $av + bw$ *is called a* **linear combination** *of the vectors* v *and* w*.*

(c) *The parentheses for function evaluation for linear functions, as mentioned above, are sometimes replaced with square brackets. Sometimes when a linear function is applied to a single element of its domain no grouping symbols are used (at all).*

$$L\left(\sum_{j=1}^{k} a_j v_j\right) = \sum_{j-1}^{k} a_j L v_j$$

whenever $a_1, a_2, \ldots, a_k$ *are scalars and* $v_1, v_2, \ldots, v_k \in V$ *(are vectors).*

Exercise 5 *Let* $L : \mathbb{R}^1 \to \mathbb{R}^1$ *be a linear function. Prove that there exists some number* $m \in \mathbb{R}$ *such that* L *has the form*

$$Lx = mx.$$

2 Partial Derivatives

One of the first and easiest generalizations of differentiation is to real valued functions of several real variables. Let's just start with **two** real variables. Some preliminary attention needs to be paid to the domain of such a function. For the domain, we will want an open subset of $\mathbb{R}^2$. Let's call this set $\mathcal{U}$. Thus, we consider a function $f : \mathcal{U} \to \mathbb{R}$. The value of f is written $f(x, y)$ and a point in the domain where we wish to differentiate is $(x_0, y_0) \in \mathcal{U} \subset \mathbb{R}^2$. We fix one of the variables, say y_0, and apply our previous 1-D definition directly:

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

is called the **partial derivative of f in the x-direction** when the limit exists. One thing to note about increasing the number of dimensions in the domain is the following meta-principle:

Anything that can go wrong in one dimension...
can also go wrong and will go wrong in higher dimensions.

As an application, there are of course at least as many ways for this limit to not exist as there were in one dimension.

Exercise 6 *Give examples corresponding to each of the three non-differentiable functions given in the previous section for which the partial derivative $\partial f/\partial x$ does not exist. Some of those examples admit more than one obvious generalization.*

Exercise 7 *Use the limit definition above to calculate $\partial f/\partial x$ when $f : \mathbb{R}^2 \to \mathbb{R}^1$ by $f(x, y) = x^2 + y^2$.*

There is, in principle, another **first partial** (derivative) of a function $f : \mathcal{U} \to \mathbb{R}$ when $\mathcal{U} \subset \mathbb{R}^2$. This is

$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}.$$

This is, of course, the first partial of f with respect to y, when it exists.

2.1 Picture

Drawing the picture illustrating the partial derivative of a function of two variables is perhaps a bit more challenging than the pictures drawn for 1-D calculus. There are some good pictures for this in most calculus books. You can look one up, but it very well likely might be good for you to learn to draw one yourself. Here are some pointers.

1. Start by drawing the graph

$$\mathcal{G} = \{(x, y, f(x, y)) \in \mathbb{R}^3 : (x, y) \in \mathcal{U}\}$$

   of a function. The graph of a function of two variables is a **surface**. The surface is in $\mathbb{R}^3$, so you'll need to start with three axes.[1] You might draw a region $\mathcal{U}$ for the domain in the x, y-plane. Or you might just leave the region $\mathcal{U}$ unspecified in your picture. You can take, for example, $\mathcal{U} = \mathbb{R}^2$, the entire x, y-plane. Again,

[1] A lot of students don't know how to draw a reasonable set of right-handed coordinate axes for $\mathbb{R}^3$. I hope you do. If not, check a book or ask me about it. The positive x-axis should point mostly toward you—out of the paper. The positive y-axis should point mostly to the right, but maybe a little toward you. And the third axis, along which the values of f are taken, should point straight up.

if you lack imagination, start with $f(x, y) = x^2 + y^2$ and take, say, the (open) first quadrant or the **unit disk**

$$B_1(0) = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}.$$

The unit disk is also called the **unit ball**, and the "0" in $B_1(0)$ represents the zero (vector) in $\mathbb{R}^2$, namely $(0, 0)$; the "1" represents the (unit) radius. In your picture, the domain will appear in the x, y-plane as it sits embedded in $\mathbb{R}^3$, so technically the first quadrant will be

$$\{(x, y, 0) : x, y > 0\}.$$

If you use the unit ball, you will draw $\{(x, y, 0) : x^2 + y^2 < 1\}$.

2. Now, take a value of y_0 fixed along the y-axis at a point $(0, y_0, 0)$. Draw the line parallel to the x-axis through $(0, y_0, 0)$. This line contains the points $(x, y_0, 0)$, and hopefully the graph you have drawn indicates positive values for f along this line: $f(x, y_0) > 0$. (This isn't always true, of course, for every function, but it will make your picture easier to draw.)

3. Next you want to draw the set

$$\mathcal{G}_0 = \{(x, y_0, f(x, y_0)) : x \in \mathbb{R} \text{ and } (x, y_0) \in \mathcal{U}\}.$$

This will be a **curve** in the **plane** $y = y_0$ in $\mathbb{R}^3$. The is called the **restriction** of f to the line $y = y_0$. More properly, the curve is the restricted graph, but it corresponds to the restriction of the function f to the line. In any case, it is this restricted graph you need to use to illustrate the partial derivative with respect to x.

4. Don't try to draw secant lines in your picture, just imagine them in your mind. Draw only a tangent line understanding its meaning and its relation to the derivative (see the geometric interpretation of the 1-D derivative).

5. One is interested in the slope of the tangent line to $\mathcal{G}_0$ with respect to the plane $y = y_0$. You can think of the plane $y = y_0$ as having a "x" axis pointing along the line $y = y_0$ and $z = 0$, though this is not the actual x-axis in $\mathbb{R}^3$. Similarly, you can think of the plane $y = y_0$ as having a "z" axis pointing up along $x = 0$ and $y = y_0$.

This reference frame for the slope, and the actual slope of the tangent line to the restriction curve, are difficult to illustrate in your picture, but you it would be good to have them firmly in your mind.

6. Incidentally, I didn't mention that the tangent line will go through the point $(x_0, y_0, f(x_0, y_0))$, and you'll probably want to indicate and label the point (x_0, y_0) or, more properly, $(x_0, y_0, 0)$. It's also usually nice to draw lines (or dashed lines according to your artistic sensibilities) between the following pairs of points: (1) $(0, y_0, 0)$ and $(x_0, y_0, 0)$, (2) $(x_0, 0, 0)$ and $(x_0, y_0, 0)$, and (3) $(x_0, y_0, 0)$ and $(x_0, y_0, f(x_0, y_0))$. I figured you could probably figure that out on your own.

Now, you can draw a similar picture for $\partial f/\partial y$.

2.2 More Variables, More Derivatives, and More Notation

Essentially the same definitions apply if $f : \mathcal{U} \to \mathbb{R}$ with $\mathcal{U} \subset \mathbb{R}^3$ or $\mathcal{U} \subset \mathbb{R}^n$ for $n = 4, 5, 6, \ldots$. For $\mathcal{U} \subset \mathbb{R}^3$, we have

$$
\begin{aligned}
\frac{\partial f}{\partial x}(x_0, y_0, z_0) &= \lim_{h\to 0} \frac{f(x_0 + h, y_0, z_0) - f(x_0, y_0, z_0)}{h} \\
\frac{\partial f}{\partial y}(x_0, y_0, z_0) &= \lim_{h\to 0} \frac{f(x_0, y_0 + h, z_0) - f(x_0, y_0, z_0)}{h} \\
\frac{\partial f}{\partial z}(x_0, y_0, z_0) &= \lim_{h\to 0} \frac{f(x_0, y_0, z_0 + h) - f(x_0, y_0, z_0)}{h}
\end{aligned}
$$

where $(x_0, y_0, z_0) \in \mathcal{U}$ and, as usual, these limits happen to exist. You can imagine how this goes for higher dimensional domains. In particular, if $\mathcal{U} \subset \mathbb{R}^n$, then f will have n first partials (one for each variable). One thing to note is that when $n \geq 3$ one usually starts using subscripts to denote the variables. So for $n = 3$ we write $f = f(x, y, z)$, but for $n = 5$, we write

$$f = f(x_1, x_2, x_3, x_4, x_5),$$

and $f = f(x_1, x_2, \ldots, x_n$ when $\mathcal{U} \subset \mathbb{R}^n$.

You can't draw any reasonable picture for these cases because the graph is in $\mathbb{R}^{n+1}$ which is difficult to draw for $n \geq 3$. You can do a little something, however, if you're willing to stretch your geometric intuition.

Take $(x_0, y_0, z_0) \in \mathcal{U} \subset \mathbb{R}^3$. If we fix y_0 and z_0, then the set of points (x, y_0, z_0) with x free to move makes a line. Now, you can pair this line with a **fourth axis** representing the values of f. This fourth axis is somewhere outside the domain $\mathbb{R}^3$, but if you just pair the line

$$\{(x, y_0, z_0) : x \in \mathbb{R}\}$$

with this fourth axis, you get a (two dimensional) plane in which you can plot the restricted function values x versus $f(x, y_0, z_0)$, and the slope of the tangent line to the resulting curve at $(x_0, f(x_0, y_0, z_0))$ (with respect to this plane) is the value of $\partial f/\partial x(x_0, y_0, z_0)$. Thus, the geometric interpretation goes through in this sense.

There is no real problem with the physical interpretation. Values can change as points move in a domain $\mathcal{U}$ in any dimension. For example, if $[f]$ = Temperature and $[x] = [y] = [z]$ = Length with f measuring the temperature in a room, then

$$\frac{\partial f}{\partial z}(x_0, y_0, z_0)$$

measures the instantaneous rate at which the temperature changes as one moves straight up from (or through) the point (x_0, y_0, z_0).

I didn't mention it in the 1-D case, but derivatives of functions (when considered as dependent on the point of evaluation (x_0, y_0, z_0) for the derivative) can **themselves** be considered as functions—and often differentiated again. That is, for a function of two variables $f = f(x, y)$, we can take

$$\frac{\partial f}{\partial y}$$

and consider this partial derivative as another function

$$\frac{\partial f}{\partial y} : \mathcal{U} \to \mathbb{R}.$$

And this function often has derivatives as well, that is, partial derivatives:

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) \qquad \text{and} \qquad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right).$$

These two derivatives are usually indicated using a version of Newton's notation:

$$\frac{\partial^2 f}{\partial x \partial y} \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2}.$$

In general,

$$\frac{\partial^2 f}{\partial x \partial y} \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y \partial x}$$

take different values. Notice the difference is the order in which the derivatives are taken. These are called **mixed partials**, and the other ones are called **homogeneous**

partials. There is, however, a well-known theorem giving conditions under which the order of differentiation does not matter and all mixed partials are equal (as long as they are taken with respect to the same variables). I won't state the mixed partials theorem here, but you can look it up and see what conditions it requires of a function f of several variables.

Exercise 8 *How many different second partials are there for $f = f(x, y, z)$ (assuming differentiability)? How does your answer change if you assume f satisfies the hypotheses of the mixed partials theorem?*

When there are lots of variables (and lots of partial derivatives and lots of subscripts) an alternative notation is used. Let me quickly tell you about it.

First of all, there is no real analog of Leibniz' notation for partial derivatives. The closest things are

$$f_x \quad \text{and} \quad D_x f \qquad \text{for} \qquad \frac{\partial f}{\partial x}.$$

A variation on the second one, when using subscripts, is

$$D_j f \qquad \text{for} \qquad D_{x_j} f = \frac{\partial f}{\partial x_j}.$$

All of these notations become a bit cumbersome to express something like

> *the 7-th order partial derivative of f, two times with respect to x_4, two times with respect to x_6, and three times with respect to x_8.*

To be explicit, we would have

$$\frac{\partial^7 f}{\partial x_8^3 \partial x_6^2 \partial x_4^2} = D_{x_8} D_{x_8} D_{x_8} D_{x_6} D_{x_6} D_{x_4} D_{x_4} f = f_{x_4 x_4 x_6 x_6 x_8 x_8 x_8}.$$

When things get this complicated, one almost always assumes the conclusion of the mixed partials theorem, so no effort is made to distinguish between

$$\frac{\partial^7 f}{\partial x_8^3 \partial x_6^2 \partial x_4^2} \qquad \text{and} \qquad \frac{\partial^7 f}{\partial x_8 \partial x_6 \partial x_8 \partial x_4 \partial x_8 \partial x_4 \partial x_6}.$$

With this in mind, here is the "simple" notation used for this kind of thing: A **multi-index** is like a point in $\mathbb{R}^n$, but all the entries are natural numbers (including 0). Thus, if $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, then a multi-index is an ordered n-tuple in $\mathbb{N}^n$. For example, $(0, 0, 0, 2, 0, 2, 0, 3)$ is a multi-index of order eight. If $f : \mathcal{U} \to \mathbb{R}$ and $\mathcal{U} \subset \mathbb{R}^8$, then the derivative above is expressed in **multi-index notation** as

$$D^{(0,0,0,2,0,2,0,3)} f.$$

This means "no derivatives in the first, second, third, fifth, or seventh variables, two partials with respect to x_4, two partials with respect to x_6, and three partials with respect to x_8." This may not strike you as a stupendous improvement, but in general, if $\beta = (\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{N}^n$ is a multi-index of order n, then taking β_j derivatives with respect to x_j for each $j = 1, 2, \ldots, n$ can be written as

$$D^\beta f,$$

and that's pretty simple. There are other really good reasons for the use of multi-indices as well. Let me mention one. You may be familiar with the Taylor series expansion for a function of one variable:

$$\sum_{j=0}^{\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j.$$

There is a version of this series expansion for functions of several variables (and a Taylor approximation theorem to go along with it). Writing down the expansion without multi-indices is a nightmare. With multi-indices we can write for a function $f = f(x_1, x_2, \ldots, x_n)$

$$\sum_{\beta \in \mathbb{N}^n} \frac{D^\beta f(\mathbf{p})}{\beta!} (\mathbf{x} - \mathbf{p})^\beta.$$

Here we are writing $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and the expansion is at the point $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ (which plays the role of x_0 from the 1-D case). Aside from that, the formulas are essentially identical. Of course, you need to know how to calculate the factorial of a multi-index, and you need to understand multi-index powers of vector variables, but learning these things is a small price to pay to avoid the nightmare of having to do multivariable power series without knowing them.

3 Directional Derivatives

These are also derivatives of a function $f : \mathcal{U} \to \mathbb{R}$ with $\mathcal{U}$ an open subset of $\mathbb{R}^n$. To begin the discussion, let's back up to the definition of partial derivatives (and even to the case of $\mathcal{U} \subset \mathbb{R}^2$) and introduce (or point out) the role played by **parametric curves** and **composition**. We can think of the difference quotient

$$\frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

as constructed using the composition of the function f and the parameterized curve $\gamma(t) = (x_0 + t, y_0)$. That is,

$$\frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} = \frac{f \circ \gamma(h) - f \circ \gamma(0)}{h}.$$

You should have a picture in your mind (or on your paper) at this point. It's a picture of the image of γ in the domain $\mathcal{U}$ of f. That simple picture is an image of a directed segment starting at x_0 and pointing horizontally to the right.

Similarly, replacing $\gamma(t) = (x_0 + t, y_0)$ with $\gamma(t) = (x_0, y_0 + t)$ we get

$$\frac{\partial f}{\partial y} = \lim_{t \to 0} \frac{f \circ \gamma(t) - f \circ \gamma(0)}{t}.$$

Directional derivatives generalize this construction to any direction $(\cos\theta, \sin\theta)$ starting at (x_0, y_0). To be precise, the directional derivative of $f : \mathcal{U} \to \mathbb{R}$ in the direction $(\cos\theta, \sin\theta)$ at $(x_0, y_0) \in \mathcal{U}$ is

$$\frac{d}{dt} f \circ \gamma(t)\Big|_{t=0} = \lim t \to 0 \frac{f \circ \gamma(t) - f \circ \gamma(0)}{t} \tag{5}$$

where $\gamma(t) = (x_0 + t\cos\theta, y_0 + t\sin\theta)$.

Exercise 9 *Prove the equality in (5).*

That is the definition of a directional derivative for $f : \mathcal{U} \to \mathbb{R}$ when $\mathcal{U} \subset \mathbb{R}^2$.

I need to say (at least) two more things. First, I should introduce the notation for directional derivatives and second I should address the case $f : \mathcal{U} \to \mathbb{R}$ when $\mathcal{U} \subset \mathbb{R}^n$. If you understand the 2-D case above, this should all be easy/straightforward.

Definition 3 *Given*

1. *a unit vector* $\mathbf{u} \in \mathbb{R}^n$*, and*

2. *a function* $f : \mathcal{U} \to \mathbb{R}$ *defined on an open set* $\mathcal{U} \subset \mathbb{R}^n$

the **directional derivative** *of f in the direction $\mathbf{u}$ at the point $\mathbf{x} \in \mathcal{U}$ is*

$$D_{\mathbf{u}} f(\mathbf{x}) = \frac{d}{dt} f \circ \gamma(t)\Big|_{t=0} = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{u}) - f(\mathbf{x})}{h}$$

where $\gamma : \mathbb{R} \to \mathbb{R}^n$ by $\gamma(t) = \mathbf{x} + t\mathbf{u}$ (when the derivative and the limit exist).

There are a few more things to say about this. Here is one of them:

$$D_{\mathbf{u}} f(\mathbf{x}) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\mathbf{x})\, u_j \tag{6}$$

when $\mathbf{u} = (u_1, u_2, \ldots, u_n)$.

Exercise 10 *Prove the formula (6) giving the directional derivative in terms of the partial derivatives using the chain rule.*

Exercise 11 *Rewrite (6) using the Euclidean dot product.*

Exercise 12 *Draw a picture in the domain of a function $f : \mathbb{R}^3 \to \mathbb{R}$ illustrating the construction of a directional derivative.*

Maybe one last thing: Many texts, especially elementary calculus texts, restrict directional derivatives to only apply to **unit vector** directions $\mathbf{u}$. And this makes some sense. There can arise a conflict of notation concerning directional derivatives along the following lines. Some elementary calculus texts define the directional derivative of a function f in the direction of any arbitrary nonzero vector $\mathbf{v} \in \mathbb{R}^n \backslash \{0\}$ to be the directional derivative of f in the direction $\mathbf{u} = \mathbf{v}/|\mathbf{v}|$. Thus, one can encounter

$$D_{\mathbf{v}} f = D_{\mathbf{u}} f \qquad \text{where } \mathbf{u} = \mathbf{v}/|\mathbf{v}|.$$

In other, sometimes more advanced, contexts it is natural to define

$$D_{\mathbf{v}} f = Df \cdot \mathbf{v} \tag{7}$$

where Df is the vector of first partial derivatives

$$\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right). \tag{8}$$

The approach taken in elementary calculus preserves the interpretation of the directional derivative as the rate of change in the direction of $\mathbf{v}$ (with respect to unit speed displacement in the domain). However, the formula in (7) has the advantage of being linear in $\mathbf{v}$. Using the definition (7) one must interpret the directional derivative to be measuring the rate of change of the function f with respect to motion at speed $|\mathbf{v}|$. In either case, one needs to know what is meant by the notation $D_{\mathbf{v}} f$ when $\mathbf{v}$ is not a unit vector.

4 Total Derivatives and the Gradient

The vector of first partial derivatives appearing in (8) is called the **gradient** of f, and it also represents the **total derivative** of a function $f : \mathcal{U} \to \mathbb{R}$ when $\mathcal{U} \subset \mathbb{R}^n$. Another common notation for the gradient is

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

The gradient $Df : \mathcal{U} \to \mathbb{R}^n$ is an example of a **vector field** on $\mathcal{U} \subset \mathbb{R}^n$, and we will discuss such functions in more detail below.

You may recall that the gradient has some nice properties:

1. The gradient is a vector pointing in the direction of maximum rate of increase of the value of f at a point.

2. The gradient $Df(\mathbf{p})$ is orthogonal to the level set

$$\mathcal{L}_c = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = c\}$$

   at the point $\mathbf{p} \in \mathcal{U}$ with $c = f(\mathbf{p})$.

It is not our purpose to give a full review of such properties. We're mostly interested in solidifying the definitions and the basic meanings of various derivatives. You should feel free, however, to prove these properties and draw appropriate pictures illustrating them.

4.1 Affine Approximation

We do want to address what makes the gradient Df a **total derivative**. The basic point here is that this vector is integral to what it means for a function $f : \mathcal{U} \to \mathbb{R}$ with $\mathcal{U} \subset \mathbb{R}^n$ to be **differentiable**.

...

5 Vector Valued Functions of One Variable

I mentioned above that partial derivatives of $f : \mathcal{U} \to \mathbb{R}$ where $\mathcal{U} \subset \mathbb{R}^n$ were among the first and easiest generalization of 1-D calculus differentiation. There is another generalization which is, in some respects, even simpler. This is for functions

$$\mathbf{r} : (a, b) \to \mathbb{R}^k.$$

These functions, furthermore, play a central role in physics (analyzing the motion of point masses, extended bodies, and other things in space, usually $\mathbb{R}^2$ or $\mathbb{R}^3$) and in the study of **ordinary differential equations** which we are going to review in this course.

...

6 Vector Valued Functions of Several Variables

...

7 Summary and Conclusion

This completes our catalog of the first and easiest kinds of real differentiation. The next step might be to consider the differentiation of real valued functions defined on (real) surfaces or the analog of vector valued functions on (real) surfaces, which are called tensors.

We are taking a different route and considering a generalization of 1-D differentiation to functions $f : \mathcal{U} \to \mathbb{C}$ where $\mathcal{U}$ is an open subset of $\mathbb{C}$ using

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}.$$

This looks almost identical to our alternative form for the 1-D derivative (2), but we shall see it is quite different. One should expect comparison to the 1-D case, of course. But in another sense, this kind of differentiation should have certain things to compare to the differentiation of sections on real valued functions of several variables and the last one on vector valued functions of several variables. This is because in the domain a complex variable $z = x + iy$ is rather like a point/vector valued argument $(x, y) \in \mathbb{R}^2$, and a complex value $f(z) = f_1(z) + if_2(z)$ is also rather like a real vector value (u, v) where $u = f_1$ and $v = f_2$. These comparisons form a large part of the material we want to cover on complex analysis. But, of course, you need to understand the real differentiations before you can understand and appreciate the comparisons.