# Differentiation in Lots of Dimensions

*Nueva Multivariable Calculus*

We've now hacked our way through some basic linear algebra. We don't have a true understanding of that subject—but, as I've said many times, this isn't a linear algebra course. This is a multivariable calculus course. So, having completed the preliminaries, let's start actually thinking about how to do **calculus in higher dimensions**!

## Partial Derivatives

OK, real fast, we're going to learn how to take what's called a **partial derivative**. Then we'll talk in more depth about what it *means* to think about derivatives and slopes in higher dimensions. Ordinarily I'd do it in the opposite order—but taking partial derivatives is a very easy skill, so let's get it out of the way, quickly and clearly.

It's really no different than taking ordinary derivatives. I vividly remember my father explaining partial differentiation to me while changing lanes on I-81 in Syracuse in the summer of 2006—that's all the time it took—and being disappointed at how simple it was.

So. Partial derivatives. You just take a normal derivative, but you use a fancy symbol, and you call it a "partial derivative."

Example. Here's a function:

$$f : \mathbb{R}^3 \to \mathbb{R}^1$$
$$f(x, y, z) = x^2 - 3y + 2xy - \sin(z) + e^{y^3}$$

To take its partial derivative with respect to $x$, we just treat $x$ as the variable, and everything else as a constant. And instead of writing it as $\frac{df}{dx}$, we write it as $\frac{\partial f}{\partial x}$:

$$\begin{aligned}
\substack{\text{the partial derivative of } f \\ \text{with respect to } x} &= \frac{\partial f}{\partial x} \\
&= \frac{\partial}{\partial x} f(x, y, z) \\
&= \frac{\partial}{\partial x}\left[x^2 - 3y + 2xy - \sin(z) + e^{y^3}\right] \\
&= 2x - 0 + 2y - 0 + 0 \\
&= 2x + 2y
\end{aligned}$$

So we just pretend that everything *other* than the variable we're differentiating against is constant. We pronounce this "partial $f$ partial $x$." (As opposed to how we pronounce $\frac{df}{dx}$: "d f d x.") And we often refer to partial derivatives as just "partials." Note how this cool symbol, $\partial$, is unique to partial derivatives! It's probably loosely modeled after the letter "d" or the Greek $\delta$, but it's its own thing!

We could take the partial derivative of this function with respect to $y$:

$$\begin{aligned}
\substack{\text{the partial derivative of } f \\ \text{with respect to } y} &= \frac{\partial f}{\partial y} \\
&= \frac{\partial}{\partial y} f(x, y, z) \\
&= \frac{\partial}{\partial y}\left[x^2 - 3y + 2xy - \sin(z) + e^{y^3}\right] \\
&= 0 - 3 + 2x - 0 + e^{y^3} \cdot 3y^2 \\
&= 3y^2 e^{y^3} + 2x - 3
\end{aligned}$$

And we could take the partial derivative of this function with respect to $z$:

$$\begin{aligned}\underset{\text{with respect to } z}{\text{the partial derivative of } f} &= \frac{\partial f}{\partial z} \\ &= \frac{\partial}{\partial z} f(x, y, z) \\ &= \frac{\partial}{\partial z}\left[x^2 - 3y + 2xy - \sin(z) + e^{y^3}\right] \\ &= 0 - 0 + 0 - \cos(z) + 0 \\ &= -\cos(z)\end{aligned}$$

For that matter, we could even take a partial derivative with respect to some variable that doesn't even show up! The partial will then just be zero, since the function will just be constant with respect to that variable. So, for example, if we partially-differentiate this function with respect to $q$:

$$\begin{aligned}\underset{\text{with respect to } q}{\text{the partial derivative of } f} &= \frac{\partial f}{\partial q} \\ &= \frac{\partial}{\partial q} f(x, y, z) \\ &= \frac{\partial}{\partial q}\left[x^2 - 3y + 2xy - \sin(z) + e^{y^3}\right] \\ &= 0\end{aligned}$$

So. That's it, computationally.

Sometimes people use subscripts to indicate partial derivatives, e.g.:

$$\frac{\partial f}{\partial x} = f_x$$

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

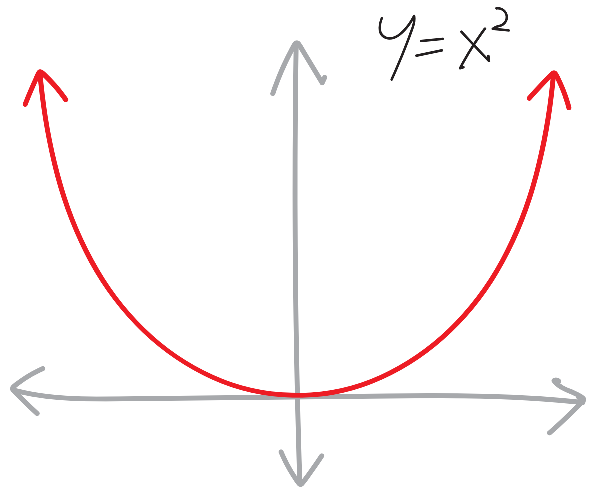
I don't love this notation, but nor do I abhor it. (Occasionally it's convenient.)

There's a very important question we haven't answered, or even asked: what *is* a partial derivative? We can compute them, sure, but what *are* they? And how do they relate to our bigger questions: what, exactly, is the slope of a high-dimensional function? How can we think about them? How can we visualize them? How do we calculate them? In what ways are they distinct from the slopes of lower-dimensional functions?

## Derivatives in one dimension

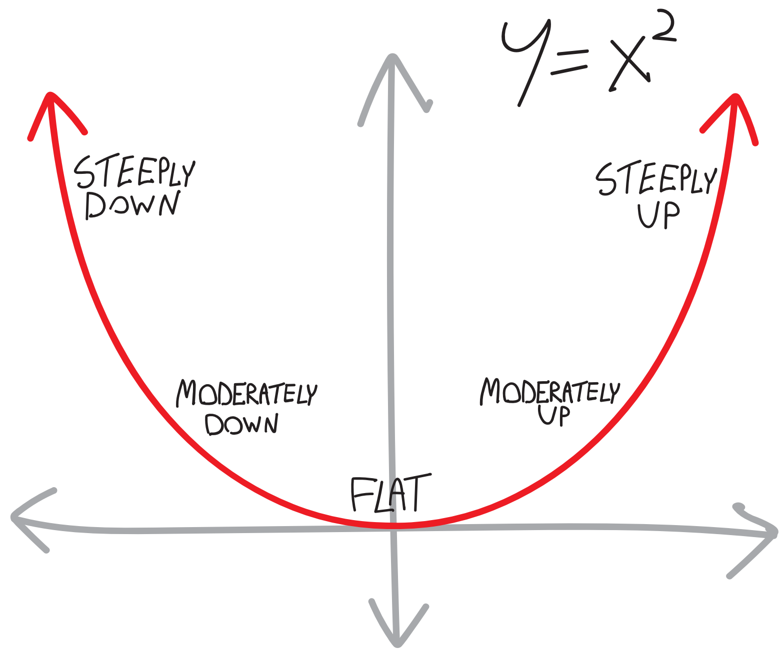
But first, let's remind ourselves about how derivatives and slopes work in one dimension! (Groan all you want, but I haven't talked about calculus with you all in quite some time! And it's always important to be on the same page before turning the page, if you will.)

Let's suppose we have our favorite function, a parabola:



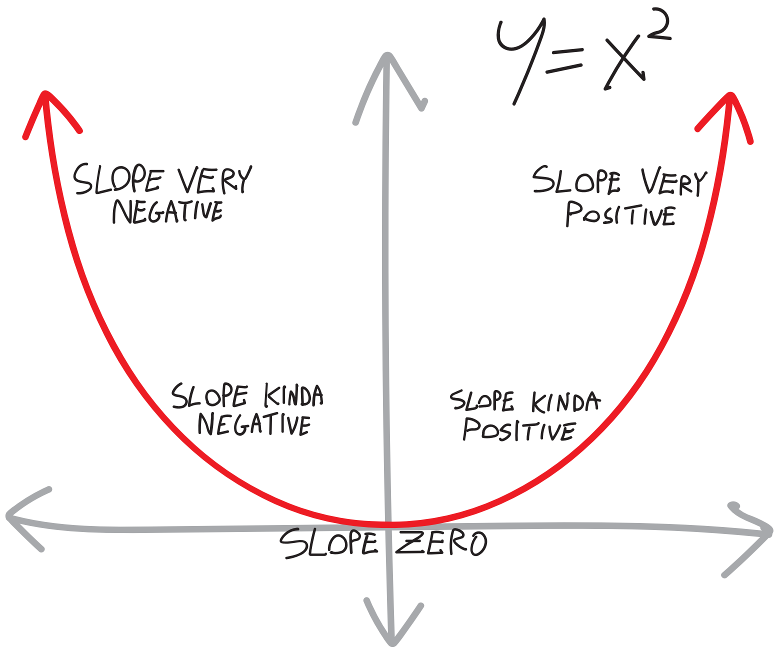

What's going on with its slope?

- At the far left, the function is going down steeply
- Then it goes down less steeply
- Then it's flat for a moment
- Then it starts going up, shallowly
- Then it starts going up steeply

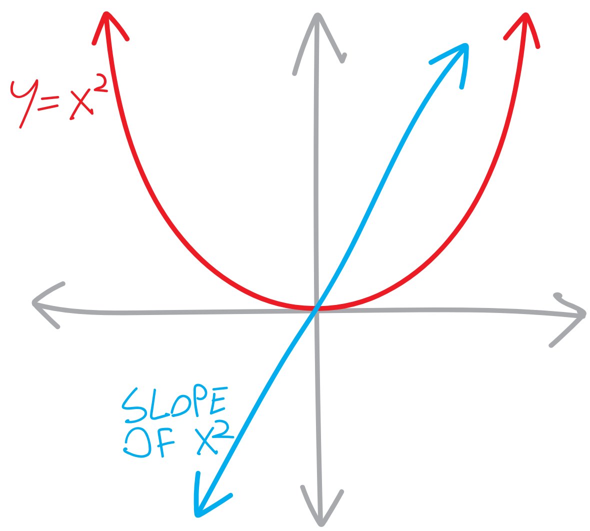



So then, if we think about what its slope is as a number:

- At the far left, the function is going down steeply, so its slope is very negative
- Then it goes down less steeply, so the slope is still negative, but less so
- Then it's flat for a moment, at the minimum, so the slope is 0 there
- Then it starts going up, shallowly, so the slope is positive, but only slightly
- Then it starts going up steeply, so the slope is very positive!




So, we can doodle the slope on the same axes as the original function:




Pictures are nice. What about the slope of $x^2$, as an equation? We know, from our extensive experience calculating slopes/derivatives, that the slope/derivative of $x^2$ is:

$$f'(x) = 2x$$

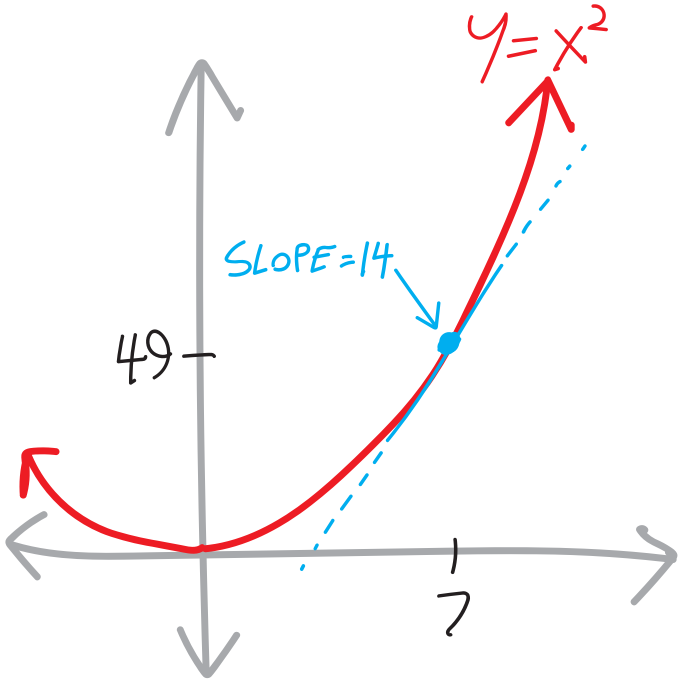
That equation tells us the slope at *any* point $x$. It's just like how the original equation, $y = x^2$, tells us the $y$-value at *any* point $x$.

$$\begin{array}{c}\text{the slope of } f(x) = x^2 \\ \text{at any point } x\end{array} \quad = \quad \underbrace{\frac{df}{dx} \quad = \quad f'(x)}_{\text{the most common notations}} \quad = 2x$$

If, for some reason, we want to find the slope at a *specific* point, we can just plug in a number:

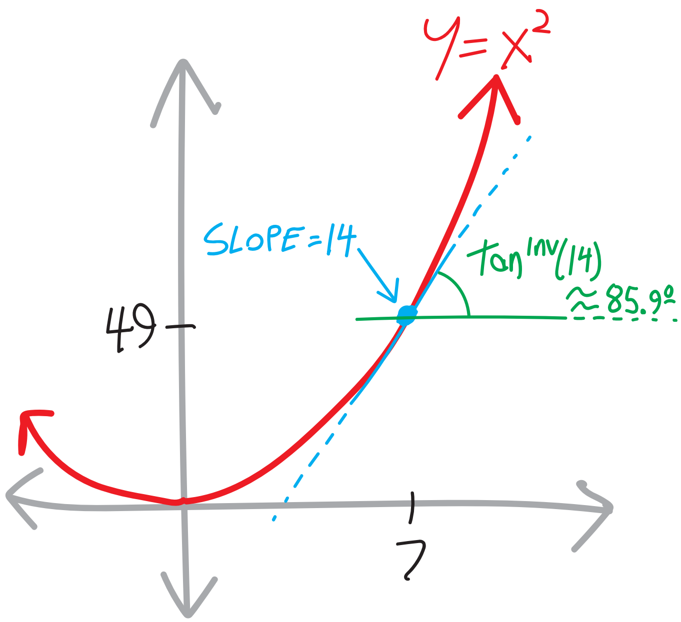
$$\begin{aligned}\begin{array}{c}\text{the slope of } f(x), \\ \text{at the specific point } x = 7\end{array} &= 2 \cdot 7 \\ &= 14\end{aligned}$$

So at the point $x = 7$, the function $f(x) = x^2$ is going up fourteen units for every one unit it's going to the right.




Alternatively, if we want to think of slope as an angle (why not?), we can make a right triangle with legs of length 14 (in the up direction) and 1 (in the right), and use inverse tangent to calculate it:

$$\begin{aligned}\substack{\text{the slope of } f(x),\\ \text{at the specific point } x=7\\ \text{as an angle}} &= \tan^{\text{inv}}\left(\substack{\text{the slope of } f(x),\\ \text{at the specific point } x=7}\right)\\ &= \tan^{\text{inv}}(14)\\ &\approx 1.499 \text{ radians}\\ &\approx 85.91^\circ\end{aligned}$$




Yay! All this stuff is fun and beautiful.

## Derivatives in more than one dimension

That's enough math for today. Let's go skiing! As a younger adventurer of the abstract, I spent a lot of time on the slopes of Whiteface Mountain, the claims to fame of which are having hosted the Winter Olympics in 1932 and 1980 as well as having the highest vertical drop of any resort on the East Coast. I'm particularly fond of *Skyward*, the route used for the "Downhill," which is the Olympic event that basically consists of "ski as fast as you possibly can." You can go really, really fast on Skyward! It's pretty steep!

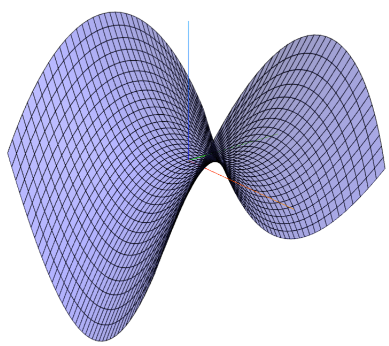
Or... is it??? I mean, sure, if you face downhill in the steepest direction, then yeah, it's pretty steep! But sometimes it gets icy, or barren, and you don't really want to do that. So you want to switchback, side to side, at an angle to the steepest direction. Then it's less steep! In fact, if you face directly to the side, then it's flat! And if you turn all the way around, then it's uphill! The point being: *how steep the slope is depends on which direction you're looking in.*

Slope, then, is not always a single number. In higher-dimensional situations, it's also a function of which direction we're looking in.

Let's use a specific (and more algebraic) example. Here's a function:

$$f : \mathbb{R}^2 \to \mathbb{R}^1$$

$$f(x, y) = x^2 - y^2$$



This is a **hyperbolic paraboloid**—a really cool shape that looks basically like a Pringles chip. How can we think about its slope?

Suppose we're standing somewhere on it. Depending on which direction we're facing, the slope will be different! We can stand in a specific point, rotate ourselves 360 degrees, and the slope will change as we turn. Sometimes it'll be going steeply up, sometimes it'll be going steeply down, sometimes it'll be shallow, sometimes it'll be flat.

In 1D calc, we don't have this problem! If we're talking about a one-dimensional (two-dimensional?) function from $\mathbb{R}^1$ to $\mathbb{R}^1$, every point only has a single slope! Or rather, maybe each point has two slopes: one facing in positive direction, and the other facing in the negative direction. But by convention, in one-dimensional calculus, we always assume we're facing in the positive direction. (And the slope in the negative direction is just the negative of the slope in the positive direction.)

So how can we get some sort of mathematical/symbolic grasp on what the slope of $f(x, y) = x^2 - y^2$ is?

Let's back up a bit. We can think about its slope in a moment, but what about its derivatives? We can find the two partial derivatives:

$$\frac{\partial}{\partial x}\left(x^2 - y^2\right) = 2x$$
$$\frac{\partial}{\partial y}\left(x^2 - y^2\right) = -2y$$

But what about the *full* derivative? Is that even a concept that exists for multivariable functions? Yes! The usual way we think of it is this. If we want to take the derivative of this function—not the partial derivative, *the* derivative—it's just both the partial derivatives, nicely organized into a vector:

$$\begin{aligned} f'(x,y) &= \left\langle \frac{\partial f}{\partial x},\ \frac{\partial f}{\partial y} \right\rangle \\ &= \left\langle \frac{\partial}{\partial x}\left(x^2-y^2\right),\ \frac{\partial}{\partial y}\left(x^2-y^2\right) \right\rangle \\ &= \langle 2x,\ -2y \rangle \end{aligned}$$

Tragically, this means we can't visualize it! The original function only required three dimensions to visualize (two inputs and one output):

$$f : \mathbb{R}^2 \to \mathbb{R}^1$$

But its derivative is a function in FOUR dimensions! Two inputs, two outputs:

$$f' : \mathbb{R}^2 \to \mathbb{R}^2$$

So we definitely can't visualize its derivative. In 1D we were drawing all those nice pictures. In 2D, we can't![1]

It gets worse. Let's imagine we have an even crazier function, that takes three inputs and begets four outputs:

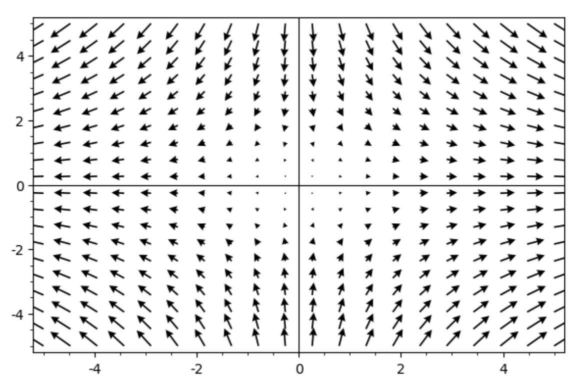
$$g : \mathbb{R}^3 \to \mathbb{R}^4$$

$$g(x,y,z) = \begin{bmatrix} x^2+3 \\ x+y+z \\ \sin(z)-2x^3 \\ e^x+72y \end{bmatrix}$$

The function itself is already impossible to visualize—we definitely don't have the seven dimensions that would require. How about its derivative? We definitely won't be able to visualize it, but how do we even calculate it?

Here's the idea, which is the most general idea of taking a derivative: **the derivative is the matrix of all the partial derivatives.** (This is true of lower-dimensional functions, too.) Every pair of one input dimension together with one output dimension gives us a partial derivative! For any combination of an input and an output, we can measure the slope—we can measure how much that output changes when we change that input. And we can organize them all nicely together using a matrix!

[1]OK, there are bad ways to visualize things that are more than three dimensional, as we've discussed. All those ways are mediocre—none are perfect. One way you could visualize this derivative is with what's known as a "**vector field**," where you plop onto the 2D plane little vectory arrows corresponding to the output vector. Here's what that'd look like for this function:



So, in this case, we have:

$$\text{the derivative } = \begin{bmatrix} \frac{\partial}{\partial x}\left(x^2+3\right) & \frac{\partial}{\partial y}\left(x^2+3\right) & \frac{\partial}{\partial z}\left(x^2+3\right) \\ \frac{\partial}{\partial x}\left(x+y+z\right) & \frac{\partial}{\partial y}\left(x+y+z\right) & \frac{\partial}{\partial z}\left(x+y+z\right) \\ \frac{\partial}{\partial x}\left(\sin(z)-2x^3\right) & \frac{\partial}{\partial y}\left(\sin(z)-2x^3\right) & \frac{\partial}{\partial z}\left(\sin(z)-2x^3\right) \\ \frac{\partial}{\partial x}\left(e^x+72y\right) & \frac{\partial}{\partial y}\left(e^x+72y\right) & \frac{\partial}{\partial z}\left(e^x+72y\right) \end{bmatrix}$$

$$= \begin{bmatrix} 2x & 0 & 0 \\ 1 & 1 & 1 \\ -6x^2 & 0 & \cos(z) \\ e^x & 72 & 0 \end{bmatrix}$$

The standard convention is that the input dimensions go along the horizontal, and the output dimensions go along the vertical.

So, in general, we have:

$$\textbf{the derivative } = \text{the matrix of all the partial derivatives}$$

$$= \begin{bmatrix} \frac{\partial(\text{the first output dimension})}{\partial(\text{the first input dimension})} & \frac{\partial(\text{the first output dimension})}{\partial(\text{the second input dimension})} & \frac{\partial(\text{the first output dimension})}{\partial(\text{the third input dimension})} & \cdots \\ \frac{\partial(\text{the second output dimension})}{\partial(\text{the first input dimension})} & \frac{\partial(\text{the second output dimension})}{\partial(\text{the second input dimension})} & \frac{\partial(\text{the second output dimension})}{\partial(\text{the third input dimension})} & \cdots \\ \frac{\partial(\text{the third output dimension})}{\partial(\text{the first input dimension})} & \frac{\partial(\text{the third output dimension})}{\partial(\text{the second input dimension})} & \frac{\partial(\text{the third output dimension})}{\partial(\text{the third input dimension})} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial(\text{every output dimension})}{\partial(\text{every input dimension})} \end{bmatrix}$$

So a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ will have $nm$ partial derivatives, and a derivative matrix with $n$ rows and $m$ columns. People usually call this **the Jacobian** or **the Jacobian matrix**, but I prefer to just call it **the derivative**, or maybe **the derivative matrix**. It's the most natural generalization of our one-dimensional idea of differentiation to higher dimensions.[2]

It's exactly the same as what we've been doing in 1D calc. It's just that in 1D calc there's only one input dimension and only one output dimension, so there's only one "partial" derivative, and the derivative itself we can think of as being a one-column, one-row, one-entry matrix! For instance:

$$\begin{aligned} f : \mathbb{R}^1 &\to \mathbb{R}^1 \\ f(x) &= x^2 \\ \frac{\partial f}{\partial x} &= 2x \\ f'(x) &= \begin{bmatrix} 2x \end{bmatrix} \end{aligned}$$

So that's how we can calculate derivatives in higher dimensions.

[2] Robert Ghrist, an idol of mine who among many other distinctions teaches UPenn's multivariable calculus class, also calls it that.

What about how we visualize them? In the first example I showed you, the derivative was a vector—so we can think of that as being some point in high-dimensional space. We know how we *could* visualize it, if somehow we could visualize high-dimensional space. (In particular, it was a two-entry vector, so we could think of that as being a point in 2D space.) But in this second, more complicated example, the derivative is a matrix. What does that even mean in terms of "dimensions"?!?? Or trying to visualize things?!?! How do we visualize matrices??!!? Vectors we can think of as just being objects in high-dimensional space—magnitudes and directions; little arrows pointing from the origin—but matrices?!?

Plus, all of this is only the first derivative! What if we take second or third or fourth, etc., derivatives? I guess we take each partial derivative, differentiate it a second time, and get another matrix of partial derivatives? So we have matrices filled with matrices?!?

## Derivatives in specific directions

That all seems very complicated. Let's back up yet further. What *is* a partial derivative? We saw how easy it is to calculate them, but what *are* they? How do they relate to slopes? How can we calculate slopes of high-dimensional functions—like, actual numbers?

In class, I made you calculate a bunch of specific slopes of some functions from $\mathbb{R}^2$ to $\mathbb{R}^1$. Most of you figured this out in various creative, complicated ways! I was proud.

The most common way you all did it was to parameterize/slice the functions in some direction (to "one-dimensionalize" them), and then take a one-dimensional derivative. That's fine! That works. On the other hand, it does require a bit of effort. There are also a lot of potential pitfalls: for instance, if you just want a function for the slope at any input point $(x, y)$, then you somehow have to convert the 1D parameterization function, in terms of $t$, back to a 2D function in terms of $x$ and $y$. And depending on what parameterization you choose, the "length" of one unit of $t$ might not be the same as the length of one unit of $x$ and $y$, so you need to account for that. (This is why, for instance, so many of you were willy-nilly dividing by $\sqrt{2}$ when we did these problems in class—you had to account for the fact that in your paramaterization, one unit of $t$ was $\sqrt{2}$ units of $x$ and $y$.)

Good news: there's an easier way! To fully write up why and how this works would take more time then I think I have the ability for right now—*very sad*—but here's the basic idea: we take a **weighted average of the partial derivatives, in which the weights are the components of the direction/unit vector** in which we're looking:

$$\substack{\text{the slope}\\ \text{in a certain direction}} = \left(\substack{\text{the partial}\\ \text{in the } x\text{-direction}}\right)\cdot\left(\substack{\text{the unit/direction vector}\\ \text{component in the } x\text{-direction}}\right) + \left(\substack{\text{the partial}\\ \text{in the } y\text{-direction}}\right)\cdot\left(\substack{\text{the unit/direction vector}\\ \text{component in the } y\text{-direction}}\right)$$

The components of the direction/unit vector all add up to 1, so it's kind of like they're all weights that add up to 100%! Well, they don't add up to 1 *directly*—they add up if we square them, add them up, and then square-root them. So it's the Pythagorean Theorem equivalent of a bunch of things that add up to 1! (People call this operation **Pythagorean addition**.)

Another way of putting this is to describe this as a dot product! We're dot-producting the derivative vector/the vector of the partial derivatives, with a unit vector:

$$\substack{\text{the slope}\\ \text{in a certain direction}} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} \text{the unit/} \\ \text{direction vector} \\ \text{in that direction} \end{bmatrix}$$

If you have a good, solid, physics-inspired intuition for the dot product, you can think about how this works—we're projecting the derivative vector onto a unit vector.

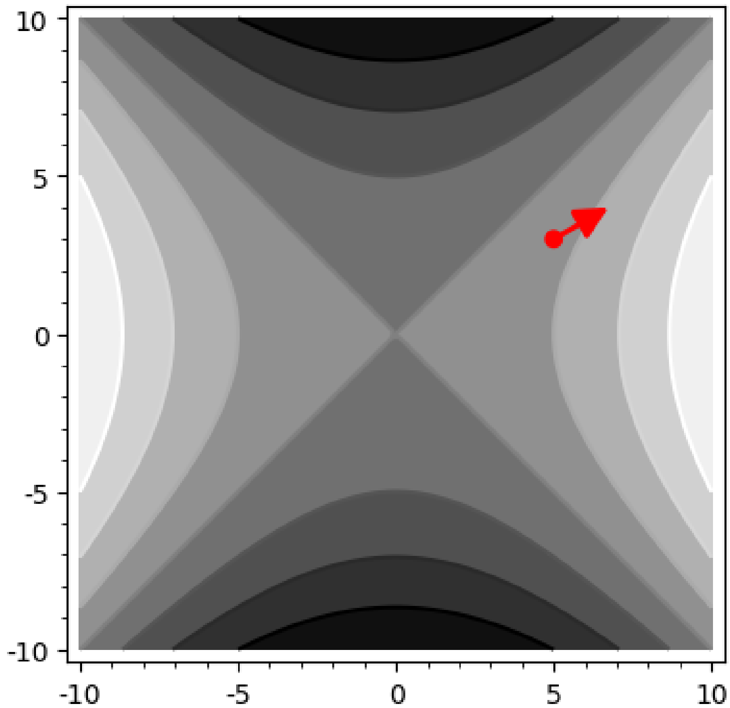
Anyway, let's get back to a concrete example. Where were we? We were skiing on a hyperbolic paraboloid:

$$f : \mathbb{R}^2 \to \mathbb{R}^1$$

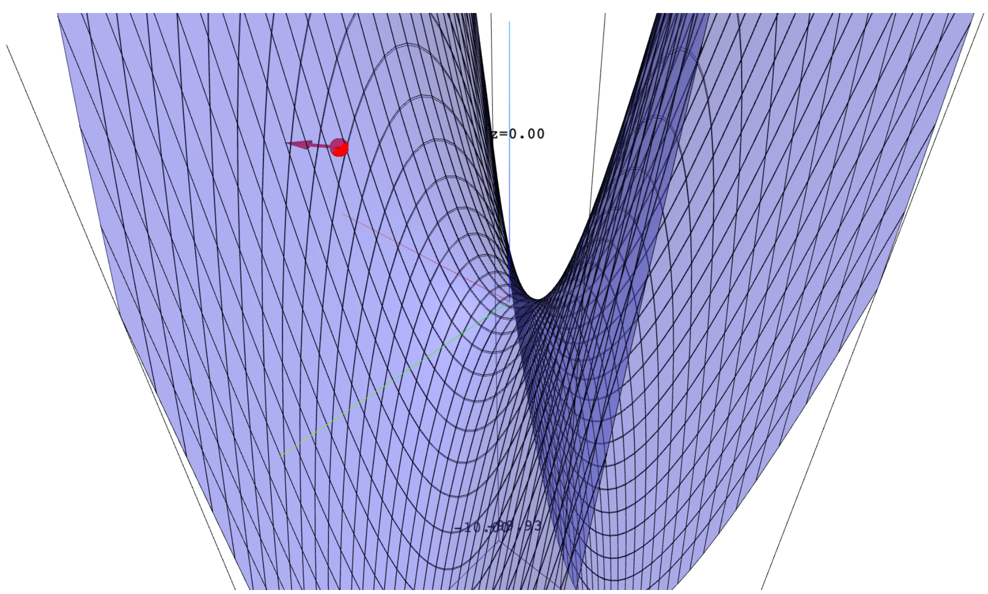
$$f(x, y) = x^2 - y^2$$

How steep is the slope? Well that depends on what direction we're looking in, and also on where we are. Suppose we're looking in the $\pi/6$ direction (on the $xy$-plane), and we're directly above the point $(5, 3)$. (So the full $(x, y, z)$-coordinates of our location on the hyperbolic paraboloid are $\left(5, 3, 5^2 - 3^2\right) = (5, 3, 16)$.)

Here's a top-down view of the situation, with the elevation represent by shading. Lighter colors are higher elevation (larger $z$-value); darker colors are lower elevation:




Here's a somewhat bad 3D visualization (looking from the northwesty part of $xy$-plane, facing southeasty):




See the dot and arrow? It looks like we're facing uphill? But maybe not in the steepest possible direction?

So, if we're standing there, facing in that direction, how steep is the slope? Let's work it out. For the partials, we have:

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\left(x^2 - y^2\right)$$

$$= 2x - 0$$
$$= 2x$$

$$\begin{aligned}\frac{\partial f}{\partial y} &= \frac{\partial}{\partial y}\left(x^2 - y^2\right) \\ &= 0 - 2y \\ &= -2y\end{aligned}$$

We're facing in the direction of a $\frac{\pi}{6} = 30^\circ$ angle on the $xy$-plane. That's the same as a unit/direction vector of $\left\langle \frac{\sqrt{3}}{2}, \frac{1}{2} \right\rangle$. So then the slope will become:

$$\begin{aligned}\begin{matrix}\text{the slope in the} \\ \left\langle \frac{\sqrt{3}}{2}, \frac{1}{2} \right\rangle \text{ direction} \\ \text{at any point}\end{matrix} &= \begin{pmatrix}\text{the partial in the } x\text{-direction} \\ \text{at that point}\end{pmatrix} \cdot \begin{pmatrix}\text{the unit/direction vector} \\ \text{component in the } x\text{-direction}\end{pmatrix} + \begin{pmatrix}\text{the partial in the } y\text{-direction} \\ \text{at that point}\end{pmatrix} \cdot \begin{pmatrix}\text{the unit/direction vect} \\ \text{component in the } y\text{-direc}\end{pmatrix} \\ &= \frac{\partial f}{\partial x} \cdot \begin{pmatrix}\text{the unit/direction vector} \\ \text{component in the } x\text{-direction}\end{pmatrix} + \frac{\partial f}{\partial y} \cdot \begin{pmatrix}\text{the unit/direction vector} \\ \text{component in the } y\text{-direction}\end{pmatrix} \\ &= 2x \cdot \left(\frac{\sqrt{3}}{2}\right) \quad + \quad (-2y) \cdot \left(\frac{1}{2}\right) \\ &= x\sqrt{3} - y\end{aligned}$$

So then at any point $(x, y)$, when facing in the direction of a $30^\circ$ angle on the $xy$-plane, the slope is $x\sqrt{3} - y$! Great.

What about if we're at a specific point? We've been saying that we're at the point $(5, 3, f(5, 3)) = (5, 3, 16)$. Then the slope at that point is:

$$\begin{aligned}\begin{matrix}\text{the slope} \\ \text{in the } \left\langle \frac{\sqrt{3}}{2}, \frac{1}{2} \right\rangle \text{ direction} \\ \text{at the point } (5, 3, 16)\end{matrix} &= x\sqrt{3} - y \\ &= 5\sqrt{3} - 3 \\ &\approx 5.660\end{aligned}$$

If we think of this as an angle up from the $xy$-plane, it becomes:

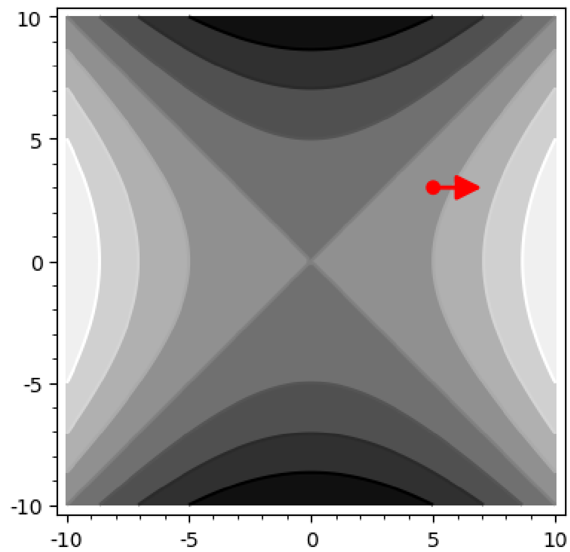
$$\begin{aligned}\text{as an angle, this is: } & \tan^{\text{inv}}\left(5\sqrt{3} - 3\right) \\ &\approx 1.396 \text{ radians} \\ &\approx 79.980^\circ\end{aligned}$$

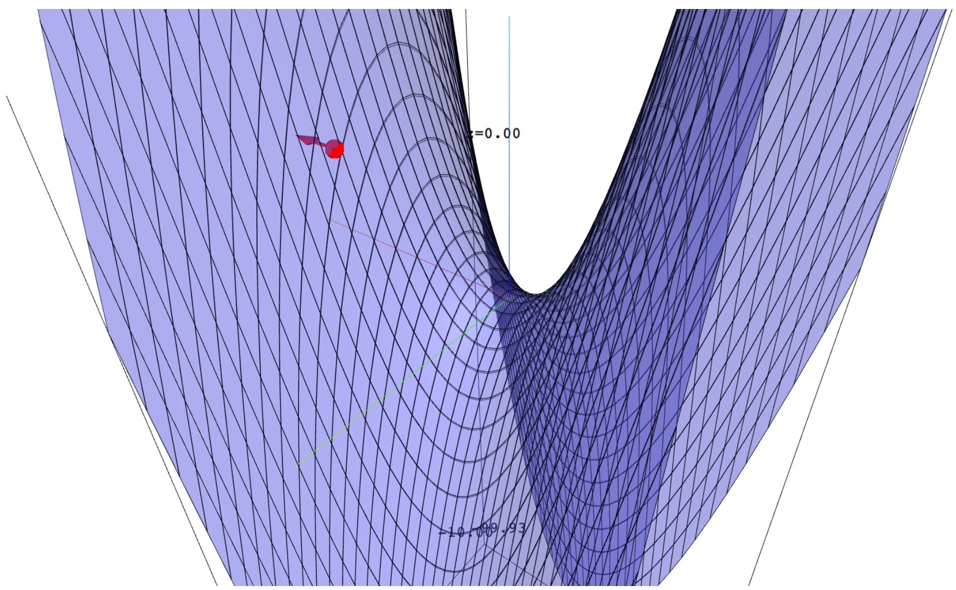
Yay!

So in general, we have something like this:

$$\boxed{\begin{matrix}\text{the slope} \\ \text{at a certain point} \\ \text{in a certain direction}\end{matrix} = \begin{pmatrix}\text{the partial in the } x\text{-direction} \\ \text{at that point}\end{pmatrix} \cdot \begin{pmatrix}\text{the unit/direction vector} \\ \text{in the } x\text{-direction}\end{pmatrix} + \begin{pmatrix}\text{the partial in the } y\text{-direction} \\ \text{at that point}\end{pmatrix} \cdot \begin{pmatrix}\text{the unit/direction vector} \\ \text{in the } y\text{-direction}\end{pmatrix}}$$

What if we turn in a different direction? Let's suppose we're looking in the $+x$ direction—i.e., directly along the $x$-axis, facing the positive direction.




The unit/direction vector will then be just $\langle 1, 0\rangle$. (You could also think of this as being a $0^\circ$ angle on the $xy$-plane.) So then the slope at that point will be:

$$
\begin{aligned}
\begin{matrix}\text{the slope}\\ \text{in the } +x \text{ direction}\\ \text{at any point}\end{matrix} &= \begin{pmatrix}\text{the partial in the } x\text{-direction}\\ \text{at that point}\end{pmatrix}\begin{pmatrix}\text{the unit/direction vector}\\ \text{in the } x\text{-direction}\end{pmatrix} + \begin{pmatrix}\text{the partial in the } y\text{-direction}\\ \text{at that point}\end{pmatrix}\begin{pmatrix}\text{the unit/direction vector}\\ \text{in the } y\text{-direction}\end{pmatrix}\\
&= \frac{\partial f}{\partial x}\cdot\begin{pmatrix}\text{the unit/direction vector}\\ \text{in the } x\text{-direction}\end{pmatrix} + \frac{\partial f}{\partial y}\cdot\begin{pmatrix}\text{the unit/direction vector}\\ \text{in the } y\text{-direction}\end{pmatrix}\\
&= \frac{\partial f}{\partial x}\cdot 1 + \frac{\partial f}{\partial y}\cdot 0\\
&= \frac{\partial f}{\partial x}\\
&= 2x
\end{aligned}
$$

So if we're looking in the $+x$ direction, the slope is just $2x$—which is just the partial in the $x$-direction! That's what a partial derivative really *is*: it's the slope in the direction of its variable/axis/dimension!

**what**
**partial derivatives**
**∼ *are* ∼**
**is**
**derivatives/slopes in the directions of their variables/axes/dimensions**

It's like we're taking this multidimensional function, taking a one-dimensional slice along the axis, and then finding that one-dimensional slope! Or, put differently, a partial derivative is the slope of a multidimensional function, only along a single axis (in the positive direction along that axis). (The adjective **axial** might be useful here: a partial derivative is the slope in one of the axial directions.)

The somewhat more general way to think about these derivatives in a particular direction is that they're the derivative matrix/vector, dot-product'd with a unit/direction vector. We multiply together each partial derivative with its corresponding unit/direction vector component, and add them all up to get a single number (aka a scalar):

$$
\begin{aligned}
\substack{\text{the slope}\\ \text{in a certain direction}} = \quad & \left(\substack{\text{the partial}\\ \text{in the } x_1\text{-dimension}}\right) \cdot \left(\substack{\text{the direction vector}\\ \text{component in the } x_1\text{-dimension}}\right) \\
+ & \left(\substack{\text{the partial}\\ \text{in the } x_2\text{-dimension}}\right) \cdot \left(\substack{\text{the direction vector}\\ \text{component in the } x_2\text{-dimension}}\right) \\
+ & \left(\substack{\text{the partial}\\ \text{in the } x_3\text{-dimension}}\right) \cdot \left(\substack{\text{the direction vector}\\ \text{component in the } x_3\text{-dimension}}\right) \\
+ & \quad \text{et cetera...}
\end{aligned}
$$

$$
= \begin{bmatrix} \text{the} \\ \text{derivative} \\ \text{vector} \end{bmatrix} \underset{\substack{\uparrow \\ \text{DOT PRODUCT!!!}}}{\cdot} \begin{bmatrix} \text{the} \\ \text{direction} \\ \text{vector} \end{bmatrix}
$$

Note that this only works if we have a function whose derivative is a vector! In other words, it only works if we have a function that only has one output—i.e., a function from $\mathbb{R}^n$ to $\mathbb{R}^1$. (People call these **scalar functions**; I often call them **surfaces**.) If we have a function that has more than one output, well, then its derivative is a matrix, things get more complicated, and we won't worry about that for now. (We'll save that for when we talk about **vector calculus**!)

Also note that this is the same as how we do things in 1D! It's just that in 1D, we only have one dimension, and the unit/direction vector in our one dimension is always $\langle +1 \rangle$. (Unless, that is, we want to look in the left/negative direction, in which case it's $\langle -1 \rangle$.) For example, if we have the 1D parabola:

$$f : \mathbb{R}^1 \to \mathbb{R}^1$$

$$f(x) = x^2$$

And we want to find its slope in the $+x$ direction, at the point $x = 5$, we have:

$$
\begin{aligned}
\substack{\text{the slope}\\ \text{in the } +x \text{ direction}\\ \text{at } x = 5} &= \left(\substack{\text{the partial in the } x\text{-direction}\\ \text{at } x = 5}\right) \cdot \left(\substack{\text{the unit/direction vector}\\ \text{component in the } x\text{-direction}}\right) \\
&= \frac{\partial f}{\partial x} \cdot \left(\substack{\text{the unit/direction vector}\\ \text{component in the } x\text{-direction}}\right) \\
&= 2x \cdot (+1) \\
&= 2 \cdot 5 \cdot (+1) \\
&= +10
\end{aligned}
$$

The name that people give to this stuff is **directional derivatives**. Sometimes people like to use fancy notation for them. Here's one I've seen a lot:

$$\substack{\text{the derivative of } f \\ \text{in the direction of } \hat{u} \\ \text{at the point } P} = D_{\hat{u}} f(x, y)\Big|_P$$

I don't like this notation and never use it. It's too complicated! Too much stuff going on! Too many symbols! Just use English instead. This is true for a lot of the stuff we'll do in multivariable calculus. The notation gets so complicated, and so many people use totally different notations. It's all a bunch of confusing hieroglyphics to me! We need to remember that *notation is not mathematics*. So feel free to make up your own notations, or just write things out in English. (The latter is what I usually do when it comes to directional derivatives.)

Actually, if you look on Wikipedia, the first few lines of the directional derivatives article has a "helpful" list of a bunch of common ways of denoting it (the derivative of the function $f$, in the direction of $\vec{v}$):

$$\nabla_{\mathbf{v}} f(\mathbf{x}) = f'_{\mathbf{v}}(\mathbf{x}) = D_{\mathbf{v}} f(\mathbf{x}) = Df(\mathbf{x})(\mathbf{v}) = \partial_{\mathbf{v}} f(\mathbf{x}) = \mathbf{v} \cdot \nabla f(\mathbf{x}) = \mathbf{v} \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$$

What a total mess!

## The Gradient

When it comes to functions from $\mathbb{R}^n$ to $\mathbb{R}^1$—**scalar functions**, a/k/a **surfaces**—their derivative matrix, as we've seen, is just a $1 \times n$ matrix, i.e., a vector. Since its derivative is a vector, we can think about its magnitude and direction, and it happens to have this really, really cool property:

| |
|---|
| thought of as a vector,<br>the derivative of a scalar function/surface<br>points in the direction of steepest ascent. |

This is pretty cool! I'll try to make some arguments for why this is the case in a moment. But first, a vocab word. Because people love talking about scalar functions, and because their derivatives we can just treat like vectors, they get a special name: a **gradient**:

| |
|---|
| **the gradient** = the derivative of a scalar function/surface |

So, to go back to our hyperbolic parabaloid example:

$$f : \mathbb{R}^2 \to \mathbb{R}^1$$

$$f(x, y) = x^2 - y^2$$

The gradient of $f$, which often gets written with a little upside down triangle $\nabla$ (a "nabla," typeset with `\nabla` in TEX), is:

$$\text{the gradient of } f = \nabla f = \text{grad} f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x \\ -2y \end{bmatrix}$$

More generally, if we have an $n$-dimensional surface embedded in $(n+1)$-dimensional space:

$$g : \mathbb{R}^n \to \mathbb{R}^1$$

$$g(x_1, x_2, x_3, \cdots, x_n) \quad = \quad \text{something}$$

Then its gradient is an $n$-dimensional vector:

$$\text{the gradient of } g = \nabla g = \text{grad } g = \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \\ \frac{\partial g}{\partial x_3} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{bmatrix}$$

So **gradient** is just a fancy name for "the derivative of a scalar function." I don't use the word that much, because why not just say 'derivative"? But lots of people like it.

And because it's a vector, it has a magnitude and a direction. And here's the fun fact, again:

| |
|---|
| thought of as a vector,<br>the gradient<br>points in the direction of steepest ascent. |

Why is this true? Well, I can make a couple of arguments. Here's one of them:

> The slope in a certain direction is just the dot product of the gradient/derivative vector with the unit vector in that direction. How does dot-producting work? Suppose you're dot-producting two vectors, and you're rotating the vectors, so that the angle between them changes (but not their lengths). How does the dot product change? Well, dot product of $\vec{a}$ and $\vec{b}$ is, in its trigonometric formulation, $||\vec{a}|| \cdot ||\vec{b}|| \cos\theta$. So then:
>
> - When $\theta = 0$, then $\cos\theta = 1$, and the dot product is just the product of the two magnitudes
> - As the angle between them increases, $\theta$ increases, and $\cos\theta$ decreases, so the dot product decreases!
> - When the angle between them gets to be $\frac{\pi}{2} = 90°$, then we have $\cos\frac{\pi}{2} = 0$, and then the dot product is zero!
>
> (You might remember that I did this in class, using two PVC pipes as vectors, and rotating them relative to each other.)
>
> So if we want the dot product to be the largest it can be—and we can't change the lengths of the vectors—then we need the angle between the two vectors to be zero. So for the slope to be the biggest it can be, the derivative vector/gradient needs to be pointing in the same direction as whatever unit vector we're dotting it with.

There's another cool argument I like that works only for functions from $\mathbb{R}^2$ to $\mathbb{R}^1$; I think I'll assign it as a homework problem!

## Repeated/Higher-Order Partial Derivatives

When we take normal derivatives, we can keep going. We can take second derivatives (derivatives of derivatives), third derivatives (derivatives of derivatives of derivatives), $n$'th derivatives, and so on. (When it comes to physics, these have beautiful interpretations: position, velocity, acceleration, jerk, etc.!)

We can take repeated partial derivatives, too! For instance, let's suppose we have this function:

$$f : \mathbb{R}^2 \to \mathbb{R}^1$$

$$f(x, y) = x^4 + 5xy + y^3$$

I made it asymmetric in $x$ and $y$, just to show how things are different. We can find the first partial derivatives:

$$\frac{\partial f}{\partial x} = 4x^3 + 5y \qquad \frac{\partial f}{\partial y} = 5x + 3y^2$$

But we can keep on going! For instance, we can take the second partial derivative with respect to $x$, just by partially-differentiating again the partial derivative:

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial x}\right] \\ &= \frac{\partial}{\partial x}\left[4x^3 + 5y\right] \\ &= 12x^2 + 0 \\ &= 12x^2 \end{aligned}$$

We can take the second partial derivative with respect to $y$, too:

$$\begin{aligned} \frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial y}\right] \\ &= \frac{\partial}{\partial y}\left[5x + 3y^2\right] \\ &= 0 + 6y \\ &= 6y \end{aligned}$$

But wait! There's more! What if we take what's called a **mixed partial derivative**? Namely: what if we take the first partial derivative with respect to $x$, and then partially-differentiate it a second time, but this time, with respect to $y$???

$$\begin{aligned} \frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial x}\right] \\ &= \frac{\partial}{\partial y}\left[4x^3 + 5y\right] \\ &= 0 + 5 \\ &= 5 \end{aligned}$$

That's not the only mixed partial we could take! We could take the first partial derivative with respect to $y$, and then partially-differentiate it with respect to $x$!

$$\begin{aligned} \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial y}\right] \\ &= \frac{\partial}{\partial x}\left[5x + 3y^2\right] \\ &= 5 + 0 \\ &= 5 \end{aligned}$$

Notice how they're the same!!! This is a general truth about mixed partial derivatives: regardless of what order we take them in, they're the same:

$$\boxed{\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}}$$

(Well, there are *some* conditions that need to be met—limits and continuity and that sort of thing—but for our purposes, we won't worry so much about that.)

I don't have any sort of intuition or understanding of what a mixed partial derivative *is*, geometrically. Here's a Math StackExchange post attempting to give one.

If you're taking a lot of higher-order partial derivatives, the subscript notation ($f_x$, $f_y$) can be a bit easier, because you don't have to keep writing $\partial$ over and over. For example, $f_{xx}$ is the second partial with respect to $x$, $f_{yy}$ is the second partial with respect to $y$, and that fact about the mixed partials being equal we could phrase as:

$$f_{xy} = f_{yx}$$

If you want to take REALLY high-order partials, then this notation requires less writing! For example, here's a mixed fifth partial derivative, phrased in both notations:

$$\frac{\partial^5 f}{\partial x \partial x \partial y \partial x \partial y} = f_{yxyxx}$$

Oh, yeah... the variable we're differentiating against gets written in the opposite order in each of these notations. Eep.

## Problems

Calculate the full derivative—i.e., the full derivative matrix—for each of these functions. (Show your work for finding all the partials!)

1. $f : \mathbb{R}^1 \to \mathbb{R}^1; \quad f(x) = x^3$
2. $f : \mathbb{R}^1 \to \mathbb{R}^1; \quad f(x) = \ln\left(\sin\left(x^7\right)\right)$
3. $f(x, y) =$
4. $f : \mathbb{R}^2 \to \mathbb{R}^1; \quad f(x, y) = \cos\left(x^2 y\right) + y^3$
5. $f : \mathbb{R}^2 \to \mathbb{R}^1; \quad f(x, y) = \frac{xy}{x^2+y^2}$
6. $f : \mathbb{R}^2 \to \mathbb{R}^1; \quad f(x, y) = e^{x^2+y^2}$
7. $f : \mathbb{R}^2 \to \mathbb{R}^1; \quad f(x, y) = xy\ln(xy)$
8. $f : \mathbb{R}^2 \to \mathbb{R}^1; \quad f(x, y) = \sqrt{1 - x^2 - y^2}$
9. $f : \mathbb{R}^2 \to \mathbb{R}^1; \quad f(x, y) = x\tan(y)$
10. $f : \mathbb{R}^2 \to \mathbb{R}^1; \quad f(x, y) = \frac{1}{xy}$
11. $f : \mathbb{R}^3 \to \mathbb{R}^2; \quad f(x, y, z) = \left\langle xy + 2yz,\, 2xy^2x\right\rangle$
12. $f : \mathbb{R}^3 \to \mathbb{R}^1; \quad f(x, y, z) = x^2 + 7yz$
13. $f : \mathbb{R}^1 \to \mathbb{R}^3; \quad f(t) = t^5\hat{\imath} - 2t\hat{\jmath} + t^2\hat{k}$
14. $f : \mathbb{R}^4 \to \mathbb{R}^5; f\left(x_1, x_2, x_3, x_4\right) = \begin{bmatrix} x_1x_3 \\ \tan\left(x_4\right) \\ -\ln\left(x_2\right) \\ \left(3x_1-2\right)^4 \\ 1729 \end{bmatrix}$

Find all the first, second, and third partial derivatives (including all the mixed ones!) for the following functions (also, draw a picture of them using some sort of 3D graphing software):

15. $f(x, y) = 7x + 2x^2y^3 + 10y^2$
16. $f(x, y) = 3xy^3 + 8x^2y^4$
17. $f(x, y) = 4x^2y^5 + 3x^3y^2$
18. $f(x) = x^5$

19. Suppose that one of your classmates reports that for a particular function $f(x, y)$, the partial derivatives are:

$$\frac{\partial f}{\partial x} = 2x + 3y$$

$$\frac{\partial f}{\partial y} = 4x + 6y$$

Do you believe them? Why, or why not?

**20**. Suppose you have a function from $\mathbb{R}^2$ to $\mathbb{R}^1$. How many total first partial derivatives does it have? What about second partial derivatives (including the mixed ones, but counting the mixed ones taken in a different order as equivalent)? What about third partial derivatives (again, including the mixed ones)? What about $k$'th partial derivatives?

**21**. Suppose you have a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. How many total first partial derivatives does it have? How many total second partial derivatives does it have (including the mixed ones, but counting the mixed ones taken in a different order as equivalent)?

**22**. How many distinct mixed $k$'th partial derivatives can a function have? (Assume that we're dealing with a function with which we can take an arbitrary number of partial derivatives, and for which the theorem about order not mattering in mixed partial derivatives holds.)

**23**. Suppose you have a function $f(x, y)$; $f : \mathbb{R}^2 \to \mathbb{R}^1$. Imagine you're standing on this function, at the point $(x, y)$, facing in the direction of the angle $\theta$. What's the slope? For what value of $\theta$ is the slope greatest? Greatest upwards? Greatest downwards? When is it flat? Argue/solve this *without* using the fact that the gradient points in the direction of steepest ascent—just use your 1D calc skills!