

Directional derivatives (introduction)

 Google Classroom

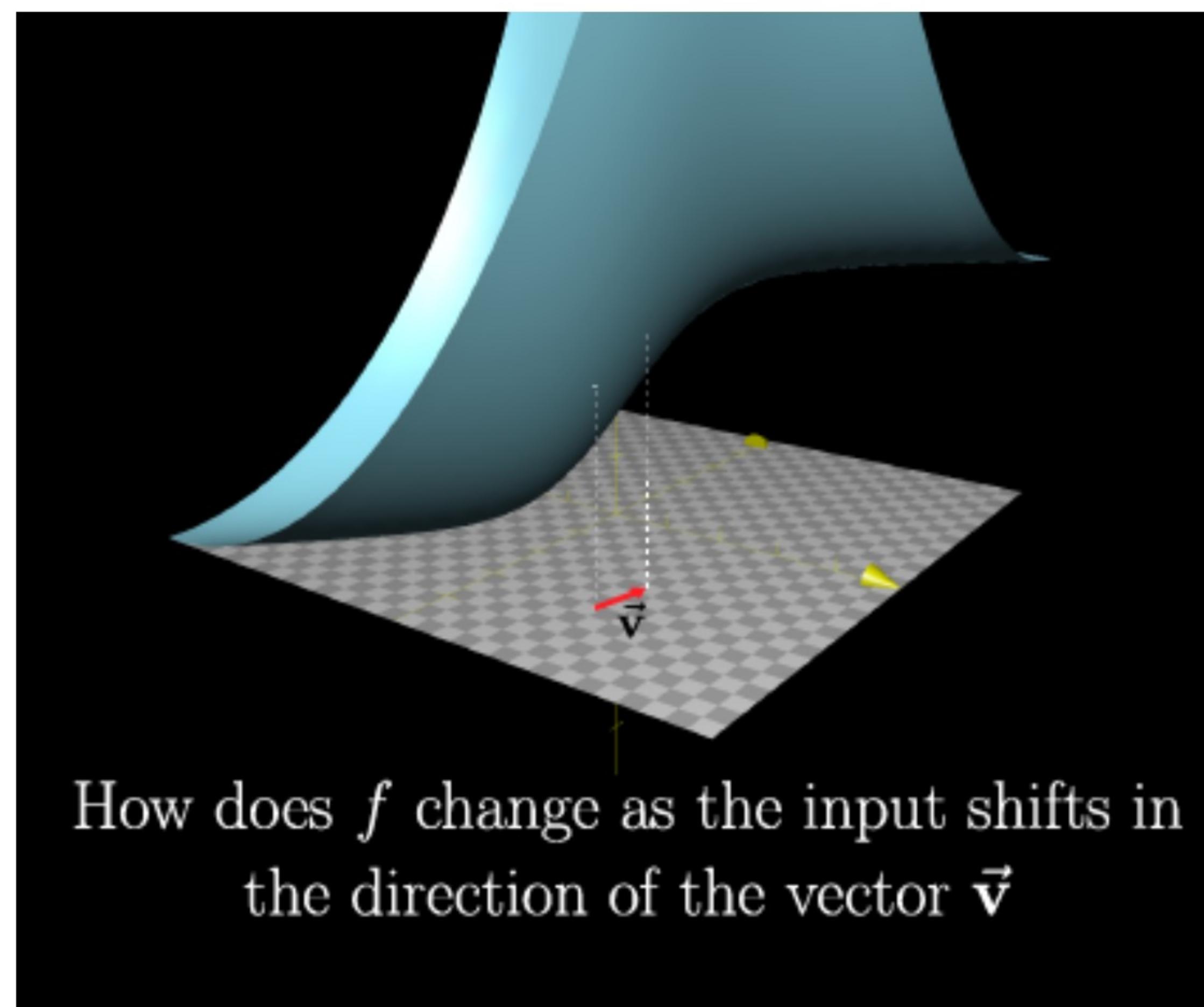
How does the value of a multivariable function change as you nudge the input in a specific direction?

Background

- [Partial derivatives](#)
- [Gradient](#)

What we're building to

- If you have some multivariable function, $f(x, y)$ and some vector in the function's input space, \vec{v} , the directional derivative of f along \vec{v} tells you the rate at which f will change while the input moves with velocity vector \vec{v} .



- The notation here is $\nabla_{\vec{v}} f$, and it is computed by taking the dot product between the gradient of f and the vector \vec{v} , that is, $\nabla f \cdot \vec{v}$
- When the directional derivative is used to compute slope, be sure to normalize the vector \vec{v} first.

Generalizing partial derivatives

Consider some multivariable function:

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

We know that the partial derivatives with respect to x and y tell us the rate of change of f as we nudge the input either in the x or y direction.

[\[Hide explanation.\]](#)

The partial derivative of f with respect to x is below:

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

If we evaluate this at some point—like $(2, 3)$ —we get the following:

$$\frac{\partial f}{\partial x}(2, 3) = 2(2) - 3 = 1$$

But what does this actually mean?

Let's say you evaluate the original function f at the point $(2, 3)$:

$$f(2, 3) = (2)^2 - (2)(3) = -2$$

Now nudge the input slightly in the x -direction, perhaps by the value 0.01, moving it to $(2.01, 3)$. The function now evaluates thus:

$$f(2.01, 3) = (2.01)^2 - (2.01)(3)$$

$$= 4.0401 - 6.03$$

$$= -1.9899$$

The total change in the output, from -2 to -1.9899 is 0.0101. The ratio between this change and the size of our nudge in the input space is below:

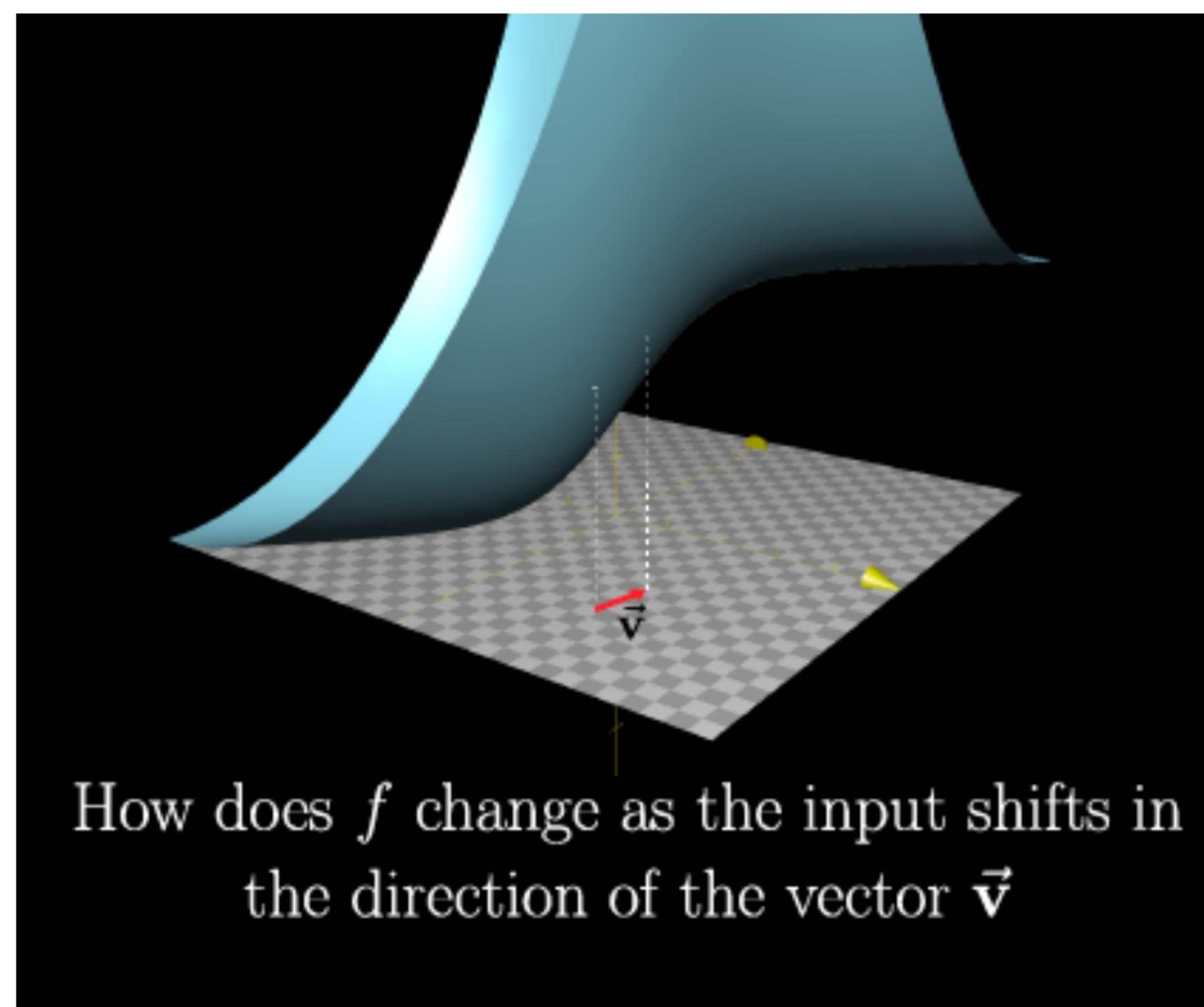
$$\frac{0.0101}{0.01} \approx 1 = \frac{\partial f}{\partial x}(2, 3)$$

The smaller our original nudge, the closer this ratio will be to $\frac{\partial f}{\partial x}(2, 3) = 1$.

The question now is what happens when we nudge the input of f in a direction which is not parallel to the x or y axes.

For example, the image below shows the graph of f along with a small step along a vector \vec{v} in the input space, meaning the xy -plane in this case. Is there an operation which tells us how the height of the graph above the tip of \vec{v}

compares to the height of the graph above its tail?



As you have probably guessed, there is a new type of derivative, called the **directional derivative**, which answers this question.

Just as the partial derivative is taken with respect to some input variable—e.g., x or y —the directional derivative is taken *along* some vector \vec{v} in the input space.

One very helpful way to think about this is to picture a point in the input space moving with velocity \vec{v} . The directional derivative of f along \vec{v} is the resulting rate of change in the output of the function. So, for example, multiplying the vector \vec{v} by two would double the value of the directional derivative since all changes would be happening twice as fast.

Notation

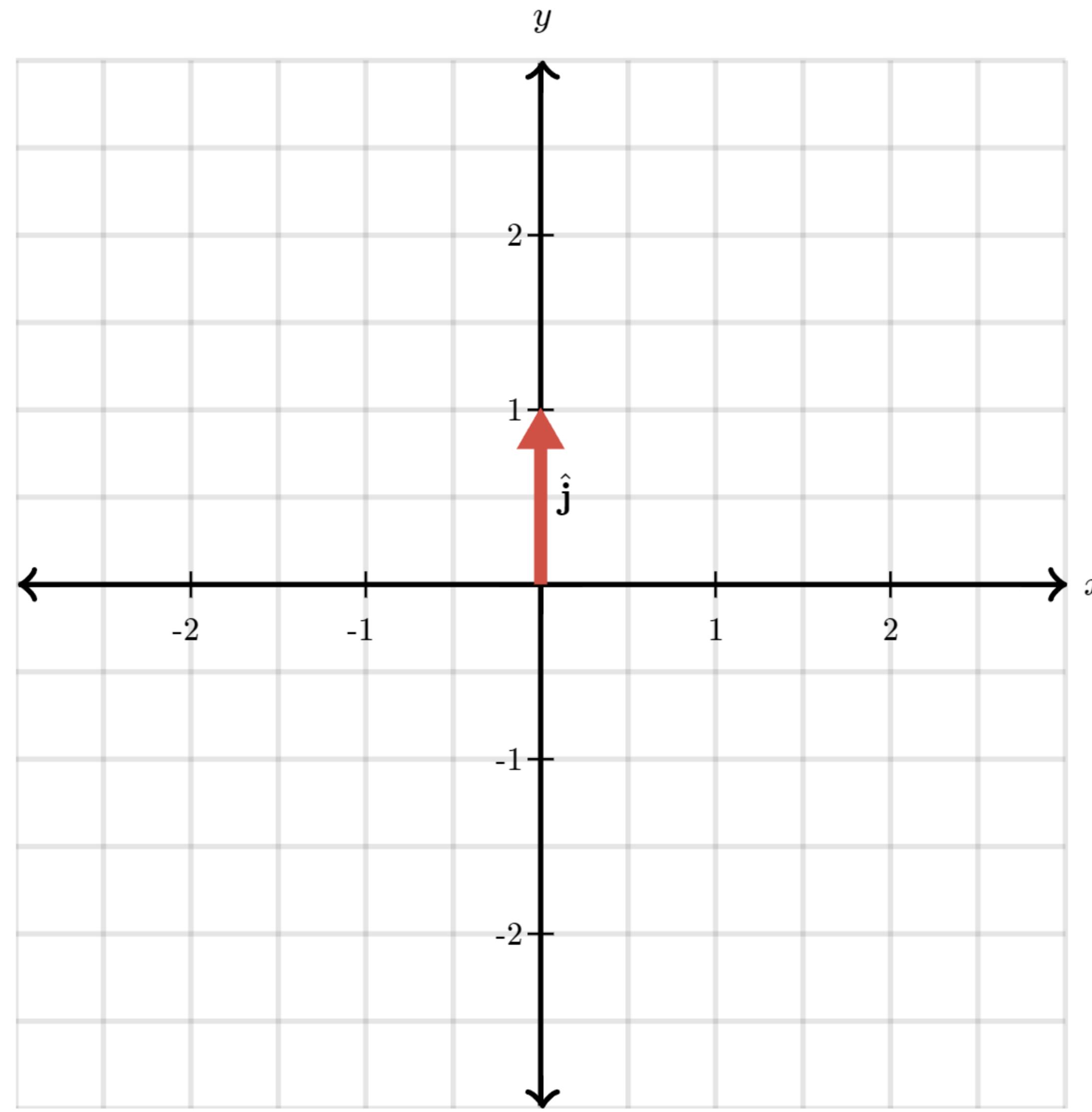
There are quite a few different notations for this one concept:

- $\nabla_{\vec{v}} f$
- $\frac{\partial f}{\partial \vec{v}}$
- $f'_{\vec{v}}$
- $D_{\vec{v}} f$
- $\partial_{\vec{v}} f$

All of these represent the same thing: the rate of change of f as you nudge the input along the direction of \vec{v} . We'll use the $\nabla_{\vec{v}} f$ notation, just because it subtly hints at how you compute the directional derivative using the gradient, which you'll see in a moment.

Example 1: $\vec{v} = \hat{\mathbf{j}}$

Before jumping into the general rule for computing $\nabla_{\vec{v}} f$, let's look at how we can rewrite the more familiar notion of a partial derivative as a directional derivative.



For example, the partial derivative $\frac{\partial f}{\partial y}$ tells us the rate at which f changes as we nudge the input in the y direction. In other words, as we nudge it along the vector $\hat{\mathbf{j}}$. Therefore, we could equivalently write the partial derivative with respect to y as $\frac{\partial f}{\partial y} = \nabla_{\hat{\mathbf{j}}} f$.

This is all just fiddling with different notation. What's more important is to have a clear mental image of what all this notation *represents*.

Reflection Question: Suppose $\vec{v} = \hat{\mathbf{i}} + \hat{\mathbf{j}}$, what is your best guess for $\nabla_{\vec{v}} f$?

[\[Hide explanation.\]](#)

$$\nabla_{\vec{v}} f = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}$$

How to compute the directional derivative

Let's say you have a multivariable $f(x, y, z)$ which takes in three variables— x , y and z —and you want to compute its directional derivative along the following

vector:

$$\vec{v} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$$

The answer, as it turns out, is

$$\nabla_{\vec{v}} f = 2 \frac{\partial f}{\partial x} + 3 \frac{\partial f}{\partial y} + (-1) \frac{\partial f}{\partial z}$$

This should make sense because a tiny nudge along \vec{v} can be broken down into *two* tiny nudges in the x -direction, *three* tiny nudges in the y -direction, and a tiny nudge backwards, by -1 , in the z -direction. We'll go through the rigorous reasoning behind this much more thoroughly in the [next article](#).

More generally, we can write the vector \vec{v} abstractly as follows:

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

The directional derivative looks like this:

$$\nabla_{\vec{v}} f = v_1 \frac{\partial f}{\partial x} + v_2 \frac{\partial f}{\partial y} + v_3 \frac{\partial f}{\partial z}$$

That is, a tiny nudge in the \vec{v} direction consists of v_1 times a tiny nudge in the x -direction, v_2 times a tiny nudge in the y -direction, and v_3 times a tiny nudge in the z -direction.

This can be written in a super-pleasing compact way using the dot product and the gradient:

$$\nabla_{\vec{v}} f(x, y, z)$$

$$= v_1 \frac{\partial f}{\partial x}(x, y, z) + v_2 \frac{\partial f}{\partial y}(x, y, z) + v_3 \frac{\partial f}{\partial z}(x, y, z)$$

$$= \begin{bmatrix} \frac{\partial f}{\partial x}(x, y, z) \\ \frac{\partial f}{\partial y}(x, y, z) \\ \frac{\partial f}{\partial z}(x, y, z) \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$= \nabla f(x, y, z) \cdot \vec{v}$$

This is why the notation $\nabla_{\vec{v}} f$ is so suggestive of the way we compute the directional derivative:

$$\nabla_{\vec{v}} f = \nabla f \cdot \vec{v}$$

Take a moment to delight in the fact that one single operation, the gradient, packs enough information to compute the rate of change of a function in *every possible direction!* That's so many directions! Left, right, up, down, north-north-east, 34.8° clockwise from the x -axis... Madness!

Example 2:

Problem: Take a look at the following function.

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

What is the directional derivative of f at the point $(2, -3)$ along the vector $\vec{v} = 0.6\hat{i} + 0.8\hat{j}$?

Solution: You can think of the direction derivative either as a weighted sum of partial derivatives, as below:

$$\nabla_{\vec{v}} f = 0.6 \frac{\partial f}{\partial x} + 0.8 \frac{\partial f}{\partial y}$$

Or, you can think of it as a dot product with the gradient, as you see here:

$$\nabla_{\vec{v}} f = \nabla f \cdot \vec{v}$$

The first is faster, but just for practice, let's see how the gradient interpretation unfolds. We start by computing the gradient itself:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x}(x^2 - xy) \\ \frac{\partial}{\partial y}(x^2 - xy) \end{bmatrix} = \begin{bmatrix} 2x - y \\ -x \end{bmatrix}$$

Next, plug in the point $(x, y) = (2, -3)$ since this is the point the question asks us about.

$$\nabla f(2, -3) = \begin{bmatrix} 2(2) - (-3) \\ -(2) \end{bmatrix} = \begin{bmatrix} 7 \\ -2 \end{bmatrix}$$

To get the desired directional derivative, we take the dot product between this gradient and \mathbf{v} :

$$\nabla_{\vec{v}} f(2, -3) = \nabla f(2, -3) \cdot (\mathbf{0.6}\hat{\mathbf{i}} + \mathbf{0.8}\hat{\mathbf{j}})$$

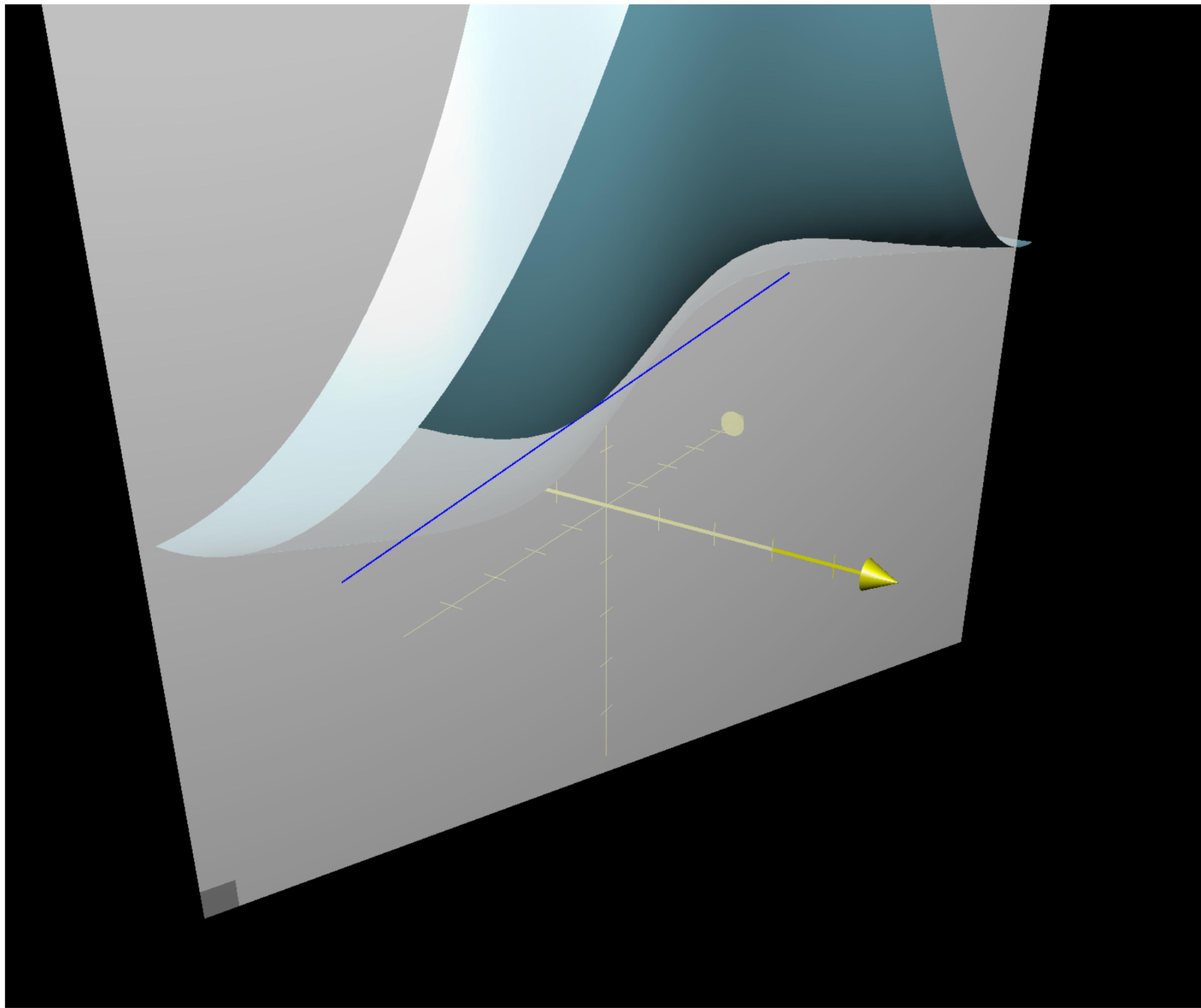
$$= \begin{bmatrix} 7 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}$$

$$= 7(0.6) + (-2)(0.8)$$

$$= 2.6$$

Finding slope

How do you find the slope of a graph intersected with a plane that is not parallel to the x or y axes?



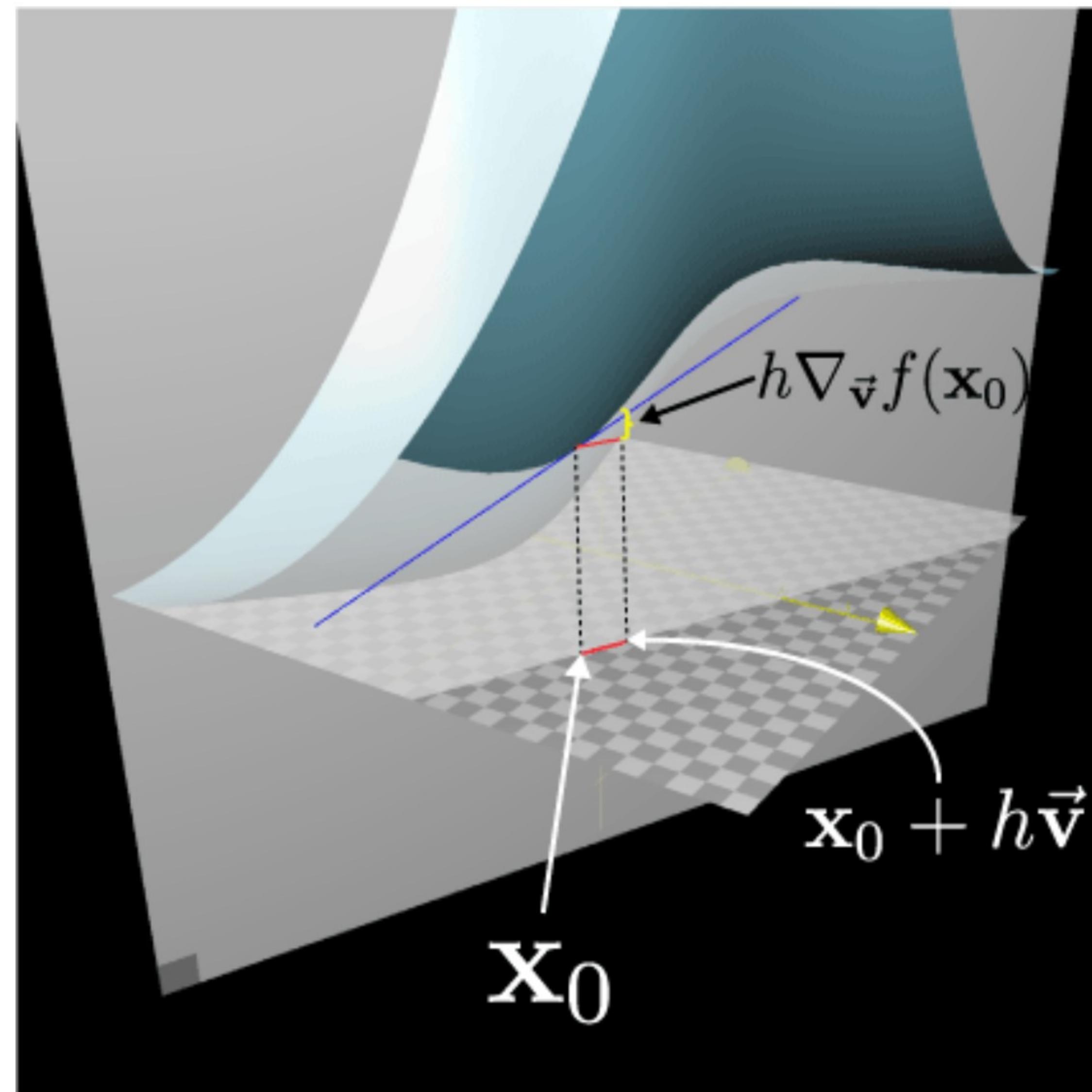
You can use the directional derivative, but there is one important thing to remember:

If the directional derivative is used to compute slope, either \vec{v} must be a unit vector or you must remember to divide by $||\vec{v}||$ at the end.

In the definition and computation above, doubling the length of \vec{v} would double the value of the directional derivative. In terms of the computation, this is because $\nabla f \cdot (2\vec{v}) = 2(\nabla f \cdot v)$.

However, this might not always be what you want. The slope of a graph in the direction of \vec{v} , for example, depends only on the direction of \vec{v} , not the magnitude $||\vec{v}||$. Let's see why.

How can we imagine this slope? Slice the graph of f with a vertical plane that cuts the xy -plane in the direction of \vec{v} . The slope in question is that of a line tangent to the resulting curve. As with any slope, we look for the rise over run.



In this case, the run will be the distance of a small nudge in the direction of \vec{v} . We can express such a nudge as an addition of $h\vec{v}$ to an input point \mathbf{x}_0 , where h is thought of as some small number. The magnitude of this nudge is $h||\vec{v}||$.

The resulting change in the output of f can be approximated by multiplying this little value h by the directional derivative:

$$h\nabla_{\vec{v}}f(x_0, y_0)$$

In fact, the rise of the tangent line—as opposed to the graph of the function—is precisely $h\nabla_{\vec{v}}f(x_0, y_0)$ due to this run of size $h||\vec{v}||$. For full details on why this is true, see the formal definition of the directional derivative in the [next article](#).

Therefore, the rise-over-run slope of our graph is

$$\frac{h\nabla_{\vec{v}}f(x_0, y_0)}{h||v||} = \boxed{\frac{\nabla_{\vec{v}}f(x_0, y_0)}{||v||}}$$

Notice, if \vec{v} is a unit vector, meaning $||\vec{v}|| = 1$, then the directional derivative does give the slope of a graph along that direction. Otherwise, it is important to remember to divide out by the magnitude of \vec{v} .

Some authors even go so far as to include normalization in the definition of $\nabla_{\vec{v}}f$.

Alternate definition of directional derivative:

$$\nabla_{\vec{v}} f(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\vec{v}) - f(\mathbf{x})}{h \|\vec{v}\|}$$

Personally, I think this definition puts too much emphasis on the particular use case of finding slope, so I prefer to use the original definition and normalize \vec{v} when necessary.

Example 3: Slope

Problem: On the stage for this problem we have three players.

Player 1, the function:

$$f(x, y) = \sin(xy)$$

Player 2, the point:

$$(x_0, y_0) = \left(\frac{\pi}{3}, \frac{1}{2} \right)$$

Player 3, the vector:

$$\vec{v} = 2\hat{i} + 3\hat{j}$$

What is the slope of the graph of f at the point (x_0, y_0) along the vector \vec{v} ?

Answer: Since we are finding slope, we must first normalize the vector in question. The magnitude $\|\vec{v}\|$ is $\sqrt{2^2 + 3^2} = \sqrt{13}$, so we divide each term by $\sqrt{13}$ to get the resulting unit vector \hat{u} in the direction of \vec{v} :

[\[Hide explanation.\]](#)

$$\hat{u} = \frac{\vec{v}}{\|\vec{v}\|} = \frac{2}{\sqrt{13}}\hat{i} + \frac{3}{\sqrt{13}}\hat{j}$$

Next, find the gradient of f :

[\[Hide explanation.\]](#)

$$\nabla f = \begin{bmatrix} \frac{\partial}{\partial x} \sin(xy) \\ \frac{\partial}{\partial y} \sin(xy) \end{bmatrix} = \begin{bmatrix} y \cos(xy) \\ x \cos(xy) \end{bmatrix}$$

Plug in the point $(x_0, y_0) = \left(\frac{\pi}{3}, \frac{1}{2}\right)$ to this gradient.

[\[Hide explanation.\]](#)

$$\nabla f \left(\frac{\pi}{3}, \frac{1}{2} \right)$$

$$= \begin{bmatrix} \left(\frac{1}{2}\right) \cos\left(\frac{\pi}{3} \cdot \frac{1}{2}\right) \\ \left(\frac{\pi}{3}\right) \cos\left(\frac{\pi}{3} \cdot \frac{1}{2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \left(\frac{1}{2}\right) \left(\frac{\sqrt{3}}{2}\right) \\ \left(\frac{\pi}{3}\right) \left(\frac{\sqrt{3}}{2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sqrt{3}}{4} \\ \frac{\pi\sqrt{3}}{6} \end{bmatrix}$$

Finally, take the dot product between $\hat{\mathbf{u}}$ and $\nabla f(\pi/3, 1/2)$:

[\[Hide explanation.\]](#)

$$\nabla f \left(\frac{\pi}{3}, \frac{1}{2} \right) \cdot \hat{\mathbf{u}}$$

$$= \begin{bmatrix} \frac{\sqrt{3}}{4} \\ \frac{\pi\sqrt{3}}{6} \end{bmatrix} \cdot \begin{bmatrix} \frac{2}{\sqrt{13}} \\ \frac{3}{\sqrt{13}} \end{bmatrix}$$

$$= \frac{\sqrt{3}}{4} \cdot \frac{2}{\sqrt{13}} + \frac{\pi\sqrt{3}}{6} \cdot \frac{3}{\sqrt{13}}$$

$$= \boxed{\frac{(1+\pi)\sqrt{3}}{2\sqrt{13}}}$$

Summary

- If you have some multivariable function, $f(x, y)$ and some vector in the function's input space, \vec{v} , the directional derivative of f along \vec{v} tells you the rate at which f will change while the input moves with velocity vector \vec{v} .
- The notation here is $\nabla_{\vec{v}} f$, and it is computed by taking the dot product between the gradient of f and the vector \vec{v} , that is, $\nabla f \cdot \vec{v}$.
- When the directional derivative is used to compute slope, be sure to normalize the vector \vec{v} first.