Partial Derivatives

Max stole Mark's template; Prepared on April 12, 2024

Part 1: Derivatives

Our goal today is to introduce derivatives as the best linear approximation of a function. We will see some other interpretations (slope of tangent line, rate of change of a function) and how they all mean the same thing, but we will avoid heavy computations (especially ones with limits).

Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$, what does it mean for f to be linear?

Definition 1:
Recall that a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **linear** if $f(\vec{x} + \vec{y}) = f(\vec{x}) + f(\vec{y})$ and $f(r\vec{x}) = rf(\vec{x})$ for $r \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$.

For simplicity, let's consider the case that $n, m = 1$, i.e., $f : \mathbb{R} \to \mathbb{R}$. Then f is linear if $f(x + y) = f(x) + f(y)$ and $f(xy) = f(x)f(y)$ for all $x, y \in \mathbb{R}$. As we learned in the Linear Algebra packet, linear functions are quite convenient to work with. We can represent them as matrices.
In our case, a linear function $f : \mathbb{R} \to \mathbb{R}$ can be represented by a 1-by-1 matrix, which is just a real number. So $f(x) = rx$ for some $r \in \mathbb{R}$.
Linear functions are extremely convenient and easy to work with, but most functions are not linear! It would be very convenient if they were though...

Problem 2:
Is the function $f(x) = x^2$ linear?

Problem 3:
Can you describe what the graph of a linear function $f : \mathbb{R} \to \mathbb{R}$ looks like geometrically? What shape(s) can it make? Try some examples and see what you can get.

It would be great if we could make functions be always linear. Unfortunately, this doesn't work because most functions are not linear. But as it turns out, most functions that you'll encounter on a daily basis are *almost* linear.

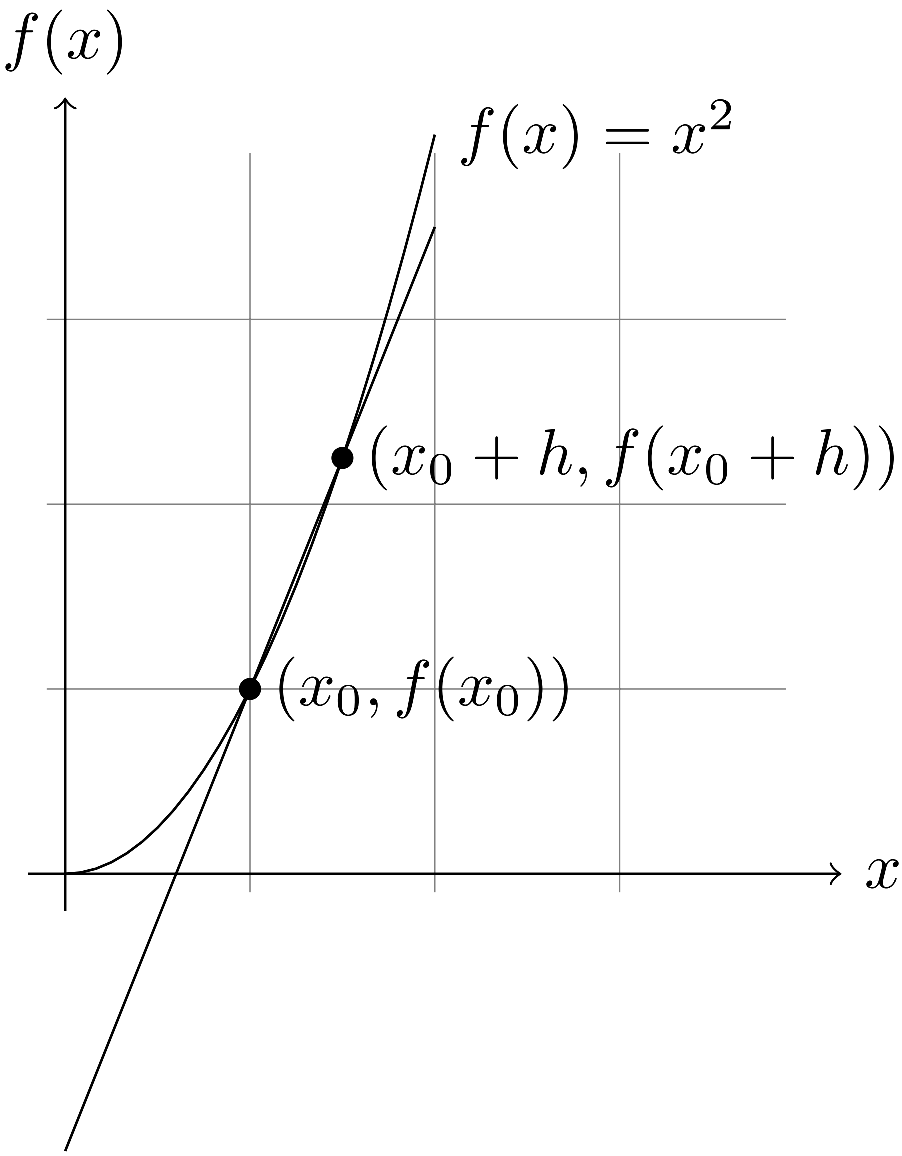
Definition 4:
Given a function $f : \mathbb{R} \to \mathbb{R}$, and a point $x \in \mathbb{R}$, we define the *derivative* of f at r to be a linear map $A : \mathbb{R} \to \mathbb{R}$ such that $f(x + h) - f(x) - A(h)$ is "as small as possible" relative to h.

Important Concept

We will define formally what "as small as possible" means, but the idea of this definition is that $f(x + h) \approx A(h) + f(x)$ is the *best linear approximation* of f around x.

When x increases by h, how much should $f(x)$ increase? The answer can be *approximated* easily by a linear approximation. If we have $f(x + h) \approx f(x) + A(h)$, then we are saying that when x increases by h, then $f(x)$ increases by approximately $A(h)$. So $A(h)$ measures the "rate of change" of f at x.

Example 5:
Let $f(x) = x^2$, $x_0 = 1, h = 0.5$.



Notice how the line between $(x_0 + h, f(x_0 + h))$ and $(x_0, f(x_0))$ is very very close to matching the graph of $f(x)$!

Problem 6:
What is the equation of the line passing through $(x_0, f(x_0))$ and $(x_0 + h, f(x_0 + h))$?

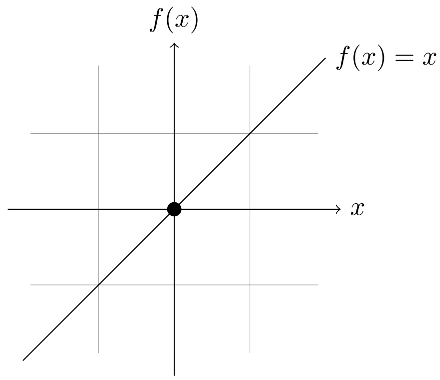
Problem 7:
Let $z \in [x_0, x_0 + h]$. Use the line to approximate $f(z)$ in terms of $f(x)$ and $z - x$.
Hint: plug in z to the equation of the line you got in the previous problem. If $f(z)$ matches the line very closely, we can approximate $f(z)$ by the y-coordinate of the line at z.

If you've seen derivatives before, you may recognise $\frac{f(x_0+h)-f(x_0)}{h}$ as being closely related to the definition of the derivative you were given. Forget that! We will define derivatives today in a way that will generalise much more easily to n dimensions.

At the beginning of this section, I said our goal was to avoid heavy computations involving limits. Unfortunately, defining derivatives without limits at all is very difficult. So let's do some examples of limits.

Definition 8:
A **limit** of a function $f(x)$ at a point x_0 is the value that f approaches at x_0.
It may be the case that $f(x_0)$ is not defined, but the limit may still be defined. It also may be the case that $f(x_0)$ exists but is not equal to the limit of $f(x)$ at x_0, although we will not be considering any functions today where that may be possible.

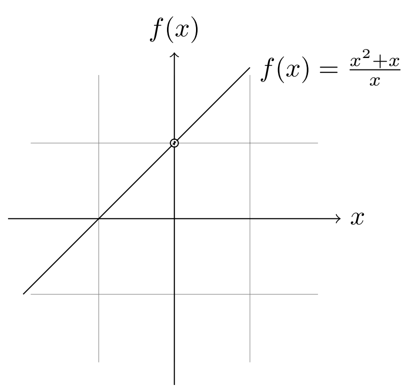
Example 9:
Consider $f(x) = x$ at $x_0 = 0$. What does $f(x)$ approach at 0?



When $x = 0.1$, $f(0.1) = 0.1$. When $x = 0.01$, $f(x) = 0.01$. So $f(x)$ gets closer and closer to 0 as x gets closer and closer to 0. Thus, we say $f(x)$ **approaches** 0 when x approaches 0. This is written as $\lim_{x \to 0} f(x) = 0$ (this is read as "the limit of f as x approaches 0 is 0").

Problem 10:
What does $f(x) = x^2$ approach at $x_0 = 0$? *Hint: this is as easy as it sounds.*

Example 11:
Let's try to evaluate $\lim_{x \to 0} f(x)$ when $f(x) = \frac{x^2+x}{x}$.



Notice that this function is not defined at 0! So we can't just plug in 0 and find the answer. Instead, we plug in values very close to 0 and see what we get: $f(0.1) = 1.1, f(0.01) = 1.01, f(0.001) = 1.001$. So we can experiment and find $\lim_{x \to 0} f(x) = 1$. We can prove that our guess is correct by noticing that, when $x \neq 0$, $f(x) = x + 1$, so we can plug in $f(0) = 1$ here.

Definition 12:
Given a function $f(x) : \mathbb{R} \to \mathbb{R}$, we say f is of **class** $o(x)$ if when x is very small, $f(x)$ is "smaller" than x (in other words, $f(x)$ goes to 0 "faster" than x does). Formally,

$$f(x) \in o(x) \iff \lim_{x \to 0} \frac{|f(x)|}{|x|} = 0$$

Problem 13:
Is $f(x) = x$ of class $o(x)$? What about $f(x) = x^2$? $f(x) = x^3$? $f(x) = \sqrt{x}$?

Problem 14:
Prove that if $f(x)$ is linear, then $f(x) \in o(x)$ if and only if $f(x) = 0$ for every $x \in \mathbb{R}$.

Problem 15:
Prove that if $f(x)$ is a polynomial, then $f(x) \in o(x)$ if and only if f has no linear or constant terms (i.e. $f(x) = a_0x^2 + a_1x^3 + \dots$).

So now that we have introduced all of the important technology, we can redefine the derivative formally.

Definition 16:

Important Definition

Given a function $f : \mathbb{R} \to \mathbb{R}$, and a point $x \in \mathbb{R}$, we define the *derivative* of f at r to be a linear map $A : \mathbb{R} \to \mathbb{R}$ such that $f(x + h) - f(x) - A(h)$ is of class $o(h)$.

Example 17:
Let $f(x) = x$. We want to find the derivative of $f(x)$ at $x = 1$. So we write

$$\begin{aligned} f(x+h) - f(x) - A(h) &= f(1+h) - f(1) - A(h) \\ &= 1 + h - 1 - A(h) \\ &= h - A(h) \end{aligned}$$

So if we let $A(h) = h$ (which is definitely linear!), we see that $f(x + h) - f(x) - A(h) = 0$, which is definitely of class $o(h)$.

Problem 18:
Compute the derivative of $f(x) = x^2$ at $x = 2$ using this method.

Problem 19:
Prove that the derivative of $f(x) = x^n$ at $x_0 = 1$ is $A(h) = nh$.
Hint: the Binomial Theorem will help.

Problem 20:
Prove that the derivative is unique. That is, if $f(x) : \mathbb{R} \to \mathbb{R}$ and $r \in \mathbb{R}$, then if we have two linear maps A and B so that $f(r+h) - f(r) - A(h) \in o(h)$ and $f(r+h) - f(r) - B(h) \in o(h)$, we have $A = B$.
This tells us that our linear approximation $f(x_0 + h) \approx f(x_0) + A(h)$ is the best possible linear approximation, since it is the unique linear approximation with error of class $o(h)$.

Definition 21:
Given a function $f(x) : \mathbb{R} \to \mathbb{R}$, we define $f'(x)$ (read as "the derivative of f") as the function where $f'(x)$ is the (unique) matrix A that is the derivative of f at x.

Example 22:
If $f(x) = x^2$, then at a point x_0, we have

$$\begin{aligned}(x_0 + h)^2 - x_0^2 - A_{x_0}(h) &= x_0^2 + 2x_0 h + h^2 - x_0^2 - A_{x_0}(h) \\ &= 2x_0 h + h^2 - A_{x_0}(h)\end{aligned}$$

So we can define $A_{x_0}(h) := 2x_0 h$ (which is linear), so that $2x_0 h + h^2 - A_{x_0}(h) = h^2 \in o(h)$. Then, our derivative is given by

$$f'(x_0) = A_{x_0} = \begin{bmatrix} 2x_0 \end{bmatrix}$$

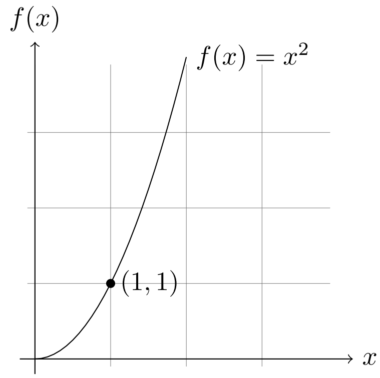
Since a 1-by-1 matrix is just a real number, we see that $f'(x_0) = 2x_0$. Thus, when $f(x) = x^2$, we have $f'(x) = 2x$.

Problem 23:
Find the derivative of $f(x) = x^n$.

Problem 24:
Prove that $(f(x)+g(x))' = f'(x)+g'(x)$ and that for $r \in \mathbb{R}$, $(r \cdot f(x))' = r \cdot f'(x)$.

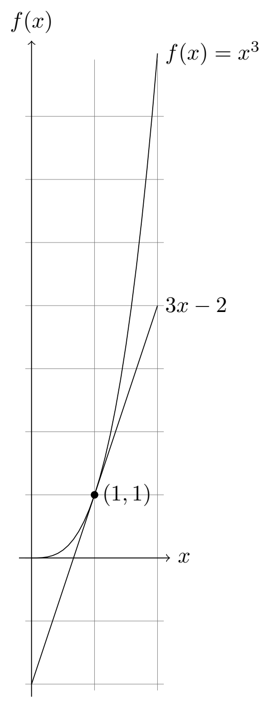
Part 2: Geometry of Derivatives

Problem 25:
Consider the graph of $f(x) = x^2$ below.



Notice that $1^2 = 1$, so $(1, 1)$ is on our graph. Draw some lines through $(1, 1)$. Is there a tangent line to our graph at $(1, 1)$? Is it unique?
So while most functions aren't linear, we can approximate most functions with linear functions.
Given a function $f : \mathbb{R} \to \mathbb{R}$, and a point $r \in \mathbb{R}$, we say that the tangent line to the graph of f at r is the **derivative** of f at r. (Yes, I know, we already defined derivatives. We'll prove they are the same at the end of this section.)

Example 26:
Let's visually find the derivative of $f(x) = x^3$ at $(1, 1)$.



If we graph the line tangent to $f(x) = x^3$ at $(1, 1)$ we see that it has a slope of 3 and hits the y-axis at $y = -2$. Thus, the line is $3x - 2$.

Problem 27:
Let $f(x) = x^2$ again. If I tell you that the slope of the tangent line of $f(x)$ at $(10, 100)$ is 20, is this enough information to find the entire line? That is, can you find the y-intercept from this information?

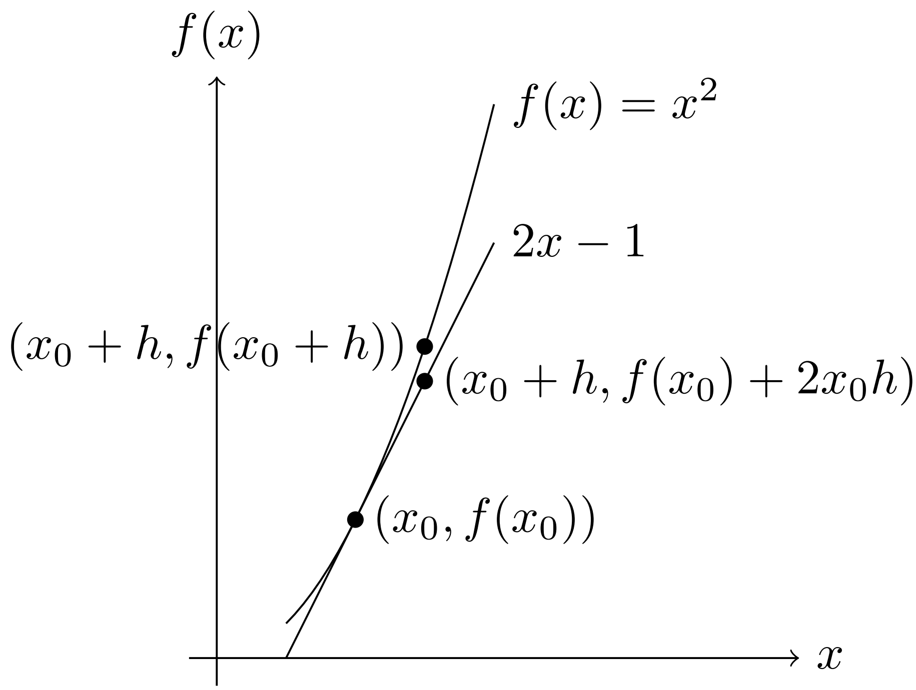
Problem 28:
Is $f(x) = 3x - 2$ linear?

So if the slope of the tangent line is enough information, let's just keep that information. So given a function $f : \mathbb{R} \to \mathbb{R}$, we can define the **derivative** $f'(x) : \mathbb{R} \to \mathbb{R}$, where $f'(x)$ is defined to be the slope of the tangent line to f at x.

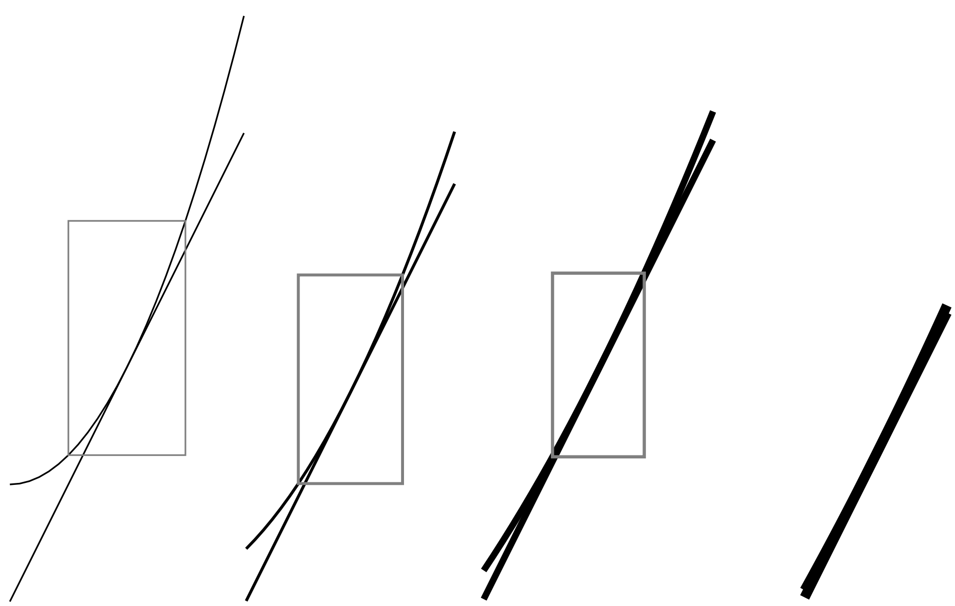
Problem 29:
Calculate the derivative of $f(x) = x^2$ by plugging in values, finding the slope, and guessing an answer.

This is really not a convenient way to calculate derivatives! But it is a very important way to *understand* derivatives that extends very naturally to higher dimensions.

How can we verify that this definition agrees with the definition we gave before? Let's look at an example where $f(x) = x^2, x_0 = 1, h = 0.5$.



In the above graph, we can see our best linear approximation $f(x_0) + 2x_0 h$ alongside $f(x_0 + h)$. Geometrically, it makes sense that the tangent line is the best linear approximation to a curve. If you zoom in on a curve enough, it starts to look like a straight line. Which line? The tangent line. Consider the graph of $f(x) = x^2$ zoomed in around $(1, 1)$. As we zoom in, the graph is indistinguishable from the tangent line! So the tangent line is definitely the best linear approximation.



Problem 30:
(For students who have seen derivatives before). You may have seen the derivative defined as

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

Prove that this definition agrees with our definition.
Hint: use Problem 7.

Part 3: Higher Dimensions

Let's consider a function $f : \mathbb{R}^n \to \mathbb{R}$ (for simplicity we will only work with scalar-valued functions today).
How can we define the derivative of f?
If we use the definition of the derivative you may have seen, we run into an immediate issue

$$f'(\vec{v}_0) \underset{?}{:=} \lim_{\vec{h} \to 0} \frac{f(\vec{v}_0 + \vec{h}) - f(\vec{v}_0)}{\vec{h}}$$

This is impossible to compute, since we can't divide vectors! We also don't know what $\vec{h} \to 0$ means!
But our definition we saw today does work.

Definition 31:
For a function $f : \mathbb{R}^n \to \mathbb{R}$ and a vector $v \in \mathbb{R}^n$, we define the **derivative** of f at r as the (unique, as we proved) linear map $A : \mathbb{R}^n \to \mathbb{R}$ so that

$$f(\vec{v} + \vec{h}) - f(\vec{h}) - A(\vec{h}) \in o(\vec{h}).$$

Note that $o(\vec{h})$ is in fact well-defined, as the class of functions $g : \mathbb{R}^n \to \mathbb{R}$ so that

$$\lim_{\vec{h} \to 0} \frac{|g(\vec{h})|}{|\vec{h}|} = 0$$

Note that now A is a matrix (and not a square matrix).

Example 32:
Let's find the derivative of $f(x, y) = x + y$ at $\vec{v} = (1, 1)$. Write $\vec{h} = (h_0, h_1)$. We have

$$\begin{aligned} f(\vec{v} + \vec{h}) - f(\vec{v}) - A(\vec{h}) &= f(1 + h_0, 1 + h_1) - f(1, 1) - A(h_0, h_1) \\ &= 1 + h_0 + 1 + h_1 - (1 + 1) - A(h_0, h_1) \\ &= h_0 + h_1 - A(h_0, h_1) \end{aligned}$$

So we can let $A = \begin{bmatrix} 1 & 1 \end{bmatrix}$ so that $A(\vec{h}) = A(h_0, h_1) = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = h_0 + h_1$. So the derivative of $f(x, y) = x + y$ is $\begin{bmatrix} 1 & 1 \end{bmatrix}$, a 1-by-2 matrix.

Problem 33:
Let $f(x), g(x) : \mathbb{R} \to \mathbb{R}$. Prove that the derivative of $h(x, y) = f(x) + g(y)$ at (x_0, y_0) is $\begin{bmatrix} f'(x_0) & g'(y_0) \end{bmatrix}$.

We can also define the derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ as a function f'. But it isn't a function from $\mathbb{R}^n$ to $\mathbb{R}$!

Problem 34:
For a function $f : \mathbb{R}^n \to \mathbb{R}$, what is the domain and codomain of f'?

Instead, we can consider the **coordinate functions** of the derivative. Recall that for a matrix $A = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$, we have $A(1, 0, \dots, 0) = a_1, A(0, 1, 0, \dots, 0) = a_2, \dots, A(0, 0, \dots, 0, 1) = a_n$. For simplicity, we write e_i for the vector that is all 0s except a 1 in the ith spot, so that $A(e_i) = a_i$.

Definition 35:
If we have a function K so that $K(x)$ is a 1-by-n matrix, we define the **coordinate function of K in the ith position** to be the function $K'_i(x) := (K'(x))(e_i)$.

Problem 36:
For a function $f : \mathbb{R}^n \to \mathbb{R}$, what is the domain and codomain of f'_1?

Problem 37:
If $f : \mathbb{R}^2 \to \mathbb{R}$ is the function given by $f(x, y) = x^2 + y$, find $f'_1(x)$ and $f'_2(x)$.

Definition 38:

Important Definition

Given a function $f : \mathbb{R}^n \to \mathbb{R}$, we define the **partial derivative of f with respect to x_i** to be $f'_i : \mathbb{R}^n \to \mathbb{R}$.

We denote this by

$$\frac{\partial f(x)}{\partial x_i} = \frac{\partial}{\partial x_i} f(x)$$

Typically, we don't use $x_1, x_2, x_3, \dots$ but $x, y, z, \dots$. Then we can say

$$\frac{\partial}{\partial x} f(x, y, z) = f'_1(x, y, z)$$

$$\frac{\partial}{\partial y} f(x, y, z) = f'_2(x, y, z)$$

$$\frac{\partial}{\partial z} f(x, y, z) = f'_3(x, y, z)$$

Note that the order of the variables matters! $\frac{\partial}{\partial x} f(x, y, z) \neq \frac{\partial}{\partial x} f(y, x, z)$!

Problem 39:
Compute $\frac{\partial}{\partial x} f(x, y)$ if $f(x, y) = 3x - y$.

Example 40:
What if we want to compute $\frac{\partial}{\partial x} f(x, y)$ if $f(x, y) = x^2 y - y^{100} \log(y) - \cosh(y)^\pi + \operatorname{arccsch}(\operatorname{erfi}(y))$?
We can't separate the x and y part in order to use Problem 33.
The terms including x are very simple, but rest of the terms are exceptionally complicated. The only process we know now is computing $f'(x, y)$ as a matrix and finding its coordinate functions. But this is very hard! It would be very convenient if we could just ignore y when computing a partial derivative with respect to x...

Problem 41:
Compute $\frac{\partial}{\partial x} x^2 y$.

We can define a new function, $g(x)$, given by $g(x) = f(x, 2) = 2x^2$.

Problem 42:
Find $g'(x)$.

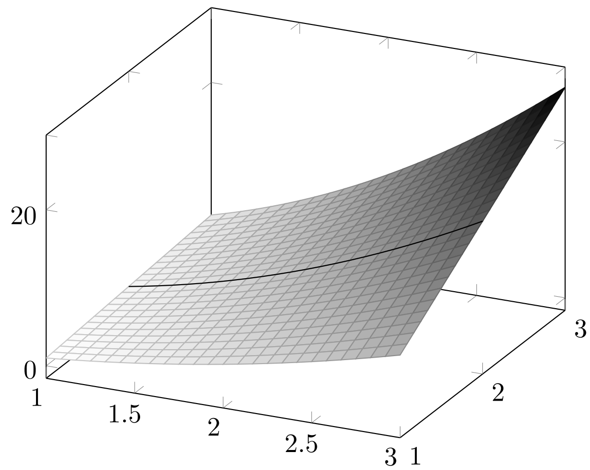
More generally, for any $y_0 \in \mathbb{R}$, we can define $g_{y_0}(x) = f(x, y_0)$. Our g before would then be $g_2(x)$.

Problem 43:
For any $y_0 \in \mathbb{R}$, find $(g_{y_0}(x))'$.

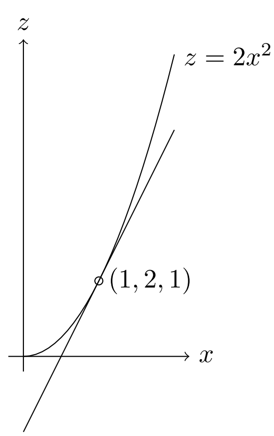
Problem 44:
Define $h(x, y) = (g_y(x))'$. Is this well-defined? Can you write an explicit formula for $h(x, y)$?

Do you notice a connection between the answers to Problem 41 and Problem 44? We will prove that these two processes produce the same output!

Example 45:
Let's consider $f(x, y) = x^2y$. Here is a plot of $f(x, y)$. The curve is the cross-section at $y = 2$, given by the curve $z = 2x^2$.

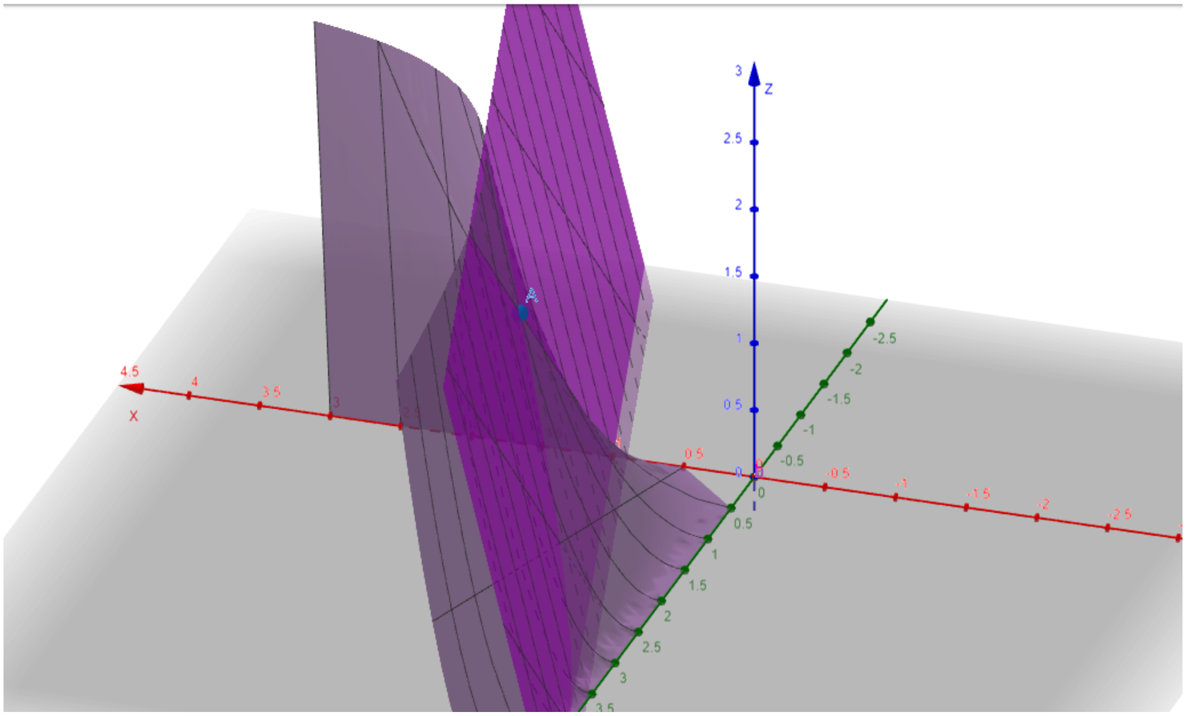


This curve is the graph of the function $g_2(x) = 2x^2$; in general, the graph of $g_{y_0}(x)$ is precisely the cross-section of the graph of $f(x, y)$ at $y = y_0$.
We don't know what it means to ask what the derivative of this curve is in three dimensions. But if we ignore the y-axis (since we set $y = 2$), we can look at this curve in the $x - z$ plane (scaled down so it doesn't take up the entire page):

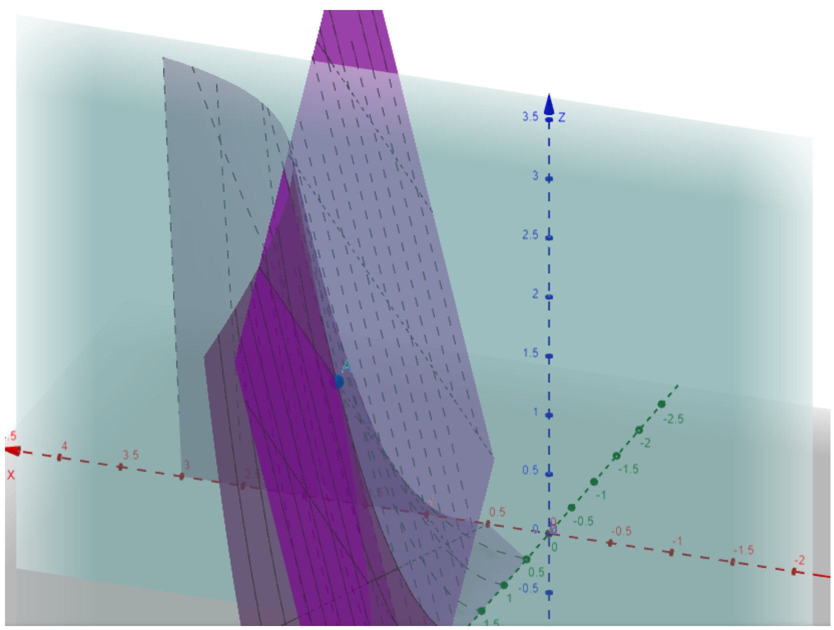


Here, the derivative is something we know how to find. For our example, let's find it at $x = 1$ – it is $g_2'(1) = 4x$. In general, we can take the cross-section at $y = y_0$ and compute the derivative as we did before, to get $(g_{y_0})'(x) = 2xy_0$.

On the other hand, the derivative of $f(x, y) = x^2y$ at $(1, 2)$ can be thought of as the *tangent plane* to the surface at $(1, 2)$. (This is hard to visualise!)



Remember that $\frac{\partial}{\partial x} f(x,y)$ is defined to be the first coordinate of $f(x,y)$. This is the same thing as taking the cross-section of our plane in the x axis (i.e. taking the intersection of this plane with the plane $y = 2$). The intersection of two planes is a line, and the line is exactly the tangent line we found before!



So our two methods produced the same line, and therefore produce the same value of the derivative. Nothing is special about $f(x,y) = x^2y$ – this works in general. So we see that for any $f : \mathbb{R}^2 \to \mathbb{R}$, with $g_{y_0}(x) := f(x, y_0)$, that

$$\left(\frac{\partial}{\partial x} f(x,y)\right)(x_0, y_0) = (g_{y_0})'(x_0)$$

Problem 46:
Calculate $\frac{\partial}{\partial x} f(x,y)$ and $\frac{\partial}{\partial y} f(x,y)$ for $f(x,y) = xy$ using both methods and verify they are the same.

Problem 47:
Calculate $\frac{\partial}{\partial x} f(x,y)$ for the function $f(x,y)$ given in Example 40.