**Exercise 2.** Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a real-valued function on $\mathbb{R}^n$, and $\mathbf{x}_0 = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ a point.

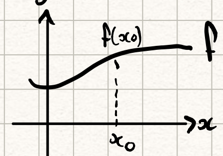
**(a)** If $f$ is differentiable at $\mathbf{x}_0$, explain what it means to say that the function $z = h(\mathbf{x})$, where

$$h(\mathbf{x}) = f(\mathbf{x}_0) + Df(\mathbf{x}_0)\,(\mathbf{x} - \mathbf{x}_0)$$

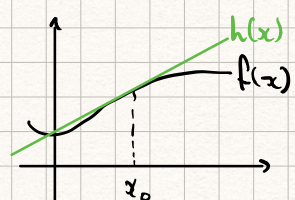
is the best linear function to approximate $f$ at $\mathbf{x}_0$.

This is described on page 109 (§2.3) of the textbook. We flesh it out here.

First, what is this question asking? In the case $n=1$, suppose the graph of our function looks like this:



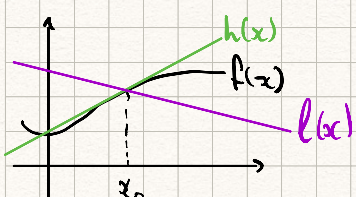

We may superimpose $h(x)$ on this,




and the question is then, of all other straight lines passing through $(x_0, f(x_0))$, that is, functions of the form

$$l(x) = f(x_0) + m(x - x_0),$$

why is the case where $m = f'(x_0)$ the best linear approximation to $f$?




Some other linear approximation.

To answer this question we have to decide how we measure "best".

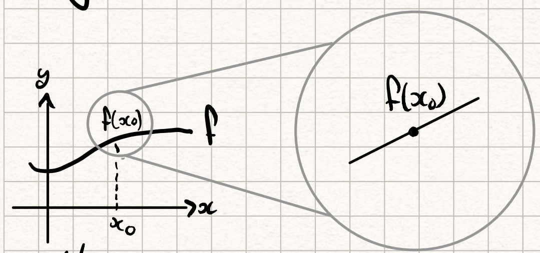
One way to think about what best should be is to measure how well the line approximates the function close to $x_0$, that is, look at

$$f(x) - l(x)$$

but amplify any errors in the region. If $x$ is very close to $x_0$ then $x - x_0$ is very small and so $1/x - x_0$ is rather large.

So the expression $\dfrac{f(x) - l(x)}{x - x_0}$ measures how well $l(x)$ approximates $f(x)$ near $x_0$, and amplifies any errors more and more as $x \to x_0$.

In another view, when we zoom in a lot to around $x_0$, $f(x)$ should begin to behave like a line passing through $f(x_0)$




and so $\dfrac{f(x) - l(x)}{x - x_0}$ is looking at the difference in 'gradients' between $f(x)$ and $l(x)$ very close to $x_0$.

thus

Thus, for $l(x)$ to be the best linear approximation to $f$ at $x_0$ we ask for $\lim\limits_{x \to x_0} \dfrac{f(x) - l(x)}{x - x_0} = 0$, that is, for the 'gradient' of $f$ to match that of $l$.

If $l(x) = f(x_0) + m(x - x_0)$ we can now solve for $m$ under the constraint that $l$ is the best linear approximation to $f$ at $x_0$.

$$0 = \lim_{x \to x_0} \frac{f(x) - l(x)}{x - x_0} = \lim_{x \to x_0} \frac{f(x) - f(x_0) - m(x - x_0)}{x - x_0}$$

$$= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} - \lim_{x \to x_0} m$$

thus $m = \lim\limits_{x \to x_0} \dfrac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$ and so the tangent line is the best linear approximation to $f(x)$ at $x_0$.

The same is true in higher dimensions, $\mathbb{R}^n \to \mathbb{R}$, where tangent line becomes plane becomes hyperplane and the textbook discusses this briefly on page 110.