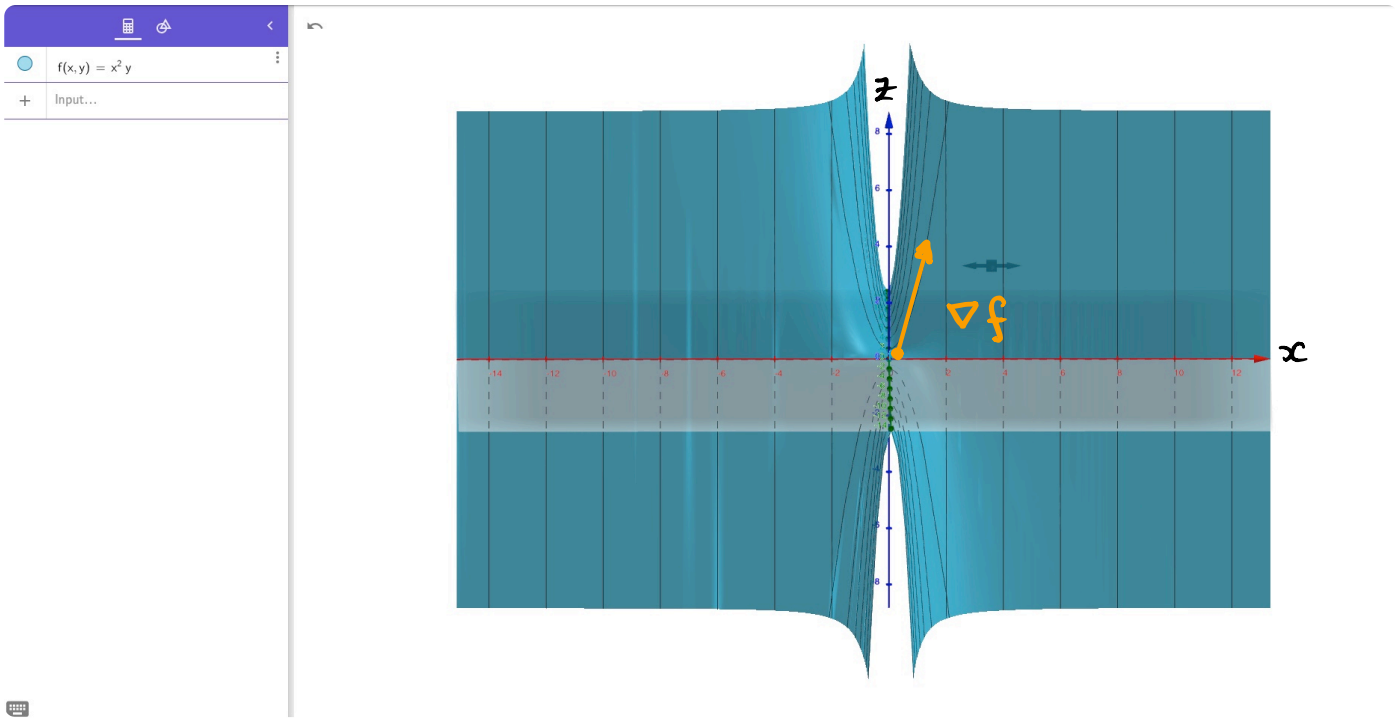
The Newton-Raphson method used the gradient to iteratively solve a 1D function. That process can be generalized to a multivariate function – the process of iteratively using the gradient.

A direct application of this technique is finding maxima and minima in optimization problems.



The grad is oriented in the direction of the greatest increasing gradient.

It can be calculate from partial derivatives.

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

$$\frac{\partial f}{\partial x} = 2xy$$

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

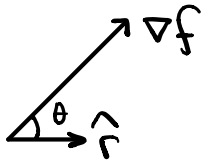
$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 2xy \\ x^2 \end{bmatrix}$$

The grad can be used as a multiplier to calculate the next parameter (e.g $x, y, z, \ldots$) values closer to the local maxima/minima. The evaluation of this technique produces the directional gradient

$$u = \nabla f \cdot \hat{r}, \text{ where } \hat{r} \text{ is a unit vector}$$

$$= \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} r_0 \\ r_1 \end{bmatrix}, \quad \sum_{i=0}^{N} r_i^2 = 1$$

① Question #1 - How great can the magnitude of a directional unit vector be?



$\nabla f \cdot \hat{r} = \|\nabla f\| \|\hat{r}\| \cos\theta$. $\|\hat{r}\|$ can't change bc. its a unit vector.

Only the angle can change - $\cos\theta$.
The directional gradient will be at it's greatest when,

$$\cos\theta = 1 \iff \theta = 0.$$
$$\iff \hat{r} \parallel \nabla f, \text{ parallel.}$$

This means the max directional gradient is when,

$$\hat{r} = \frac{\nabla f}{\|\nabla f\|} \iff \nabla f \cdot \hat{r} = \nabla f \cdot \frac{\nabla f}{\|\nabla f\|} = \frac{\|\nabla f\|^2}{\|\nabla f\|} = \|\nabla f\|$$

A corollary of this is that is the max size of the directional vector is $\|\nabla f\|$.

② Question #2 - Which direction does ∇f point?

It points in the direction of steepest ascent.
That is perpendicular to the contour lines
in the direction of the nearest local maxima.

Differing from the Newton-Raphson method the application of grad is not to find the roots of a function. Instead it is for minima/maxima. This has a name - called Gradient Descent.

Gradient Descent is an iterative computation of the form.

$$S_{n+1} = S_n - \gamma \cdot \nabla f(S_n)$$, where S_n is a step,

S_{n+1} is the next step,

$\nabla f(S_n)$ is the gradient @ S_n,

γ is an amount.

Essentially a small step up/down the hill from the point S_n.

---

(1) Overshooting is okay - grad will redirect

(2) Closer to the turning point grad will become smaller!

(*) There are some problems with gradient descent - the method isn't globally aware and find itself lost in local extrema.

Grad - Gradient Descent - is a foray into vector - calculus!!!