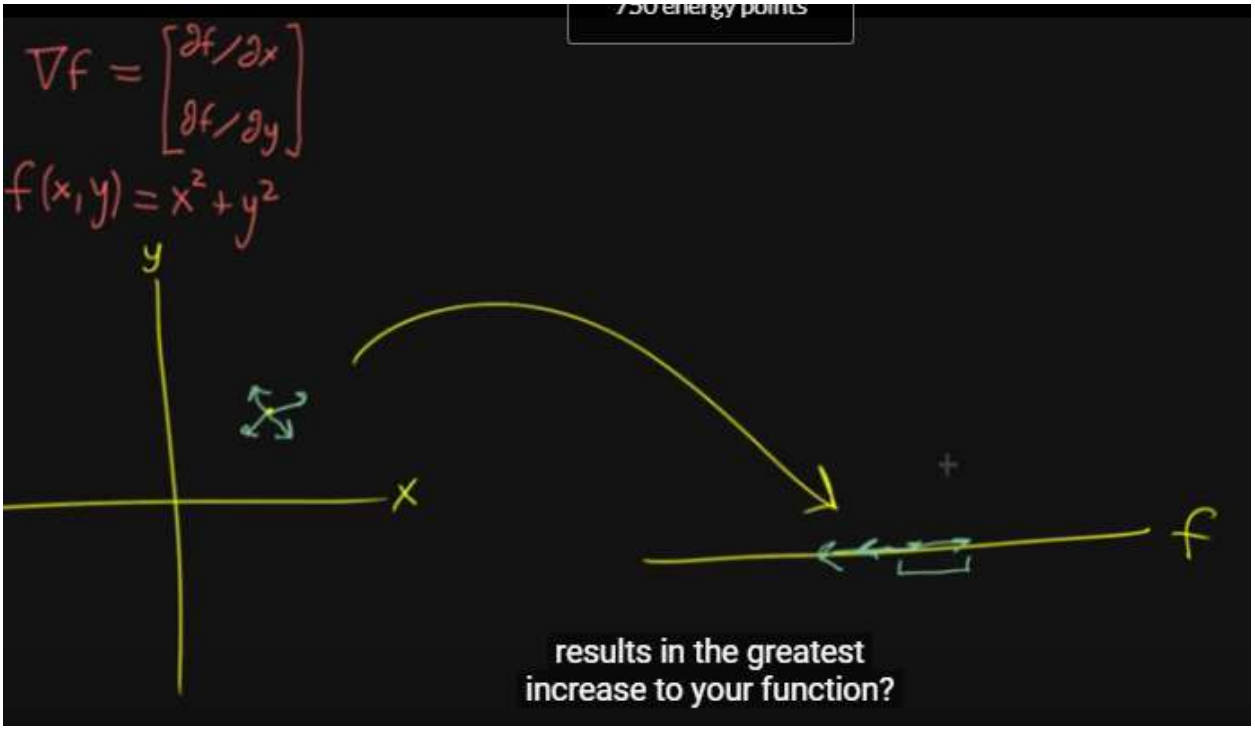
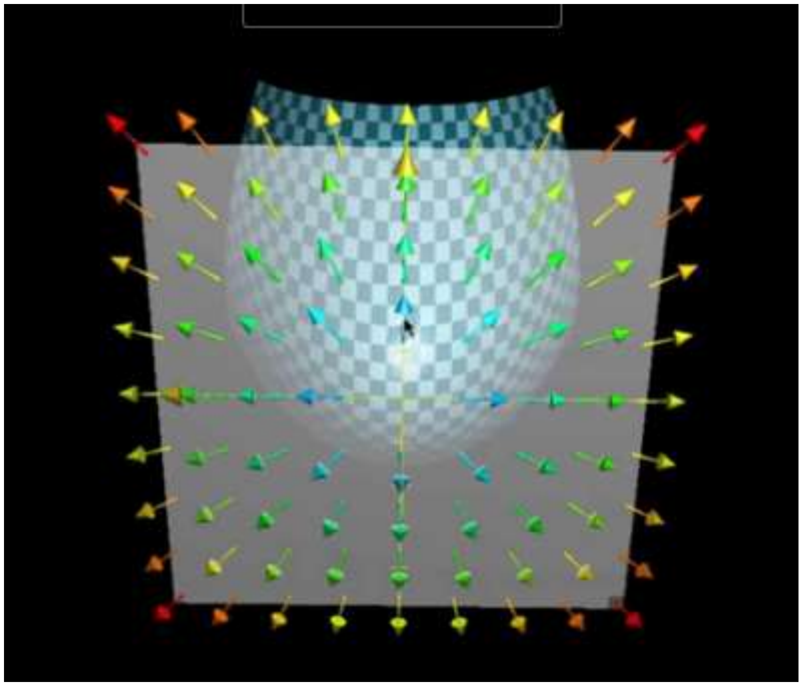
Why the gradient is the direction of steepest ascent



- In red we have the computation of the gradient for function f
    - Which is simply a vector of partial derivatives

- Graphical interpretation: The gradient always points in direction of steepest ascent
    - Out of all the different direction we can nudge the point above (In XY space), which one results in the greatest increase to your function f

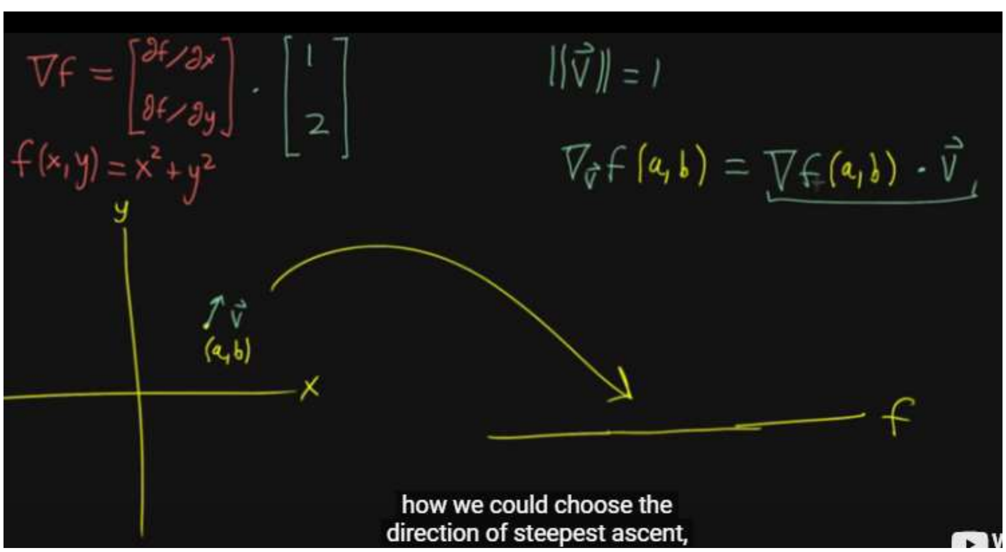


Here we have mapped out our function f and show the gradient field at the bottom

- The field shows the direction to steepest accent based on where a point lies in the function

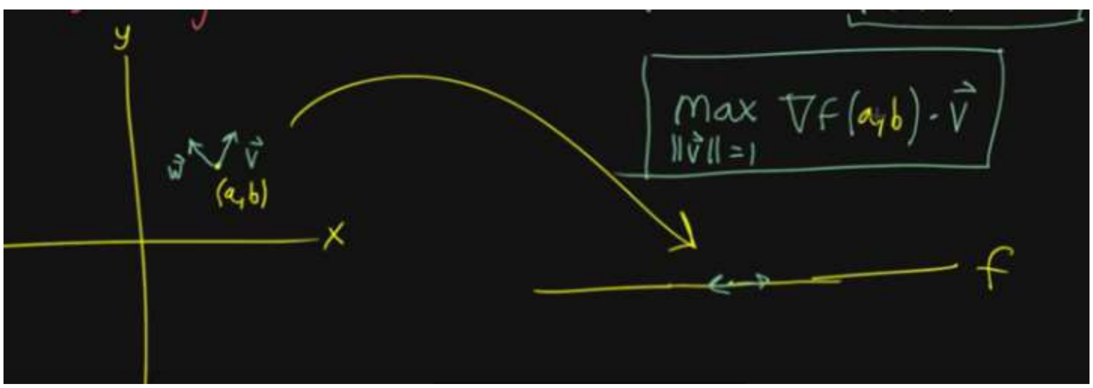
Comparing these two visuals, it may be difficult to initially see the connection

- Why does combination of partial derivatives coincide with choosing the direction of steepest accent?
  - Intuition:

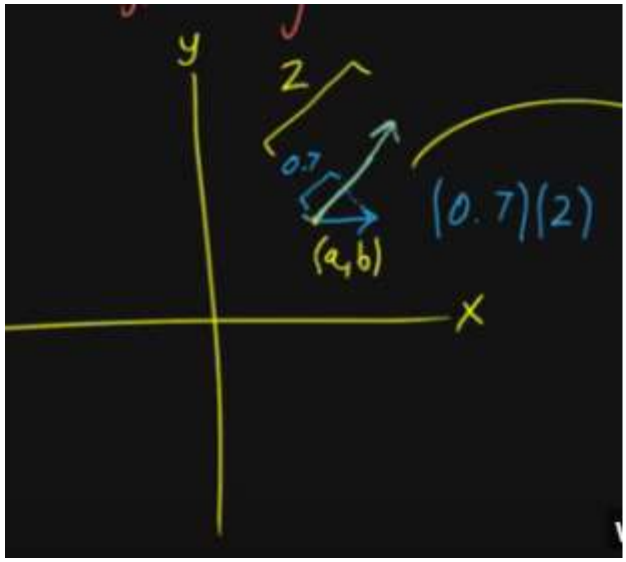


        - Lets say we have single unit vector v
            - Magnitude of 1
        - You can tell rate at which the function changes when it moves in direction of v by :
            - Taking the directional derivative of function with respect to vector v at some point (a, b)
                - The way we do this is dotting the gradient of f at point (a, b) together with unit vector v
                    - This is how you tell the rate of change
                    - Imagine you were dotting the gradient together with vector [1 2]
                        - Vector represents 1 step in x direction, 2 steps in y direction
                        - The amount that the function changes should be the partial derivative with respect to x times a 1 step change in x direction plus df/dy times 2 step change in y direction
                        -

- This starts to give us the key for how we could choose the direction of steepest ascent
    - What we are really asking when we say which direction of the unit vector at point (a,b) changes function f the most:
        - Find max for all unit vectors v (magnitude = 1) of the dot product of gradient f at point (a,b) times vector v

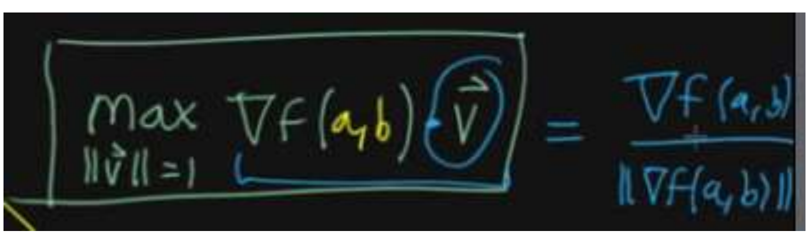


    - What does dot product represent?
        - Dot product of gradient of f at point (a,b) and vector v
            - Length of projection of vector v onto gradient vector times length of gradient vector

- 

  - Maximized when vector v is in same direction as gradient vector because then calculation would be (1)*(2) = 2 and you return length of gradient vector
    - 

- Thus: the direction of steepest accent is the direction of the gradient vector.
  - We have to normalize the gradient vector in case it has a magnitude larger than 1 because we are interested purely in direction here



- Fundamental fact: The gradient is a tool for computing directional deriatives
  - Gradient is vector that loves to be dotted with other things
    - (what a rascal that gradient)

- Because of this: the direction of steepest ascent is that vector itself: because the max of the dot product with the gradient vector and any other vector is a vector that points in same direction as gradient vector

- This leads into greater intuition for what we should do with the length of the gradient
  - The directional derivative in the direction of gradient itself has a value equal to the magnitude of the gradient
    - When you are moving in direction of gradient, the rate at which the function changes is given by the magnitude of the gradient (MAGIC VECTOR!)