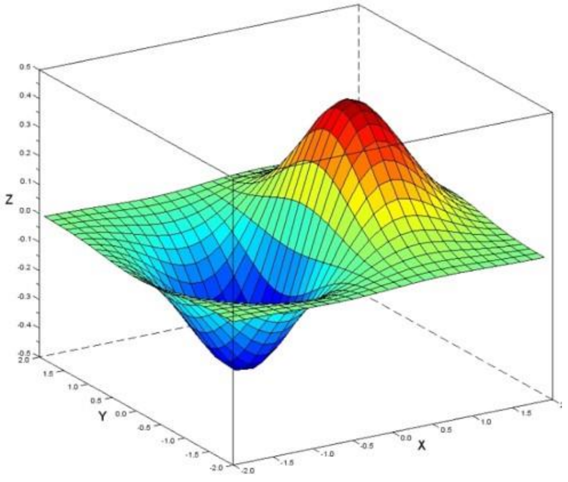
Calculus

Looking at functions in detail



Yordan Darakchiev
Technical Trainer
iordan93@gmail.com

sli.do

#MathForDevs

Table of Contents

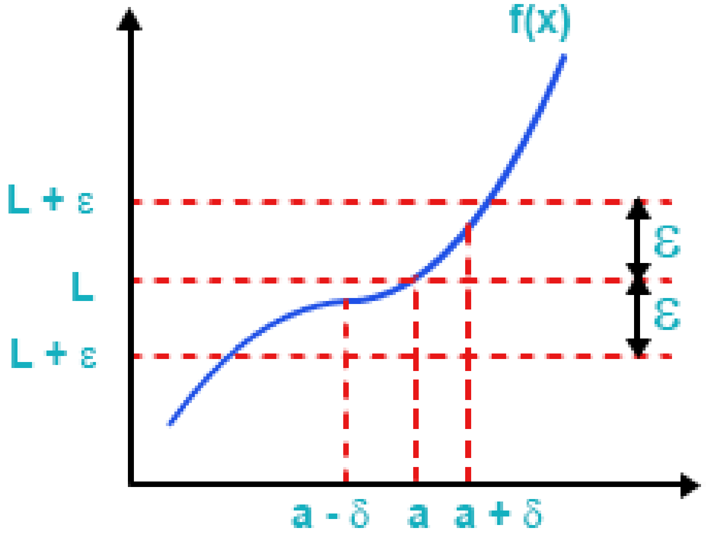
- Limits
- Derivatives
- Integrals
- Calculus in many dimensions
- Gradient descent

Limits

Approaching places

Limit

- Natural definition
  - Given a function $f(x)$, "nudge" the input around a given value a
    - As a result, the function value changes
  - Limit of $f(x)$ at the point $x = a$: what f approaches as x approaches a
- Notation: $\lim_{x \to a} f(x) = L$
- Mathematical definition
  - Gives us a nice way to define "approaching a value"
  - For any positive δ and ε
    - If $0 < |x - a| < \delta$
    - Then $|f(x) - L| < \varepsilon$
  - Also called "epsilon-delta" definition
  - What are these numbers? Arbitrary, they only need to be positive
    - It's very useful to **make them really small**



Limits in Python

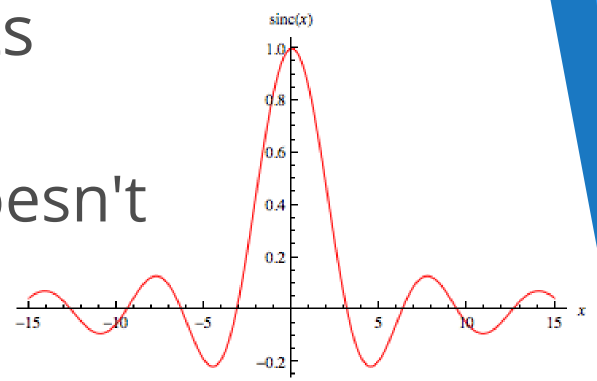
- To find the limit of a function at a point, just apply the definition
  - Generate several values of x around a
    - Don't forget to include positive and negative "nudges"
  - Print the function values at those points

```
def get_limit(f, a):
  epsilon = np.array([
    10 ** p
    for p in np.arange(0, -11, -1, dtype = float)])

  x = np.append(a - epsilon, (a + epsilon)[::-1])
  y = f(x)
  return y

print(get_limit(lambda x: x ** 2, 3))
print(get_limit(lambda x: x ** 2 + 3 * x, 2))
print(get_limit(lambda x: np.sin(x), 0))
```

More Limits

- Some functions don't have a value at certain points
  - But they are defined "around" these points
  - **The limit exists** even though the function value doesn't

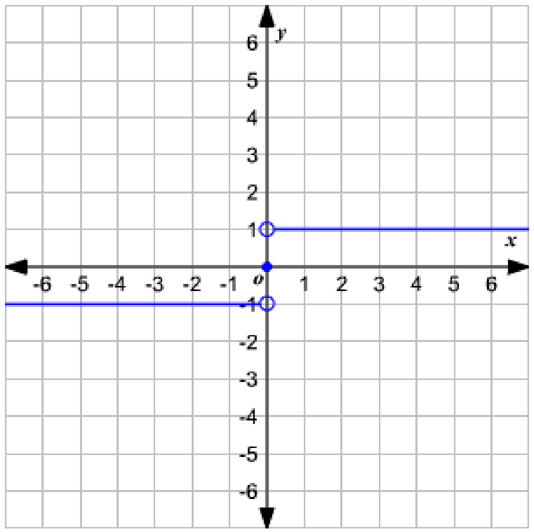
$$\lim_{x \to 0} \frac{\sin(x)}{x} = 1$$



- Some limits can be infinite: $\lim_{x \to \infty} x^2 = \infty$
- Some functions "jump"
  - The limits "from the left" and "from the right" are different
    - Therefore, the limit is not defined
    - We say the function is not continuous at that point
  - Example:
    - In this case, $f(0) = 0$ but the limit does not exist

$$f(x) = \begin{cases} -1, x < 0 \\ 0, x = 0 \\ 1, x > 0 \end{cases}$$

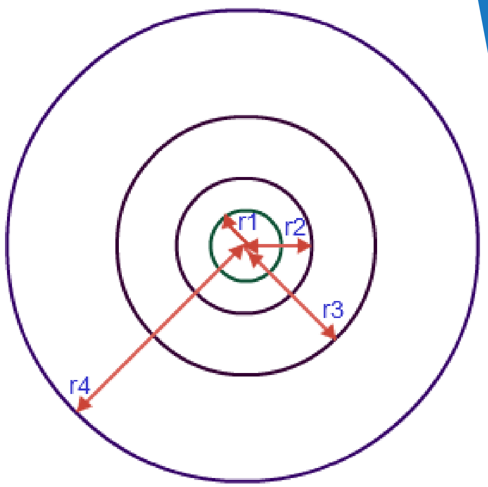
$$\lim_{x \to 0^-} f(x) = -1; \lim_{x \to 0^+} f(x) = 1$$



Derivatives

Slope and velocity

Calculus Motivation



- Say you want to compute the area of a circle
  - It is πR^2 but why?
  - Remember how you can divide a shape into simpler shapes and sum their areas to get the total area
    - One way: cut it like cake: see [this video]
    - Another way: concentric rings
  - If you "cut" and "straighten" each ring, you'll get a trapezoid
    - If your ring is very, very thin; it will actually be close to a rectangle
    - Example:

$2\pi r_2$

$r_3 - r_2$

$2\pi r_3$

    - **Set the difference to be very, very, veeeeeeery small:** $r_3 - r_2 \to 0$
    - ... and you get calculus :)
- Even in this simple example, there are the notions about derivatives and integrals; even the fundamental theorem of calculus

Derivatives and Velocity

- We all know that $v = \frac{s}{t}$
  - But that's mostly useless
  - Travelling is not done at a uniform velocity, it's not a fixed number but a function of time: $v = v(t)$
- Instantaneous velocity: $v(t_0) = v(t)|_{t=t_0}$
- Computing instantaneous velocity from travelled distance
  - Say, $s(t) = t^2$; say we start at $t = 0s$ and finish at $t = 5s$
    - Final distance: $s(5) = 5^2 = 25m$
  - Average speed: $\frac{25}{5} = 5\frac{m}{s}$
  - But we cover different distances for the same time
    - From $0 \leq t \leq 1$: $s(1) - s(0) = 1 - 0 = 1m$
    - From $3 \leq t \leq 4$: $s(4) - s(3) = 16 - 9 = 7m$
    - From $4 \leq t \leq 5$: $s(5) - s(4) = 25 - 16 = 9m$
    - And neither of these is even close to the average speed

Derivatives and Velocity (2)

- Let's calculate the instantaneous velocity
  - Fix time at $t = 3$
  - But... how can we move if time is fixed?
- Let's apply our previous idea
  - Nudge time a tiny bit and see how the distance changes
  - $t = 3{,}01$: $v \approx \frac{s(3{,}01) - s(3)}{3{,}01 - 3} = \frac{3{,}01^2 - 3^2}{0{,}01} = 6{,}01 \frac{m}{s}$
  - $t = 3{,}00001$: $v \approx \frac{s(3{,}00001) - s(3)}{3{,}00001 - 3} = \frac{3{,}00001^2 - 3^2}{0{,}00001} = 6{,}00001 \frac{m}{s}$
    - More generally, if we nudge time from $t = t_0$ to $t = t_0 + \Delta t$, we'll get an approximation of the instantaneous velocity:

$$v \approx \frac{s(t + \Delta t) - s(t)}{t + \Delta t - t} = \frac{s(t + \Delta t) - s(t)}{\Delta t}$$

    - This approximation will get increasingly more accurate as Δt becomes smaller
    - Smaller $\Delta t \Rightarrow$ better approximation of v

Derivatives and Velocity (3)

- How does the velocity behave as $\Delta t \to 0$?
  - Note that we **cannot** set $\Delta t = 0$, this will freeze time
  - Math notation: if $\Delta t \to 0$, we write it as dt

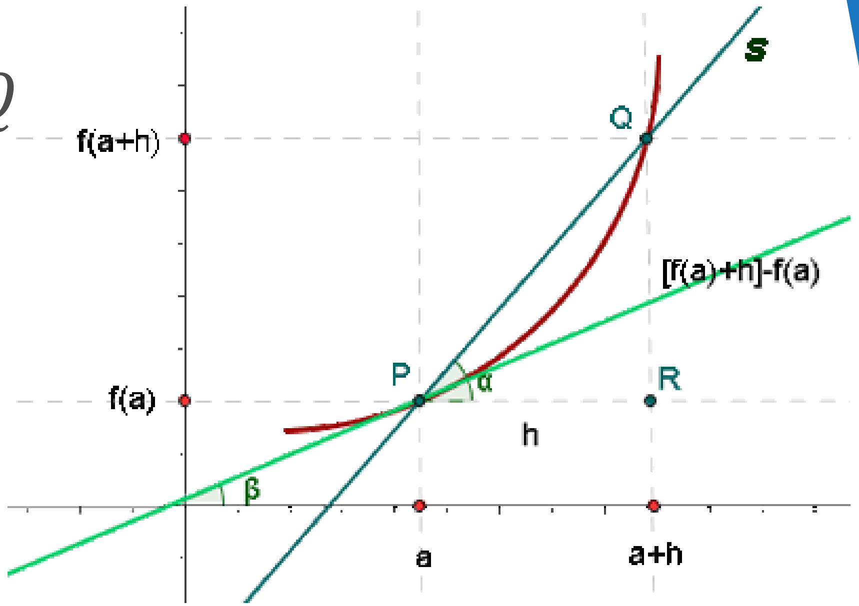
$$v(t) = \lim_{dt \to 0} \frac{s(t + dt) - s(t)}{dt}$$

- We now have a nice definition of velocity
  - But what does it mean mathematically?
    - Velocity = rate of change of travelled distance over time
  - The rate of change of a function $f(x)$ as its argument x changes, is called the **first derivative** of $f(x)$ with respect to x
  - Math notation: $f'(x)$ or $\frac{df}{dx}$
    - Note that $\frac{df}{dx}$ is only notation, it is not equal to $\frac{f}{x}$
  - Definition: $f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$

Geometric Interpretation

- $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$
- Look at the chord PQ and the triangle PRQ
- As $h \to 0$, Q approaches P
  - The chord becomes the same as the tangent line at point P
  - The angle $\alpha = \beta$: slope of the tangent line

$$\tan(\alpha) = \lim_{h \to 0} \frac{\Delta f}{h} = f'(x)$$



- Geometrically, the derivative at a given point is **equal** to the slope of the tangent line to the function at this point
- This is what calculus is all about
  - **Zooming in really close** until everything appears as a straight line

Calculating Derivatives

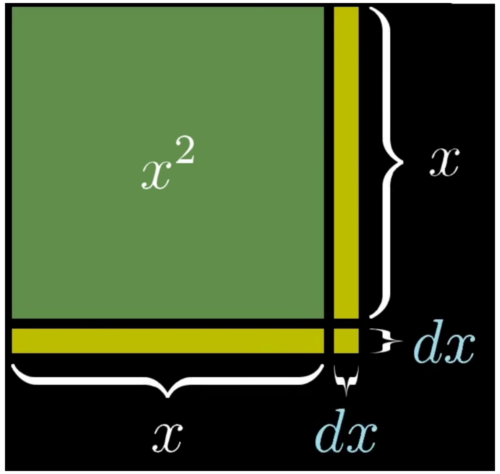
- Note that we have two definitions
  - Derivative of $f(x)$ at a fixed point x (e.g. $x = 5$): this is a number
  - Derivative of $f(x)$ at any point: this is another function
- Calculate the derivative of $3x^2 + 5x - 8$ at $x = 3$
  - We're doing a numerical approximation
  - We can't work with infinitesimally small h but we can get away with something quite small

```python
def calculate_derivative(f, a, h = 1e-7):
  return (f(a + h) - f(a)) / h

print(calculate_derivative(lambda x: 3 * x**2 + 5 * x - 8, 3))
# 23.0000026878024
```

- We can also do this **analytically**
  - A fancy term for "with pen and paper"

Calculating Derivatives Analytically

- Let's take a relatively simple function like $f(x) = x^2$



$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h}$$

$$= \lim_{h \to 0} \frac{x^2 + 2hx + h^2 - x^2}{h} = \lim_{h \to 0} \frac{2hx + h^2}{h}$$

- We're looking for approximation and h is small, so let's ignore h^2
  - Ignoring higher-order terms is completely valid (and is done often)

$$\Rightarrow f'(x) = \lim_{h \to 0} \frac{2hx}{h} = 2x$$

  - Note that the derivative **does not depend** on the tiny shift h
- We can do this for every function
  - We have precomputed [tables of derivatives]

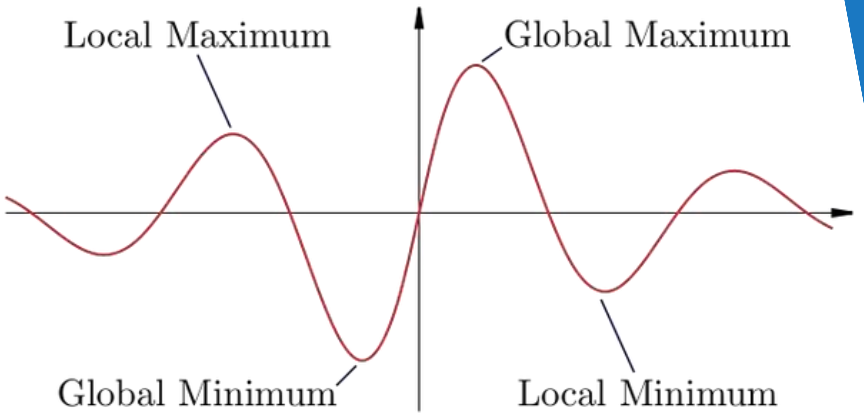
Properties of Derivatives

- The derivative of a constant ($f(x) = c$) is 0
- Derivatives are linear
  - $(f \pm g)' = f' \pm g'$
  - $(\lambda f)' = \lambda f'$
- Product rule
  - $(f.g)' = f'.g + f.g'$
  - $\left(\frac{f}{g}\right)' = \frac{f'.g - f.g'}{g^2}$
- Derivative of a function composition
  - Also called **chain rule**
  - $f\big(g(x)\big)' = f'\big(g(x)\big).g'(x)$
  - Looks better in the other notation: $\frac{df}{dx} = \frac{df}{dg}.\frac{dg}{dx}$
- We can prove these using the geometric intuition or the definition
  - This is left as an exercise for the reader :)

Higher-Order Derivatives

- The second derivative of a function is the first derivative of its first derivative
  - Interpretation: "rate of change of the rate of change"
  - ... a.k.a. acceleration
  - Notation: $f''(x) = \left(f'(x)\right)', \frac{d^2 f}{dx^2} = \frac{d}{dx}\left(\frac{df}{dx}\right)$
- This can be applied arbitrary many times
  - E.g. rate of change of acceleration: third derivative
    - a.k.a. "jerk"... don't ask me why
  - Third, fourth, etc. derivatives; n-th derivative notation: $f^{(n)}(x)$
    - E.g. $f^{(6)}(x)$

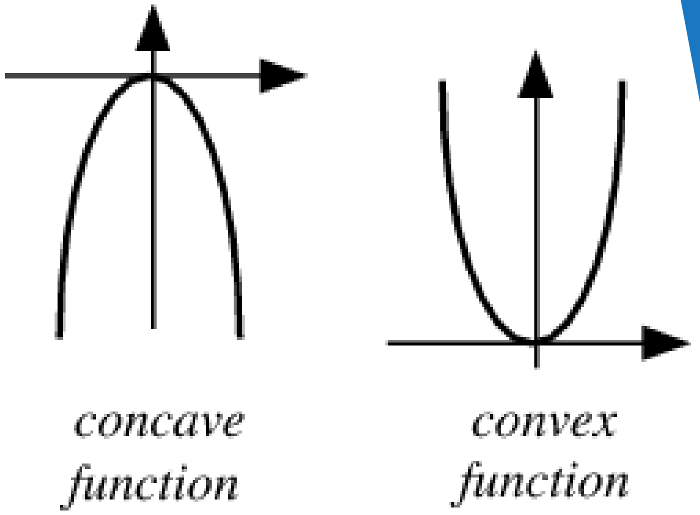
Function Extrema



- Even if we don't know the function, its derivatives give us useful information
- Consider the drawn function
  - The smallest value of $f(x)$ is called a **global minimum**
  - Conversely, largest value: **global maximum**
- These are collectively called extrema (plural of extremum)
- Smallest / largest value of $f(x)$ in a tiny range: local min / max
- More formally, we say $f(x)$ has a maximum at, say, $x = 5$ if the function value $f(5)$ is bigger than the function values immediately to the left and right
  - The complete definition involves limits
  - The points x of min / max (e.g. $x = 5$) are called **critical points**

Function Extrema (2)

- Notice how the tangent line behaves
  - At max / min, $f' = 0$
  - Around max / min, f' changes its sign
- Also notice that if $f'(x) > 0$ in a given interval, the function increases
  - If $f'(x) < 0$, the function decreases
- Therefore, if f behaves like this
  - Increasing; stop; decreasing ⇒ local maximum
  - Decreasing; stop; increasing ⇒ local minimum
- The second derivative gives us more information about whether the function is "concave up" or "concave down"
  - More specifically, its sign
  - These are sometimes called convex and concave functions



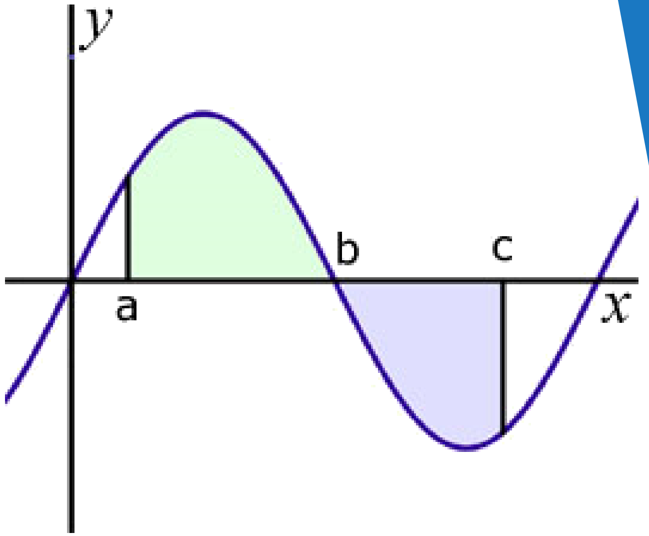
Integrals

Areas and accumulation

Area under a Function

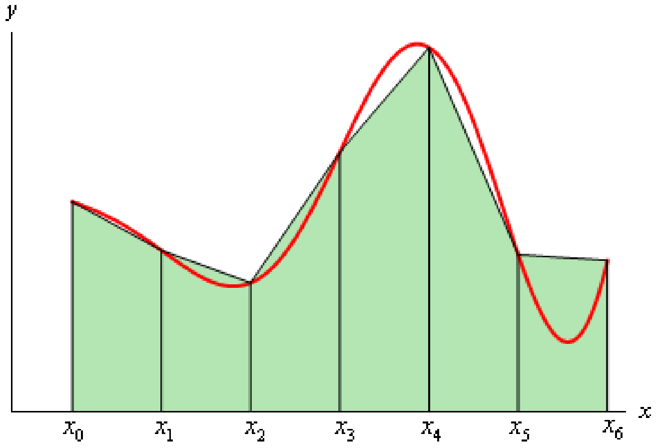
- Look back to the motivating example
- How can we find the area S "under" a curve given by a function?
  - What is the shaded area ($S < 0$ if $f < 0$)?
- **Approach:** approximate and zoom in
- Divide the x-axis into equal intervals Δx
- Approximate the area with trapezoids

$$S = \sum_i S_i$$



- If the intervals in x are really small, the trapezoids will look like rectangles

$$S_i = f(x_i)\Delta x$$



- Smaller $\Delta x \Rightarrow$ better approximation

Integral of a Function

- At the limit, $\Delta x \to 0$, so we write dx
- The sum is denoted differently: $\int_a^b f(x)dx = \lim_{\Delta x \to 0} \sum_{x=a}^{b} f(x_i)\Delta x$
  - This is called the **definite integral** of $f(x)$
  - **Note:** don't forget the dx after the function!
- **Indefinite integral**: the same, without the end points
  - Like derivatives, the definite integral is a number
  - The indefinite integral is a function of x
- Calculating integrals
  - Analytically – very difficult (unlike derivatives)
  - Numerically – apply the trapezoidal rule
    - Use a small number dx, like before

Fundamental Theorem of Calculus

Putting it all together

Antiderivatives

- The antiderivative $F(x)$ of a function $f(x)$ is such a function that $F'(x) = f(x)$
  - It's also called the primitive function of $f(x)$
  - Note that since the derivative of a constant is zero, there are many antiderivatives: $(F(x) + C)' = f(x)$
  - Therefore, we can know the antiderivative only up to an arbitrary additive constant

- If we do definite integrals, the $+\, C$ does not apply – we know the area exactly
- If we do indefinite integrals, we must always add the constant

Fundamental Theorem of Calculus

- The indefinite integral of a function is related to its antiderivative and can be reversed via differentiation
- The definite integral of a function can be computed using one of its infinitely many antiderivatives

- Simply, differentiation and integration are inverse functions
- Proof: [Khan Academy]
- Intuition
  - The sum of infinitesimal changes in a quantity over time adds up to the net change in quantity
  - Think about distance and velocity again

$$s = v(t)\Delta t \to s = \sum \frac{\Delta s}{\Delta t}\Delta t$$

$$\Delta t \to 0: \;\; s = \int \frac{ds}{dt} dt$$

Calculus in Many Dimensions

Same thing, a little different notation

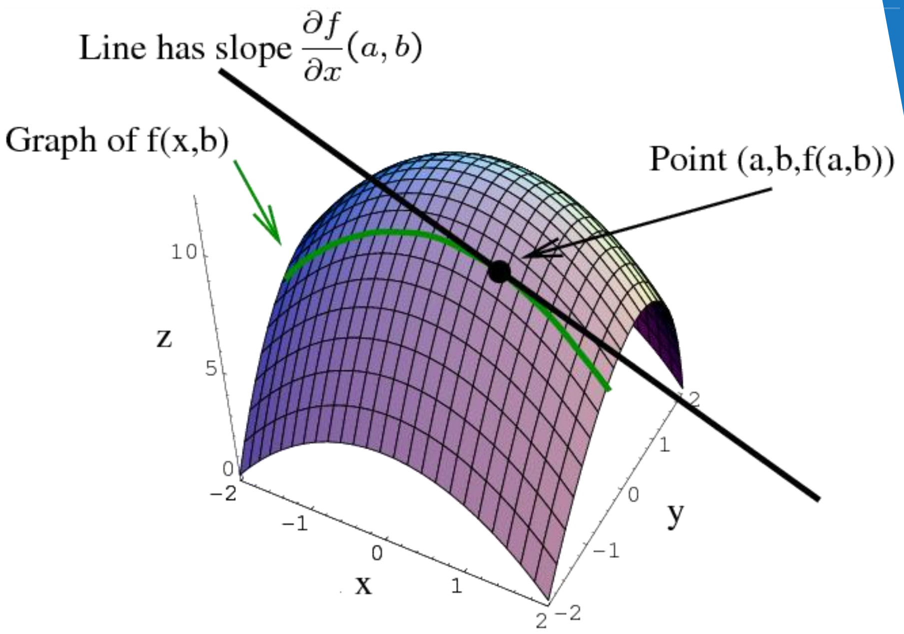
Generalizations

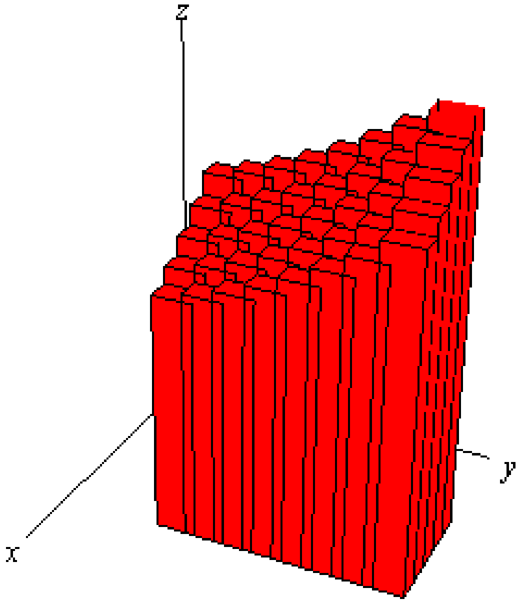
- The notions of derivatives and integrals generalize to more dimensions
  - Derivatives: take the derivative w.r.t. one variable, treat the other variables as "parameters" $\rightarrow$ **partial derivatives**

$$\frac{\partial f(x,y)}{\partial(x)} = g(y)$$

    - Yet more confusing notation: ∂ is the same as d, it's just used for many dimensions
  - Integrals: 1D intervals $[a;\ b]$ can become curves or planes
    - Apply the same "zooming in" technique

$$\iint_R f(x,y)dxdy, \text{R: 2D-region}$$





Gradient Descent

- Optimization method
  - Used for finding local extrema
- Gradient: $\text{grad}(f)$ or ∇f
  - A combination of vector and derivative: $\text{grad}(f(x,y)) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \end{bmatrix}$
    - "Multi-dimensional derivative"
    - A vector whose components are the partial derivatives w.r.t. every variable
  - Shows where the **steepest rise in slope** is
- If we follow the gradient, we'll arrive at a maximum
  - Conversely, negative gradient takes us to a minimum
- Iterative procedure
  - Continue to apply until close enough
- Not guaranteed to find global extrema
  - May get "stuck" in a local extremum

Example: Gradient Descent

- Find a local minimum of the function $f(x) = x^4 - 3x^3 + 2$
  - Start at $x = 6$

```
x_old = 0
x_new = 6
step_size = 0.01
precision = 0.00001

def df(x):
  # f'(x^4 - 3x^3 + 2) = 4x^3 - 9x^2
  y = 4 * x ** 3 - 9 * x ** 2
  return y

while abs(x_new - x_old) > precision:
  x_old = x_new
  x_new += -step_size * df(x_old)

print("The local minimum occurs at ", x_new)
```

Summary

- Limits
- Derivatives
- Integrals
- Calculus in many dimensions
- Gradient descent

Questions?