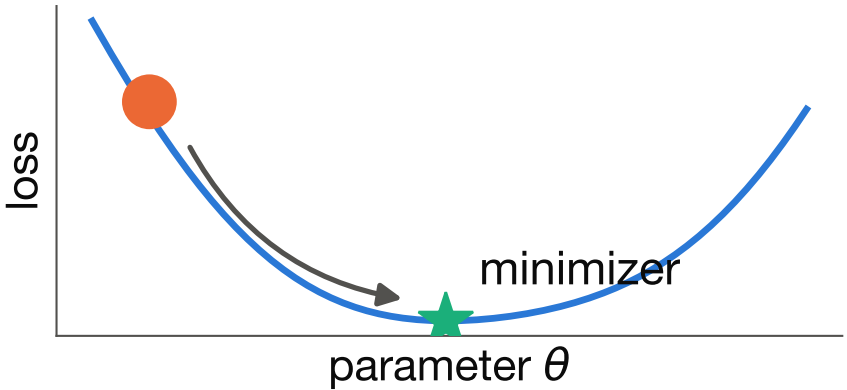
Calculus you need

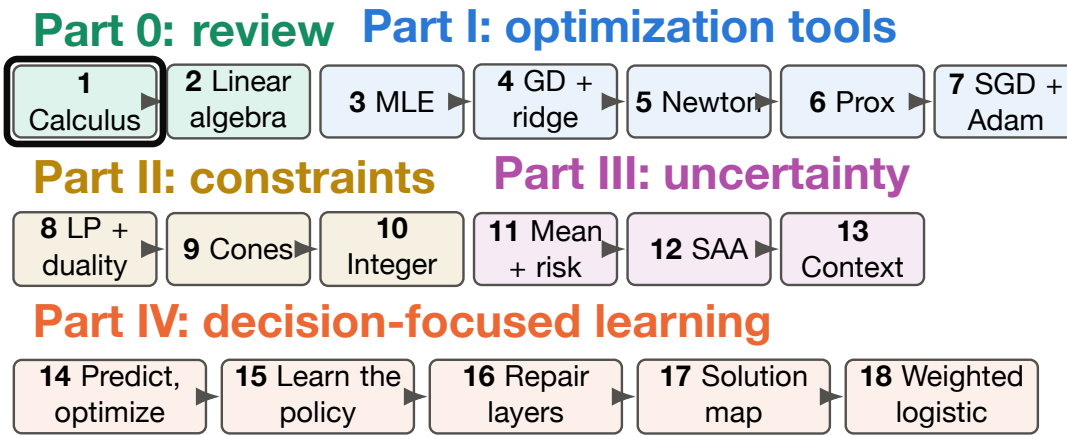
208424 Lecture 1

Optimization = find the minimum

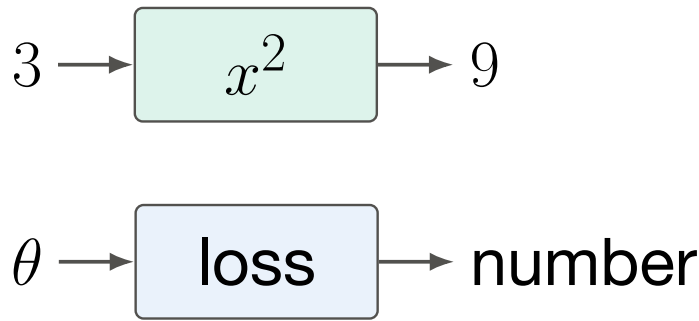


the derivative gives the descent direction

Course map

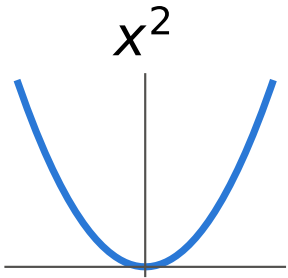

Part 0: review
Part I: optimization tools
1 Calculus
2 Linear algebra
3 MLE
4 GD + ridge
5 Newton
6 Prox
7 SGD + Adam
Part II: constraints
Part III: uncertainty
8 LP + duality
9 Cones
10 Integer
11 Mean + risk
12 SAA
13 Context
Part IV: decision-focused learning
14 Predict, optimize
15 Learn the policy
16 Repair layers
17 Solution map
18 Weighted logistic

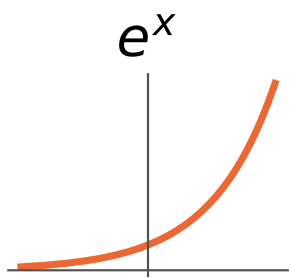
A function: input, rule, output

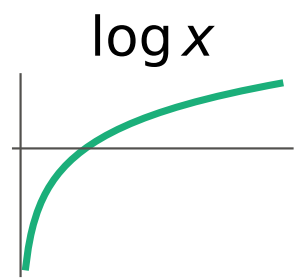


input $\rightarrow$ rule $\rightarrow$ output

Three functions we use all the time

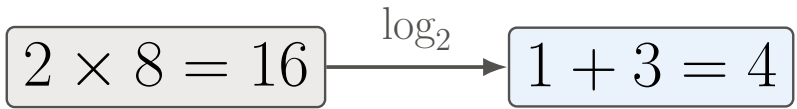






$e^x > 0$ always $\log x$ only for $x > 0$

$\log$ **turns** $\times$ **into** $+$



$$\log(ab) = \log a + \log b \qquad \log e^x = x$$

$\log$ with no base: base $e \approx 2.718$, the natural log

$\sum$ **means** *add them up*

$$\boxed{2} + \boxed{0} + \boxed{3} + \boxed{1} = 6$$

$$y_1 \quad y_2 \quad y_3 \quad y_4$$

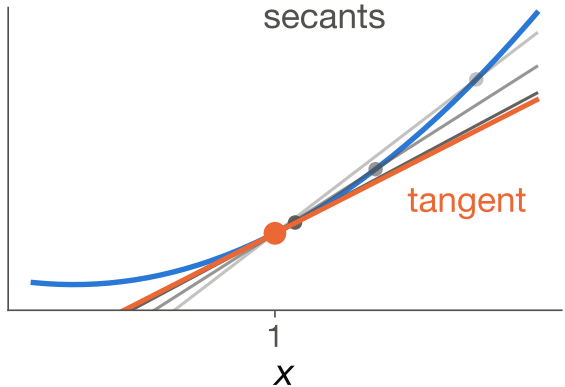
$$\sum_{i=1}^{4} y_i = y_1 + y_2 + y_3 + y_4$$

Slope of a line



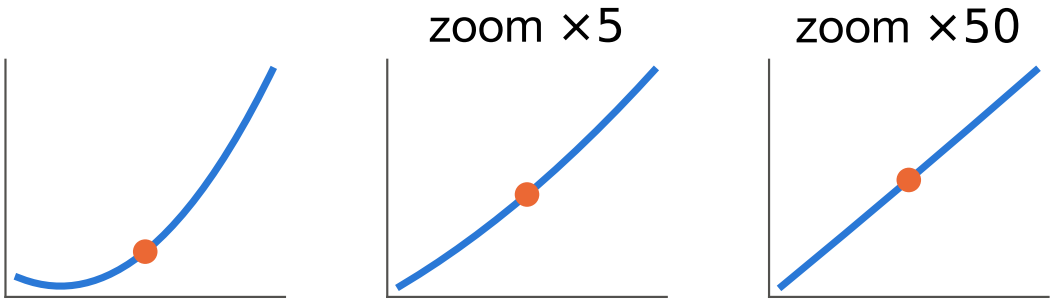
$$\text{slope} = \frac{\text{rise}}{\text{run}} = \frac{2}{1} = 2$$

Secant lines approach the tangent



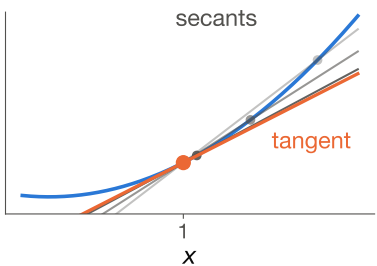
tangent = the line that the secants approach

Zoom in: a smooth curve looks straight



the slope of that straight piece is the derivative

Slope with a tiny run

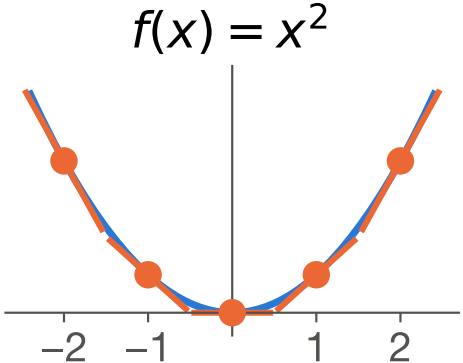


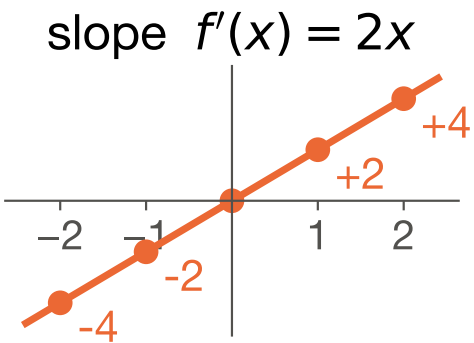
run h	$\dfrac{(1+h)^2 - 1^2}{h}$
1	3
0.1	2.1
0.01	2.01
0.001	2.001
$\to 0$	$\to 2$

$$f'(x) \approx \big(f(x+h) - f(x)\big)/h \quad \text{for small } h$$

why: $\big((1+h)^2 - 1\big)/h = 2 + h$, close to 2 for small h

The slope at every point



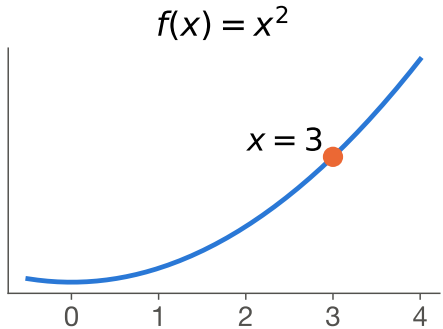


$$f(x) = x^2 \implies f'(x) = 2x$$

Summary: the derivative is a slope

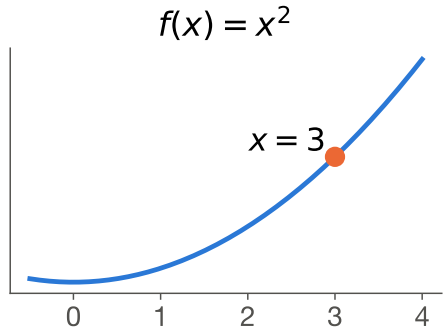
- Let f be a function, x a point, h a small run.
- We want: $f'(x)$, the slope of the tangent line at x.
- Rise over run: $f'(x) \approx \big(f(x+h) - f(x)\big)/h$ for small h.
- $f(x) = x^2$ at $x = 1$: $h = 0.1,\ 0.01,\ 0.001$ give $2.1,\ 2.01,\ 2.001$.
- At every x: $f(x) = x^2 \implies f'(x) = 2x$.

Checkpoint 1



slope at $x = 3$?

Checkpoint 1



slope at $x = 3$?

$$f'(3) = 2 \cdot 3 = 6$$

slope > 0: a step to the left goes down

Derivative rules

$$\big(x^n\big)' = n\,x^{n-1}$$

$$(x^3)' = 3x^2$$

$$\big(e^x\big)' = e^x$$

slope $= e^x$

$$\big(\log x\big)' = \frac{1}{x}$$

$$x > 0$$

$$\big(c\,f\big)' = c\,f'$$

$$(5x^2)' = 10x$$

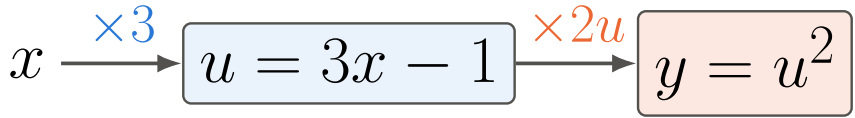
$$\big(f+g\big)' = f' + g'$$

$$(x^2 + x)' = 2x + 1$$

$$\big(c\big)' = 0$$

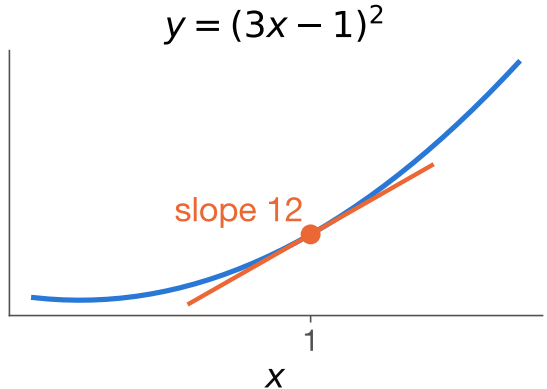
flat line

Chain rule: multiply the derivatives



$$\frac{dy}{dx} = 2u \cdot 3 = 6(3x - 1)$$

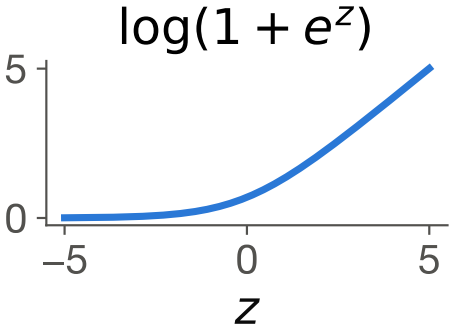
Check with numbers

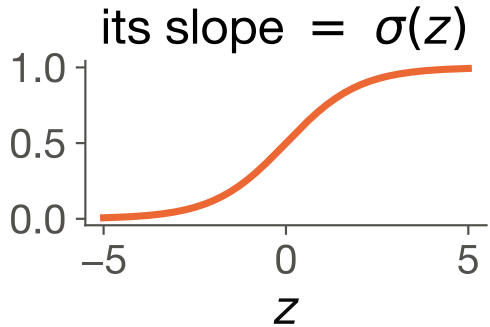


$$\frac{y(1.001) - y(1)}{0.001} = 12.009 \approx 6(3 \cdot 1 - 1) = 12$$

do not forget the $\times 3$

A slope you will meet in Lecture 3



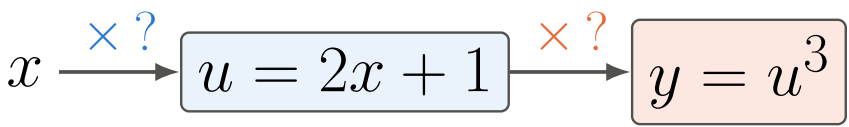


$$\frac{d}{dz}\log\left(1+e^{z}\right)=\frac{e^{z}}{1+e^{z}}=\sigma(z),$$ the sigmoid

Summary: the chain rule

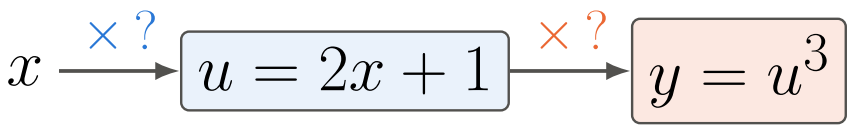
- Let $x \to u \to y$ be two functions: $u = 3x - 1,\ y = u^2$.
- We want: dy/dx, how fast y changes when x changes.
- Each step has a derivative from the rules: $\times 3$ and $\times 2u$.
- Chain rule: multiply them: $dy/dx = 2u \cdot 3 = 6(3x - 1)$.
- At $x = 1$: $dy/dx = 12$. Forget the inner $\times 3$ and you get 4: wrong.

Checkpoint 2



$$\frac{dy}{dx} = ?$$

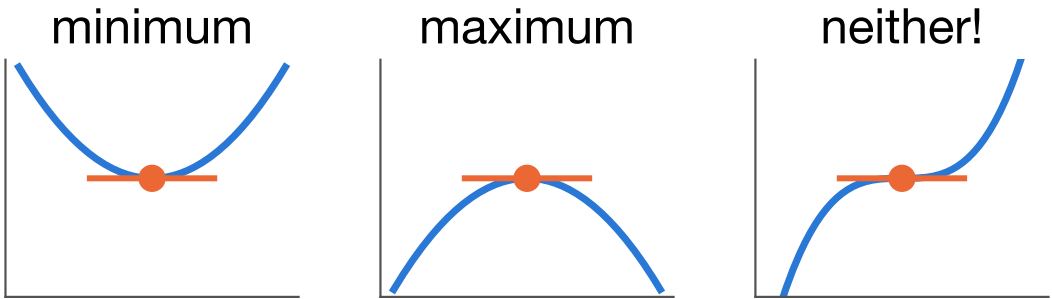
Checkpoint 2



$$\frac{dy}{dx} = ?$$

$$3u^2 \cdot 2 = 6(2x + 1)^2$$

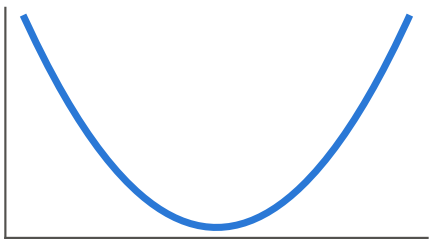
At a minimum, the slope is zero



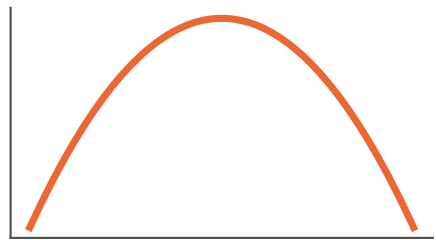
$f'(x) = 0$: stationary point, not always a minimum

Second derivative: min or max

$f'' > 0$: smile, minimum

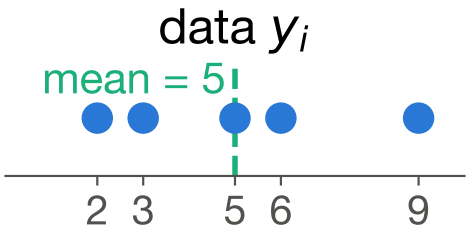


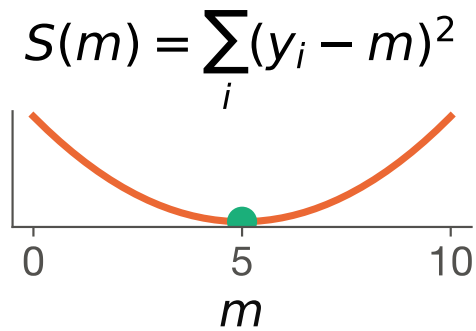
$f'' < 0$: frown, maximum



f'' = the slope of the slope · x^2: $f'' = 2 > 0$

Statistics example: the mean





$S(m) = \sum_i (y_i - m)^2$: squared distances from m

The mean by hand

data $2, 3, 5, 6, 9$

$$S(m) = (2-m)^2 + (3-m)^2 + (5-m)^2 + (6-m)^2 + (9-m)^2$$

$$S'(m) = -2\big[(2-m) + (3-m) + \cdots + (9-m)\big] = -2(25 - 5m)$$

chain rule: the slope of $y_i - m$ in m is -1

$S'(m) = 0$: $5m = 25$, so $m = 5$, the mean

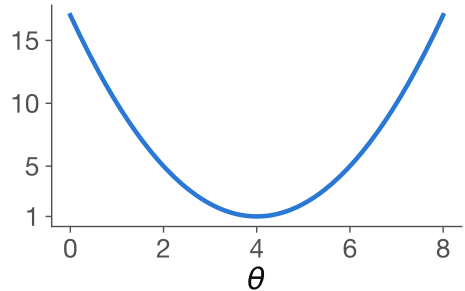
$S''(m) = 10 > 0$: a minimum

any n numbers: $\sum_i y_i = n\,m$, so $m = \bar{y}$, their mean

Summary: finding a minimum

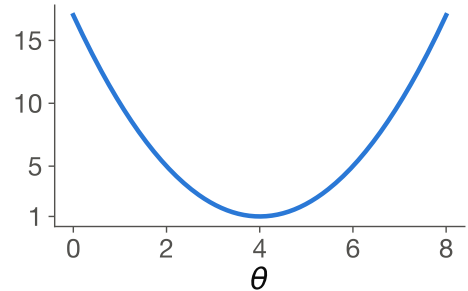
- Let f' be the slope of f, and f'' the slope of f'.
- We want: the minimum, the x where $f(x)$ is smallest.
- Solve $f'(x) = 0$; if $f''(x) > 0$ there (smile), it is a minimum.
- For data y_i and one number m: $S(m) = \sum_i (y_i - m)^2$ is smallest at $m = \bar{y}$, the mean.
- $f'(x) = 0$ also holds at a maximum and at the inflection point of x^3; if $f''(x) = 0$, the test does not decide.

Checkpoint 3



$f(\theta) = (\theta - 4)^2 + 1$. Where is the minimum, and why?

Checkpoint 3

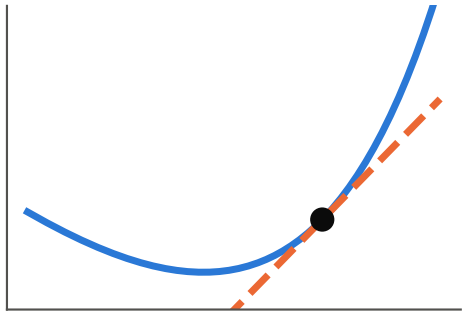


$f(\theta) = (\theta - 4)^2 + 1$. Where is the minimum, and why?

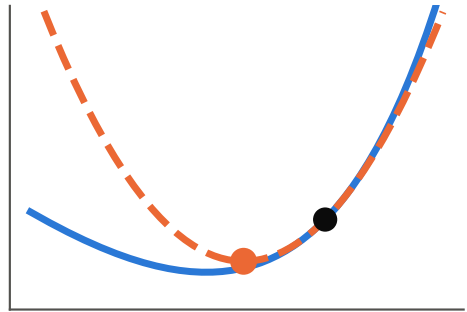
$f'(\theta) = 2(\theta - 4) = 0 \Rightarrow \theta = 4$; $f'' = 2 > 0$: minimum

Taylor approximations

line: which way is down

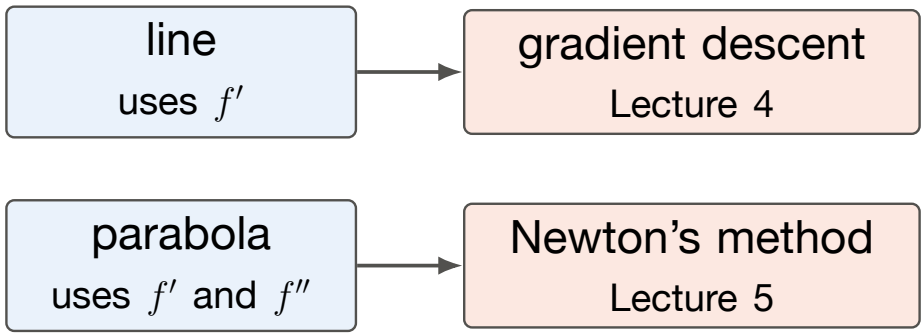


parabola: its minimum

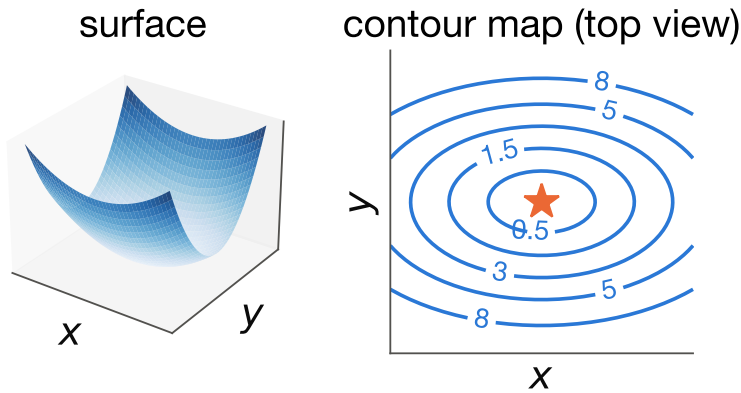


$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2$$

Two pictures, two algorithms



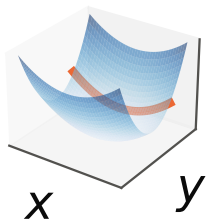
Two inputs: a surface and its map



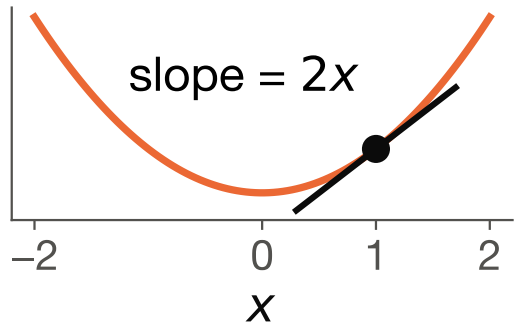
contour line = points with the same height

Partial derivative: slope of a slice

slice at $y = 1$



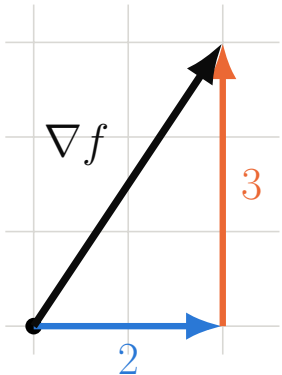
$f(x, 1) = x^2 + 3$



$f(x, y) = x^2 + 3y^2$: $\dfrac{\partial f}{\partial x} = 2x$ (treat y as a number)

in the same way: $\partial f / \partial y = 6y$ (treat x as a number)

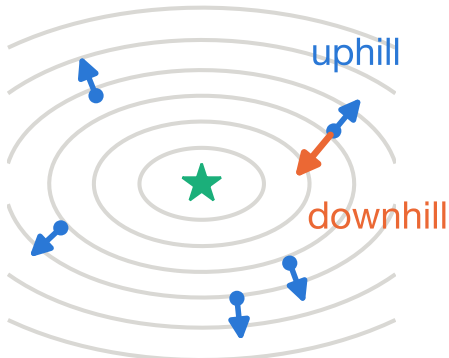
Gradient = the list of partial derivatives



$$\nabla f = \left(\frac{\partial f}{\partial x},\ \frac{\partial f}{\partial y}\right) = (2x,\ 6y)$$

at $(1,\ 0.5)$: $\nabla f = (2,\ 3)$

The gradient points uphill

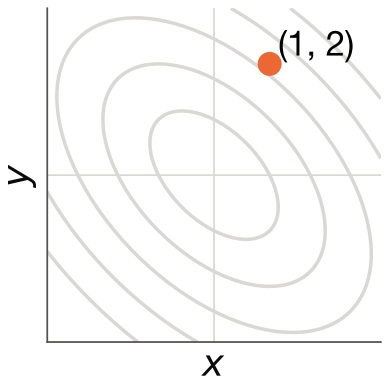


$-\nabla f$ points downhill at the minimum: $\nabla f = (0, 0)$

Summary: the gradient

- Let $f(x, y)$ have two inputs, for example $f(x, y) = x^2 + 3y^2$.
- Let $\partial f/\partial x$, $\partial f/\partial y$ be the slopes in x and in y, the other input fixed.
- We want: the descent direction, and the minimum.
- Gradient: $\nabla f = (\partial f/\partial x,\ \partial f/\partial y) = (2x,\ 6y)$.
- $-\nabla f$ points downhill; at the minimum, $\nabla f = (0, 0)$.

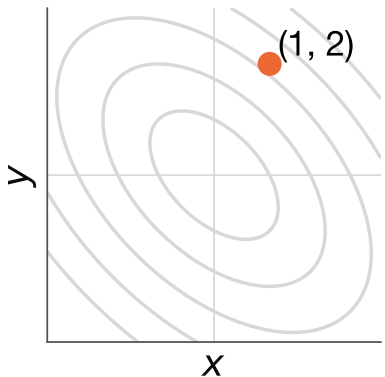
Checkpoint 4



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

$$\nabla f(1, 2) = ?$$

Checkpoint 4



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

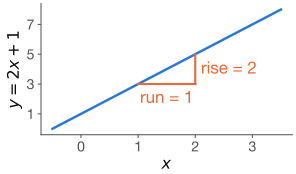
$$\nabla f(1, 2) = ?$$

$$\partial f / \partial x = 2x + y = 4$$

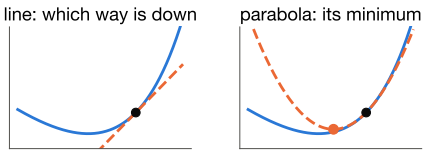
$$\partial f / \partial y = 2y + x = 5$$

$$\nabla f(1, 2) = (4,\ 5)$$

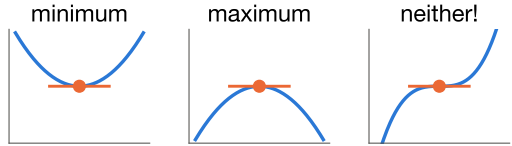
Lecture summary



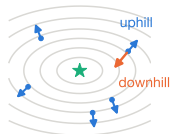
slope = derivative



linear and quadratic approximations

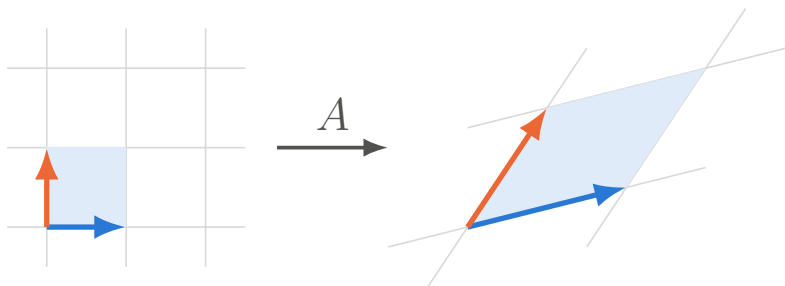


$f' = 0,\ f'' > 0$: a minimum



gradient points uphill

Next lecture: linear algebra



vectors · matrices · $X^\top X$ · the Hessian

Exercise

Find the derivative:

1. $5x^3$
2. e^{2x} (chain rule)
3. $\log(1 + x^2)$ (chain rule)

Find the gradient at $(1, 1)$:

4. $f(x, y) = x^2y + y^2$

Find the minimum, and check it:

5. $f(\theta) = (\theta - 2)^2 + 3$