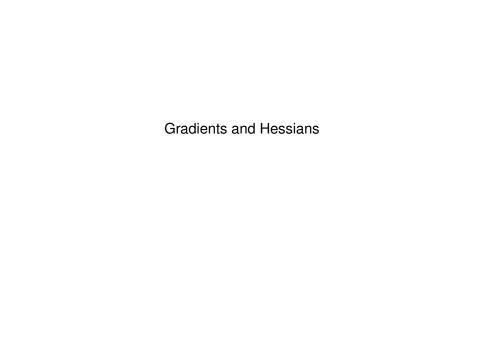
# Differentiation and Convexity

**Andrew Nobel** 

February, 2020



## Gradients and Hessians

**Definition:** Let  $f: \mathbb{R}^d \to \mathbb{R}$  be nice. The *gradient* of f is the vector of partial derivatives

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_d}(\mathbf{x})\right)^t \in \mathbb{R}^d$$

The *Hessian* of *f* is the matrix of second partial derivatives

$$H(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) : 1 \le i, j, \le d \right] \in \mathbb{R}^{d \times d}$$

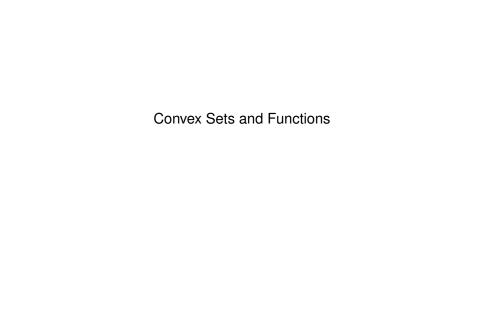
**Note:** Under mild conditions  $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ , so  $\nabla^2 f$  is symmetric

## Examples

**Ex 1:** Linear 
$$f(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle + b$$
.

- $\triangleright \nabla f(\mathbf{x}) = \mathbf{u}$
- $\nabla^2 f(\mathbf{x}) = \mathbf{0}$

## **Ex 2:** Quadratic $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$ with $\mathbf{A} \in \mathbb{R}^{d \times d}$ symmetric



### Convex Sets

**Definition:** A set  $C \subseteq \mathbb{R}^d$  is *convex* if for every  $x,y \in C$  and every  $\alpha \in [0,1]$  the point  $\alpha x + (1-\alpha)y \in C$ .

Interpretation: the line between any two points in  $\mathcal{C}$  is contained in  $\mathcal{C}$ 

### **Examples**

- ▶ Case d = 1: C = (a, b) finite or infinite interval
- ▶ Case d > 1:  $C = (a_1, b_1) \times \cdots \times (a_d, b_d)$  Cartesian product of d intervals
- $ightharpoonup B_r := \{x: ||x|| < r\}$  open ball of radius r centered at the origin
- $ightharpoonup C = \{x : w^t x b \ge 0\}$  half-space with direction w, offset b
- $ightharpoonup C = \{x: w^t x = b\}$  hyperplane, (n-1)-dimensional

**Fact:** If  $C_1, \ldots, C_n$  are convex sets then so is their intersection  $\bigcap_{i=1}^n C_i$ .

### **Convex Functions**

**Definition:** Let  $C\subseteq\mathbb{R}^d$  be convex. A function  $f:C\to\mathbb{R}$  is *convex* if for every  $x,y\in C$  and every  $\alpha\in(0,1)$ 

$$f(\alpha x + (1 - \alpha)y) \le \alpha f(x) + (1 - \alpha)f(y) \tag{0.1}$$

*Interpretation:* Line connecting (x, f(x)) and (y, f(y)) lies above graph of f

#### **Definition**

- ▶  $f: C \to \mathbb{R}$  is *concave* if (0.1) holds with  $\leq$  replaced by  $\geq$
- ▶  $f: C \to \mathbb{R}$  is *strictly convex* if (0.1) holds with  $\leq$  replaced by <

**Note:** f is convex if -f is concave and v.v.

## Examples of Convex/Concave Functions

#### Case d=1

- f(x) = |x| is convex, but *not* strictly convex
- lacksquare  $f(x)=x^2,\,e^x,\,e^{-x},\,x^{-1},\,{\rm and}\,\,x\log x$  are strictly convex
- $f(x) = \log x$ ,  $\sqrt{x}$  are strictly concave

#### Case $d \ge 2$

- f(x) = ||x|| is convex
- affine function  $f(x) = \langle x, u \rangle + b$  is convex and concave
- ▶ if  $A \subseteq \mathbb{R}^d$  is finite then  $f(x) = \min_{u \in A} \langle x, u \rangle$  is concave
- quadratic form  $f(x) = x^t A x$  is convex if  $A \ge 0$ , concave if  $A \le 0$

## **Properties of Convex Functions**

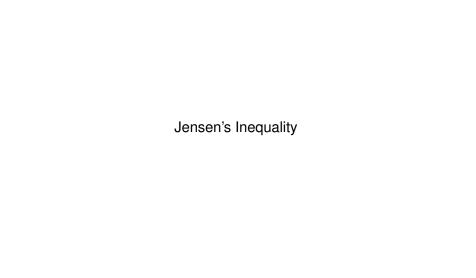
Fact: If  $f_1,\ldots,f_n:C o\mathbb{R}$  are convex so is  $f=\max\{f_1,\ldots,f_n\}$ 

## Checking convexity: Case d=1

- $f:(a,b)\to\mathbb{R}$  is convex if  $f''(x)\geq 0$  for all  $x\in(a,b)$
- $\blacktriangleright \ f:(a,b)\to \mathbb{R} \text{ is concave if } f''(x)\le 0 \text{ for all } x\in (a,b)$

### Checking convexity: Case d > 1

- ▶  $f: C \to \mathbb{R}$  is convex if  $\nabla^2 f(x)$  is non-negative definite for each  $x \in C$
- ▶  $f: C \to \mathbb{R}$  is concave if  $-\nabla^2 f(x)$  is non-negative definite for each  $x \in C$



# Jensen's Inequality in 1-Dimension

**Theorem:** Let  $X \in (a, b)$  be a random variable

- (1) The expected value  $\mathbb{E}X \in (a,c)$
- (2) If  $f:(a,b)\to\mathbb{R}$  is convex then  $f(\mathbb{E}X)\leq\mathbb{E}f(X)$ .
- (3) If  $f:(a,b)\to\mathbb{R}$  is concave then  $f(\mathbb{E}X)\geq\mathbb{E}f(X)$ .

# First Applications of Jensen's Inequality

#### **Fact**

**AM-GM inequality:** If  $a_1,\ldots,a_n>0$  then  $\left(\prod_{i=1}^n a_i\right)^{1/n}\leq \frac{1}{n}\sum_{i=1}^n a_i$ 

**Cauchy-Schwartz:** If X and Y are r.v. then  $\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2} \, \mathbb{E}Y^2$