# Convex functions, Part I

Lecture 13

November 2, 2021

Beck, sections 7.1-7.3

# Convex and strictly convex functions

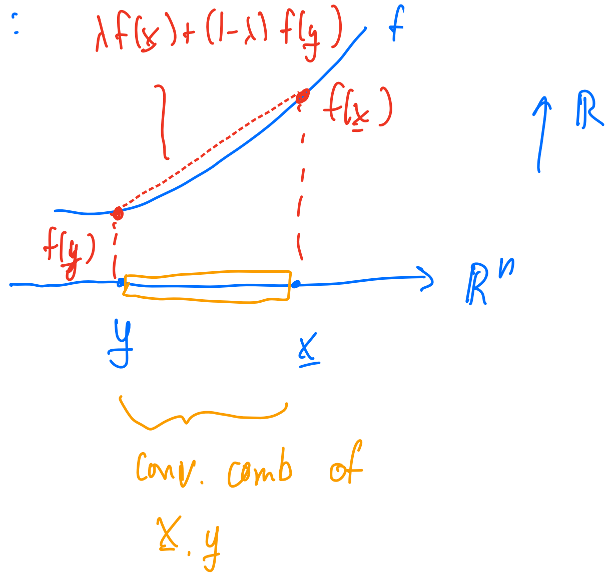
Definition: Let $f : \mathbb{R}^n \to \mathbb{R}$. $f$ is convex (on a set $S$) if

$\forall\ \underline{x}, \underline{y} \in S$, $\lambda \in (0,1)$, then

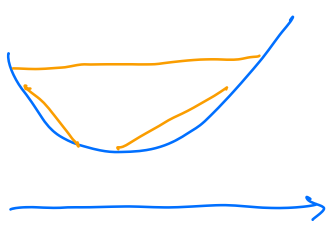
$$\underbrace{f(\lambda \underline{x} + (1-\lambda)\underline{y})}_{\text{convex comb. of } \underline{x}, \underline{y}.} \leq \underbrace{\lambda f(\underline{x}) + (1-\lambda) f(\underline{y})}_{\text{convex comb. of } f(\underline{x}) \text{ and } f(\underline{y}).}$$

This definition is geometric:

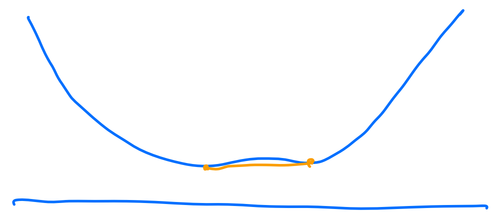
Defn of convexity: graph of $f$ between $\underline{x}$ and $y$ lies below secant line connecting $(\underline{x}, f(\underline{x}))$, $(y, f(y))$.

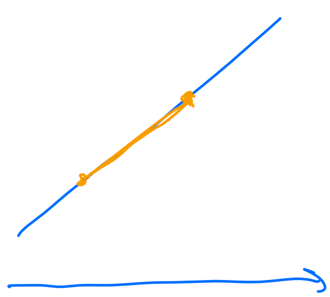



convex



not convex





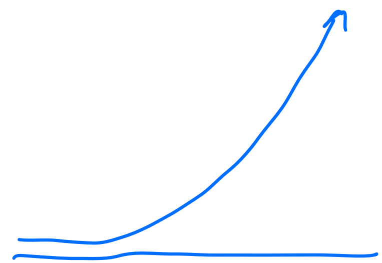
<u>Definition</u>: $f: \mathbb{R}^n \to \mathbb{R}$ is <u>strictly</u> convex if $\forall$ $\underline{x}, y$, $\lambda \in (0,1)$, then

$$f(\lambda \underline{x} + (1-\lambda) y) < \lambda f(\underline{x}) + (1-\lambda) f(y)$$

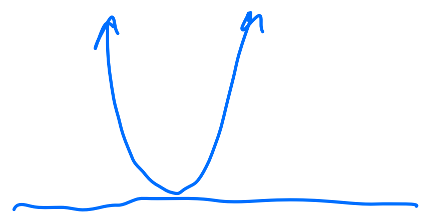
# Examples of convex/concave functions

Examples (visual) $f: \mathbb{R} \to \mathbb{R}$

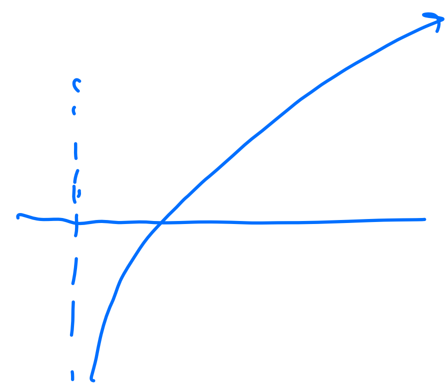
$f(x) = \exp(x)$ is convex



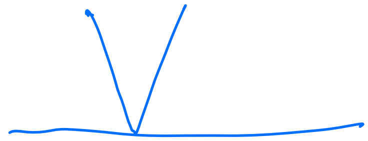
$f(x) = x^2$ is convex



$f(x) = \log x$ is not convex

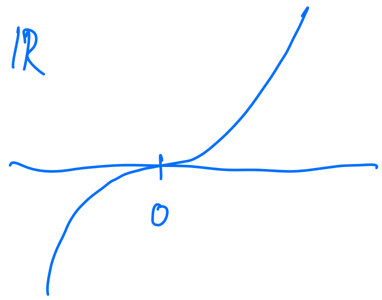


$f(x) = |x|$ is convex

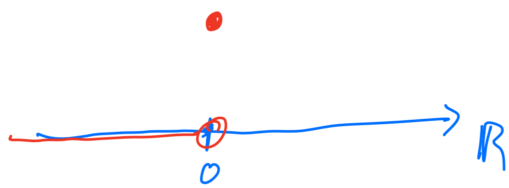


$f(x) = x^3$ is not convex on $\mathbb{R}$

$f$ is convex on $[0, \infty)$



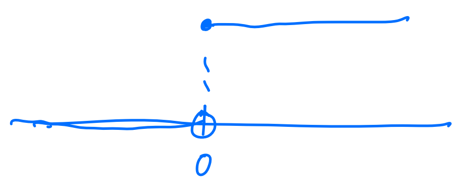

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x = 0 \end{cases}$$




$f$ <u>is</u> convex.

$f(x) = H(x)$ (Heaviside function)

$$= \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$




$f$ is not convex on $\mathbb{R}$.

---

<u>Example</u>: $f(\underline{x}) = \underline{a}^T \underline{x} + b$, $\underline{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$

("Affine" function)

$f$ is convex: let $\underline{x}, \underline{y} \in \mathbb{R}^n$, $\lambda \in (0,1)$.

$$f(\lambda \underline{x} + (1-\lambda)\underline{y}) = \underline{a}^T(\lambda \underline{x} + (1-\lambda)\underline{y}) + b$$

$$= \lambda \underline{a}^T \underline{x} + \lambda b + (1-\lambda)\underline{a}^T \underline{y} + (1-\lambda) b.$$

$$= \lambda(\underline{a}^T \underline{x} + b) + (1-\lambda)(\underline{a}^T \underline{y} + b)$$

$$= \lambda f(\underline{x}) + (1-\lambda) f(\underline{y})$$

i.e. $f(\lambda \underline{x} + (1-\lambda)\underline{y}) \leq \lambda f(\underline{x}) + (1-\lambda) f(\underline{y})$

$\Longrightarrow$ $f$ is convex.

---

<u>Definition</u>: $f$ is <u>concave</u> if $-f$ is convex.

---

<u>Example</u>: $f: \mathbb{R}^n \to \mathbb{R}$

$f$ is a <u>norm</u>

$f$ is convex: $\lambda \in (0,1)$, $\underline{x}, \underline{y} \in \mathbb{R}^n$

$f(\lambda \underline{x} + (1-\lambda) \underline{y})$

$$\leq f(\lambda \underline{x}) + \cancel{(1-\lambda) f(\underline{y})} \; f((1-\lambda)\,\underline{y})$$

triangle inequality

$$\leq |\lambda|\, f(\underline{x}) + |1-\lambda|\, f(\underline{y})$$

positive homogeneity

$$= \lambda f(\underline{x}) + (1-\lambda) f(\underline{y}) \quad \checkmark$$

# Jensen's inequality



We know: if $f$ is convex, then

$$f(\lambda \underline{x} + (1-\lambda)\underline{y}) \le \lambda f(\underline{x}) + (1-\lambda) f(\underline{y})$$

a convex set

Theorem: Let $f$ be convex on $C$. Then for any

$\underline{x}_1 \ldots \underline{x}_K \in C$, and $\underline{\lambda} \in \Delta_K$, we have

$$f\left(\sum_{i=1}^{K} \lambda_i \underline{x}_i\right) \le \sum_{i=1}^{K} \lambda_i f(\underline{x}_i) \qquad \text{(Jensen's inequality)}$$

Proof: (Induction) $K=1$, $K=2$ straightforward

Assume $k > 2$, assume inductive hypothesis:

$$f\left(\sum_{i=1}^{k-1} \lambda_i \underline{x}_i\right) \leq \sum_{i=1}^{k-1} \lambda_i f(\underline{x}_i) \qquad \underline{\lambda} \in \Delta_{k-1},\ \underline{x}, \ldots \underline{x}_{k-1} \in C$$

Let $\underline{x}_1, \ldots \underline{x}_k \in C$, $\underline{\lambda} \in \Delta_k$.

$$f\left(\sum_{i=1}^{k} \lambda_i \underline{x}_k\right) = f\left((1-\lambda_k) \sum_{i=1}^{k-1} \frac{\lambda_i}{1-\lambda_k} \underline{x}_k + \lambda_k \underline{x}_k\right)$$

Note: $\mu_i = \dfrac{\lambda_i}{1-\lambda_k}$, $i = 1 \ldots k-1$ satisfies $\underline{\mu} \in \Delta_{k-1}$

$$f\Big((1-\lambda_k) \underbrace{\sum_{i=1}^{k-1} \mu_i \underline{x}_i}_{\underline{v} \in C} + \lambda_k \underline{x}_k\Big)$$

f convex

$$\leq (1-\lambda_k) f(\underline{v}) + \lambda_k f(\underline{x}_k)$$

$$\leq (1-\lambda_k) f\left(\sum_{i=1}^{k-1} \mu_i \underline{x}_i\right) + \lambda_k f(\underline{x}_k)$$

inductive hyp.

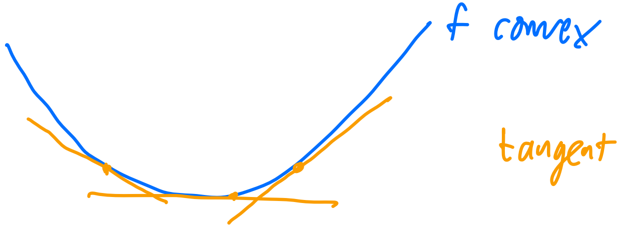
$$\leq (1-\lambda_k) \sum_{i=1}^{k-1} \mu_i f(\underline{x}_k) + \lambda_k f(\underline{x}_k)$$

$$= (1-\lambda_k) \sum_{i=1}^{k-1} \frac{\lambda_i}{1-\lambda_k} f(\underline{x}_k) + \lambda_k f(\underline{x}_k)$$

$$= \sum_{i=1}^{k} \lambda_i f(\underline{x}_i) \quad ✓ \quad \blacksquare$$

# Gradients and convex functions

Geometric characterization of convex functions




f lying "above" its tangent lines $\Longleftrightarrow$ convexity.

<u>Theorem</u>: ("Gradient inequality") Let $f$ be $C^1$ on a convex set $C \subset \mathbb{R}^n$. Then $f$ is convex on $C$ iff $\forall\ \underline{x}, \underline{y} \in C$,

$$f(\underline{y}) - f(\underline{x}) \geq \nabla f(\underline{x})^T (\underline{y} - \underline{x})$$

Note: tangent plane at $x$ is $f(\underline{x}) + \nabla f(\underline{x})^T(\underline{y} - \underline{x})$

so gradient inequality:

$$f(\underline{y}) \geq f(\underline{x}) + \nabla f(\underline{x})^T(\underline{y} - \underline{x})$$

value of f

value of tangent plane.

# Stationary points of convex functions



Gradient inequality immediately gives a connection to optimization.

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$, that is convex on $C \subset \mathbb{R}^n$. Assume $f$ is $C^1$.

Then if $\underline{x} \in C$ satisfies $\nabla f(\underline{x}) = \underline{0}$ ($\underline{x}$ is a stationary point), then $\underline{x}$ is a global minimizer of $f$ on $C$.

Proof: Let $y \in C$.

Gradient inequality: $f(\underline{y}) - f(\underline{x}) \geq \underbrace{\nabla f(\underline{x})}_{0}{}^T (\underline{y} - \underline{x})$

$\Rightarrow f(\underline{y}) \geq f(\underline{x})$. ($\underline{y} \in C$) $\blacksquare$

Thurs. Nov. 4

HW #5 posted

Office hours tomorrow (Nov 5) moved to 12pm-1pm.

---

Recall: the gradient inequality is equivalent to convexity.

$$f \text{ convex} \iff f(\underline{y}) - f(\underline{x}) \geq \nabla f(\underline{x})^T (\underline{y} - \underline{x}) \quad \forall\, \underline{x}, \underline{y}.$$

Also:

$$f \text{ strictly convex} \iff f(\underline{y}) - f(\underline{x}) > \nabla f(\underline{x})^T (\underline{y} - \underline{x}) \quad \forall\, \underline{x} \neq \underline{y}$$

# Quadratic functions with definite Hessians

Recall: $f$ is quadratic if

$$f(\underline{x}) = \underline{x}^T \underline{\underline{A}}\, \underline{x} + 2\underline{b}^T \underline{x} + c, \qquad (1)$$

$$\text{for } \underline{\underline{A}} \in \mathbb{R}^{n\times n}, \quad \underline{b} \in \mathbb{R}^n, \quad c \in \mathbb{R}$$

(We can always choose $\underline{\underline{A}}$ to be symmetric)

<u>Theorem</u>: Let $f$ be quadratic of the form (1). Then

(i) $f$ is convex iff $\underline{\underline{A}} \succeq \underline{\underline{0}}$

(ii) $f$ is strictly convex iff $\underline{\underline{A}} \succ \underline{\underline{0}}$

Proof: (i) $f$ convex $\iff$ gradient inequality.

gradient inequality: $f(\mathbf{y}) - f(\mathbf{x}) \geq \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x})$

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c$$

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + 2\mathbf{b}$$

grad. inequality:

$$\mathbf{y}^T\mathbf{A}\mathbf{y} + 2\mathbf{b}^T\mathbf{y} - \mathbf{x}^T\mathbf{A}\mathbf{x} - 2\mathbf{b}^T\mathbf{x} \geq (2\mathbf{A}\mathbf{x} + \overset{2}{\vee}\mathbf{b})^T(\mathbf{y}-\mathbf{x})$$

$$\mathbf{y}^T\mathbf{A}\mathbf{y} - \mathbf{x}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}\mathbf{y} + 2\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$$

$$\mathbf{y}^T\mathbf{A}\mathbf{y} - 2\mathbf{x}^T\mathbf{A}\mathbf{y} + \mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$$

$$(\mathbf{y}-\mathbf{x})^T\mathbf{A}(\mathbf{y}-\mathbf{x}) \geq 0 \quad \forall\, \mathbf{y}, \mathbf{x}$$

$$\Updownarrow \text{ def } \mathbf{z} = \mathbf{y} - \mathbf{x}$$

$$\mathbf{z}^T\mathbf{A}\mathbf{z} \geq 0 \quad \forall\, \mathbf{z}$$

$$\Updownarrow$$

$$\mathbf{A} \succeq \mathbf{0} \qquad \blacksquare$$

# Convex functions and their Hessians



Theorem: Suppose $f \in C^2$ on a convex, open set $C \subseteq \mathbb{R}^n$.

Then $f$ is convex on $C$ iff $\nabla^2 f(\underline{x}) \succeq \underline{\underline{0}}$ $\forall\, \underline{x} \in C$.

Proof (half) We'll show $\nabla^2 f(\underline{x}) \succeq \underline{\underline{0}} \Longrightarrow$ convexity.

Let $\underline{x}, y \in C$. By Taylor's theorem:

$$f(y) = f(\underline{x}) + \nabla f(\underline{x})^T (y - \underline{x}) + \frac{1}{2}(y - \underline{x})^T \nabla^2 f(\underline{z})(y - \underline{x})$$

for some $\underline{z} \in [\underline{x}, y]$

$$\nabla^2 f \succeq \underline{\underline{0}} \Longrightarrow \frac{1}{2}(y - \underline{x})^T \nabla^2 f(\underline{z})(y - \underline{x}) \geq 0$$

$\Longrightarrow f(\underline{y}) \geq f(\underline{x}) + \nabla f(\underline{x})^T(\underline{y}-\underline{x}) \longrightarrow$ gradient inequality

$\Longrightarrow f$ is convex. ■

Example: $f(x,y) = y^2/x$ ("quadratic over linear" function)

$y \in \mathbb{R},\ x > 0.$

$$\nabla f = \begin{pmatrix} -y^2/x^2 \\ 2y/x \end{pmatrix} \qquad \nabla^2 f = \begin{pmatrix} 2y^2/x^3 & -2y/x^2 \\ -2y/x^2 & 2/x \end{pmatrix} = 2\begin{pmatrix} y^2/x^3 & -y/x^2 \\ -y/x^2 & 1/x \end{pmatrix}$$

$$\left.\begin{aligned} \det \nabla^2 f &= 4\left[\frac{y^2}{x^4} - \frac{y^2}{x^4}\right] = 0 \\ \operatorname{trace}(\nabla^2 f) &= 2\left(\frac{y^2}{x^3} + \frac{1}{x}\right) > 0 \end{aligned}\right\} \nabla^2 f \succeq \underline{\underline{0}} \Rightarrow f \text{ convex.}$$

---

Definiteness of the Hessian results in other properties:

Proposition: Assume $f: \mathbb{R} \to \mathbb{R}$ is $C^2$ on a convex set $C$.

$f$ is convex iff $f'(x)$ is nondecreasing on $C$.

[This is true in $\mathbb{R}^n$, and for $f$ only $C^1$.]

$$(\nabla f(\underline{x}) - \nabla f(\underline{y}))^T(\underline{x}-\underline{y}) \geq 0$$

# Strictly convex functions and their Hessians

Note: Strictly convex functions need not have (strictly) definite Hessians: $f(x) = x^4$

$$f''(x) = 12x^2 \geq 0$$

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$, assume $C \subseteq \mathbb{R}^n$ is convex. Then:

$$\nabla^2 f(\underline{x}) \succ \underline{\underline{0}} \quad \forall\, \underline{x} \in C \Longrightarrow f \text{ is strictly convex on } C.$$

(The reverse is not true.)

Q: Suppose $f : C \to \mathbb{R}$, $C \subseteq \mathbb{R}^n$ is convex, $f \in C^1$.

Assume $f$ is convex.

Suppose $\underline{x}$ is a global minimizer for $f$.

$\overset{?}{\Longrightarrow} \nabla f(\underline{x}) = \underline{0}$

No: $\underline{x}$ could lie on $\partial C = bd(C)$

Proposition: Assume $f : C \to \mathbb{R}$ is $C^1$, assume $C$ is convex. If $C = \mathbb{R}^n$, then $\underline{x}$ is a global minimizer iff $\nabla f(\underline{x}) = \underline{0}$.

Assume $f$ is convex.

---

At the moment, we can only show convexity through:

(i) definition

(ii) gradient inequality

(iii) semidefiniteness of Hessian.

These are all cumbersome to operate with.

Easier strategies to verify convexity utilize knowledge of "convexity-preserving operations."