The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381K: Convex Optimization — Fall 2019

LECTURE 11

Aryan Mokhtari Monday, October 7, 2019.

Goal: In this lecture we study different definitions of convexity for functions.

1 Convex functions

Recall the following definition

Definition 1 (Convexity I)**.** *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if dom(f) is convex and* for all $\mathbf{x}_1, \mathbf{x}_2 \in$ dom$(f) \subseteq \mathbb{R}^n$, $\lambda \in [0, 1]$, *we have:*

$$f\left(\lambda \mathbf{x}_1 + (1-\lambda)\,\mathbf{x}_2\right) \;\leq\; \lambda f(\mathbf{x}_1) + (1-\lambda)\, f(\mathbf{x}_2). \tag{1}$$

We now give the first-order condition for convexity.

Definition 2 (Convexity II)**.** *Suppose a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then it is convex if and only if dom(f) is convex and*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x})\,. \tag{2}$$

Intuitively speaking, equation (2) states that the first-order Taylor approximation is in fact a global underestimator of the function, as illustrated in figure 1.

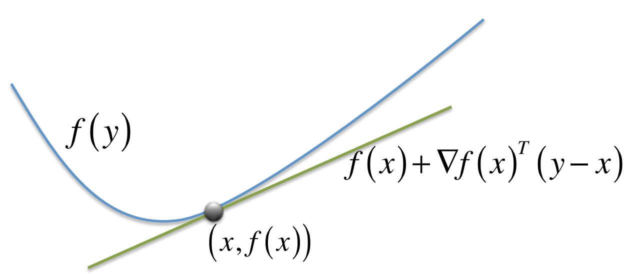

Figure 1: If function f is convex and differentiable, then eq. (2) is a global underestimator of f.

Proposition 1. *For differentiable functions, definition 1 and definition 2 are equivalent.*

Proof. If f is convex, by definition I we know that for any $\lambda \in [0, 1]$ we have

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

Rearranging terms and dividing by λ, we obtain:

$$f(\mathbf{y}) - f(\mathbf{x}) \geq \frac{f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\lambda}.$$

By sending $\lambda \to 0$ we obtain that

$$f(\mathbf{y}) - f(\mathbf{x}) \geq \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}).$$

Conversely, suppose now that function f satisfies the first-order condition. Let $\bar{\mathbf{x}} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ be some point in the convex hull ($\lambda \in (0,1)$). Hence, we have

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^\top (\bar{\mathbf{x}} - \mathbf{x}_1), \\ f(\mathbf{x}_2) &\geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^\top (\bar{\mathbf{x}} - \mathbf{x}_2). \end{aligned}$$

Multiplying the first inequality by λ, the second by $(1-\lambda)$ and adding, we obtain

$$\begin{aligned} \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) &\geq f(\bar{\mathbf{x}}) + \lambda \nabla f(\bar{\mathbf{x}})^\top (\bar{\mathbf{x}} - \mathbf{x}_1) + (1-\lambda) \nabla f(\bar{\mathbf{x}})^\top (\bar{\mathbf{x}} - \mathbf{x}_2) \\ &= f(\bar{\mathbf{x}}) + \lambda \nabla f(\bar{\mathbf{x}})^\top (\bar{\mathbf{x}} - \lambda \mathbf{x}_1 - (1-\lambda)\mathbf{x}_2) \\ &= f(\bar{\mathbf{x}}) \\ &= f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \end{aligned}$$

□

We now give the final definition of convexity.

Definition 3 (Convexity III)**.** *Suppose that a function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Then f is convex iff its Hessian is positive semidefinite:*

$$\bigtriangledown^2 f(\mathbf{x}) \succeq 0, \text{ for all } \mathbf{x} \in dom(f) \tag{3}$$

As an example, consider the function:

$$f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2.$$

Expanding, we have $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} - \mathbf{b}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{A}^\top \mathbf{b} + \|\mathbf{b}\|_2^2$. The Hessian of this is $2\mathbf{A}^\top \mathbf{A}$, which is positive semidefinite, since for any vector $\mathbf{x}$, $\mathbf{x}^\top (\mathbf{A}^\top \mathbf{A})\mathbf{x} = \|\mathbf{A}\mathbf{x}\|_2^2 \geq 0$.

Proposition 2. *The definition give above is equivalent to Definitions 1 and 2. That is, a twice differentiable function f is convex if*

$$\nabla^2 f(\mathbf{x}) \in \mathcal{S}_+^n. \tag{4}$$

Proof. One way to prove this is via Taylor's theorem, using the Lagrange form of the remainder. To prove that a function with positive semidefinite Hessian is convex, using a second order Taylor expansion we have:

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + (\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) \tag{5}$$

for some value of $\alpha \in [0,1]$. Now, since the Hessian is positive semidefinite,

$$(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) \geq 0, \tag{6}$$

which leads to

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}), \tag{7}$$

which is the first-order condition of convexity. Hence, this proves that $f(\mathbf{x})$ is convex. For the reverse direction, showing that convexity implies positive semi definiteness of the Hessian, again we can use Taylor's theorem. However, there is a slightly delicate issue because of the Hessian. For this, we need to use some properties of symmetric matrices, to claim that if for some orthonormal basis $\{\mathbf{x}_i\}$, a matrix $\mathbf{A}$ satisfies $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$, then A is positive semidefinite. This completes the proof. $\square$

1.1 Examples of Convex Functions

- ***Exponential*** $f(x) = e^{ax}, \text{for all} a \in \mathbb{R}$ is convex on $\mathbb{R}$. To show e^{ax} is convex for all$a \in \mathbb{R}$, we could simply see the second derivative of function f, which is $a^2 e^{ax} \geq 0, \text{for all} a \in \mathbb{R}$
- ***Powers*** $f(x) = x^a$ is convex on $\mathbb{R}_{++}$ when $a \geq 1$ *or* $a \leq 0$, concave otherwise. Since $f^{''}(x) = a(a-1)x^{a-2}, x \in \mathbb{R}_{++}$ is non-negative when $a \geq 1$ *or* $a \leq 0$.
- ***Negative Logarithm*** $f(x) = -logx$ is convex on its domain $\mathbb{R}_{++}$, because $f^{''}(x) = \frac{1}{x^2} > 0, \text{for all} x \in \mathbb{R}_{++}$
- ***Norms*** Every norms on $\mathbb{R}^n$ is convex. Using triangle inequality and positive homogenity, $f(\lambda x + (1-\lambda)y) \leq f(\lambda x) + f((1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y), \text{for all} x, y \in \mathbb{R}^n$ $, 0 \leq \lambda \leq 1$.
- ***Max Function*** $f(\mathbf{x}) = max\{x_1, x_2, \cdots, x_n\}$ is convex on $\mathbb{R}^n$. To show this, $f(\lambda\mathbf{x} + (1-\lambda\mathbf{y})) = max\{\lambda x_1 + (1-\lambda)y_1, \cdots, \lambda x_n + (1-\lambda)y_n\} = \lambda x_i + (1-\lambda)y_i \leq \lambda x_{max} + (1-\lambda)y_{max} = \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}), \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- ***Quadratic over linear:*** $f(x, y) = \frac{x^2}{y}$ is convex for $y > 0$ since $\nabla^2 f(x, y) = [\frac{2}{y}, -\frac{2x}{y^2}; -\frac{2x}{y^2}, \frac{2x^2}{y^3}] = \frac{2}{y^3}[y; -x][y; -x]^\top \succeq \mathbf{0}$ for $y > 0$.

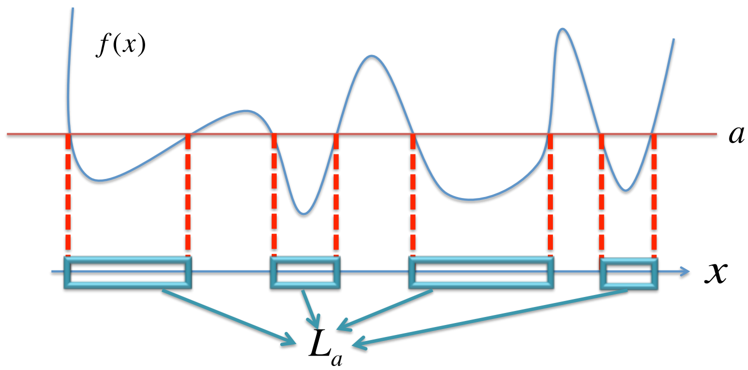

Figure 2: Sub-level set L_a of function $f(x)$

2 Sub-level set and epigraph

Definition 4. *The* ***Sub-level set*** *of function f at a is:*

$$L_a = \{x \mid f(x) \leq a\} \tag{8}$$

Sub-level set of a convex function is convex set for any value of a (note that the empty set is convex by convention). Figure 2 shows an example of a sub-level set.

Remark 1. *Sub-level sets of a convex function are convex.*

Proof. Basically, we must show that if $x_1, x_2 \in \mathcal{C}$ then any convex combination $\lambda x_1 + (1 - \lambda)x_2 \in \mathcal{C}$. If $x_1, x_2 \in \mathcal{C}$ then

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &\overset{(a)}{\leq} \lambda f(x_1) + (1-\lambda) f(x_2) \text{ for all} \lambda \in [0,1] \\ &\overset{(b)}{\leq} \lambda a + (1-\lambda)a \text{ for all} \lambda \in [0,1] \\ &= a, \end{aligned}$$

where (a) follows by using the convexity of f and (b) follows from using the definition of the level set. □

Definition 5. *The* ***epigraph*** *of function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as*

$$\mathbf{epi} f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid \mathbf{x} \in \mathbf{dom} f,\ f(\mathbf{x}) \leq t\} \tag{9}$$

Remark 2. *Function f is convex if and only if* **epi***f is convex.*

3 Jensen's inequality

The basic inequality: If f is convex and $\lambda \in [0, 1]$, then

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

The extended version I: If f is convex and $\lambda_1, \ldots, \lambda_k \geq 0$ and $\sum_{i=1}^{k} \lambda_i = 1$, then

$$f\left(\sum_{i=1}^{k} \lambda_k \mathbf{x}_k\right) \leq \lambda_1 f(\mathbf{x}_1) + \cdots + \lambda_k f(\mathbf{x}_k)$$

The extended version II: If f is convex, then

$$f\left(\mathbf{E}[\mathbf{z}]\right) \leq \mathbf{E}[f(\mathbf{z})]$$

for any random variable $\mathbf{z}$.