# 553.665 Introduction to Convexity, Fall 2022
## *Section 6*

Hongyu Cheng

October 7, 2022

**Definition 1** (Lower semicontinuous)**.** A function $f : D \to \mathbb{R} \cup \{\pm\infty\}$ is called *lower semicontinuous at* $\mathbf{x} \in D$ if

$$f(\mathbf{x}) \leq \liminf_{k\to\infty} f(\mathbf{x}_k)$$

for every sequence $\{\mathbf{x}_k\} \subseteq D$ with $\mathbf{x}_k \to \mathbf{x}$. We say that $f$ is *lower semicontinuous* if it is lower semicontinuous at each point $\mathbf{x}$ in its domain $D$. We say that $f$ is *upper semicontinuous* if $-f$ is lower semicontinuous.

**Exercise 1.** For a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$, the following are equivalent:

1. $f_\gamma = \{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) \leq \gamma\}$ is closed $\forall \gamma \in \mathbb{R}$.
2. $f$ is lower semicontinuous.
3. $\text{epi}(f) = \{(\mathbf{x}, t) \in \mathbb{R}^d \times \mathbb{R} : f(\mathbf{x}) \leq t\}$ is closed.

*Proof.* If $f(\mathbf{x}) = \infty$ for all $\mathbf{x}$, the result trivially holds. We thus assume that $f(\mathbf{x}) < \infty$ for at least one $\mathbf{x} \in \mathbb{R}^d$, so that epi$(f)$ is nonempty and there exist level sets of $f$ that are nonempty.
(i) $\implies$ (ii). Assume that the level set $f_\gamma$ is closed for every scalar $\gamma$. Suppose to the contrary that

$$f(\bar{\mathbf{x}}) > \liminf_{k\to\infty} f(\mathbf{x}_k)$$

for some $\bar{\mathbf{x}}$ and sequence $\{\mathbf{x}_k\}$ converging to $\bar{\mathbf{x}}$, and let $\bar{\gamma}$ be a scalar such that

$$f(\bar{\mathbf{x}}) > \bar{\gamma} > \liminf_{k\to\infty} f(\mathbf{x}_k).$$

Then there exists a subsequence $\{\mathbf{x}_{k_i}\}$ such that $f(\mathbf{x}_{k_i}) \leq \bar{\gamma}$ for all $i \in \mathbb{N}_+$, so that $\{\mathbf{x}_{k_i}\} \subseteq f_{\bar{\gamma}}$. Since $f_{\bar{\gamma}}$ is closed, $\bar{\mathbf{x}}$ must also belong to $f_{\bar{\gamma}}$, so $f(\bar{\mathbf{x}}) \leq \bar{\gamma}$, which leads to a contradiction.
(ii) $\implies$ (iii). Assume that $f$ is lower semicontinuous over $\mathbb{R}^d$, and let $\bar{\mathbf{x}}, \bar{t}$ be the limit of a sequence

$$\{(\mathbf{x}_k, t_k)\} \subseteq \text{epi}(f).$$

Then we have $f(\mathbf{x}_k) \leq t_k$, and by taking the limit as $k \to \infty$ and by using the lower semicontinuity of $f$ at $\bar{\mathbf{x}}$, we obtain

$$f(\bar{\mathbf{x}}) \leq \liminf_{k\to\infty} f(\mathbf{x}_k) \leq \bar{t}.$$

Hence, $(\bar{\mathbf{x}}, \bar{t}) \in \text{epi}(f)$ and epi$(f)$ is closed.

(iii) $\implies$ (i). Assume that epi$(f)$ is closed and let $\{\mathbf{x}_k\}$ be a sequence that converges to some $\bar{\mathbf{x}}$ and belongs to $f_\gamma$ for some $\gamma \in \mathbb{R}$. Then $(\mathbf{x}_k, \gamma) \in$ epi$(f)$ for all $k$ and $(\mathbf{x}_k, \gamma) \to (\bar{\mathbf{x}}, \gamma)$, so since epi$(f)$ is closed, we have $(\bar{\mathbf{x}}, \gamma) \in$ epi$(f)$. Hence, $\bar{\mathbf{x}}$ belongs to $f_\gamma$, implying that this set is closed. □

**Exercise 2.** Consider any convex function $f : \mathbb{R}^d \to \mathbb{R}$. Fix any $\bar{\mathbf{x}} \in \mathbb{R}^d$ and suppose $\mathbf{x}^* \in \mathbb{R}^d$ globally minimizes

$$\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2$$

for some $\rho > 0$. Prove that $\rho(\bar{\mathbf{x}} - \mathbf{x}^*) \in \partial f(\mathbf{x}^*)$ is a subgradient of $f$ at $\mathbf{x}^*$.

*Proof.* Let $\mathbf{g} = \rho(\bar{\mathbf{x}} - \mathbf{x}^*)$, then by definition we just need to prove that $f$ is lower bounded by the following linear function for all $\mathbf{x} \in \mathbb{R}^d$

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \mathbf{g}^T(\mathbf{x} - \mathbf{x}^*).$$

We can prove this directly. For any $\mathbf{x} \in \mathbb{R}^d$, we have

$$\begin{aligned}
f(\mathbf{x}^*) + \frac{\rho}{2}\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2^2 &\leq f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 \\
&= f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^* + \mathbf{x}^* - \bar{\mathbf{x}}\|_2^2 \\
&= f(\mathbf{x}) + \underbrace{\rho(\mathbf{x}^* - \bar{\mathbf{x}})^T}_{\mathbf{g}^T}(\mathbf{x} - \mathbf{x}^*) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2 + \frac{\rho}{2}\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2^2
\end{aligned}$$

Cancelling the last quadratic from both sides gives a weaker result than needed:

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) - \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2, \quad \forall \mathbf{x} \in \mathbb{R}^d. \tag{1}$$

For any $\lambda \in (0, 1]$, applying Equation (1) at $\lambda\mathbf{x} + (1 - \lambda)\mathbf{x}^*$, one can obtain that

$$\begin{aligned}
f(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^*) &\geq f(\mathbf{x}^*) + \mathbf{g}^T(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^* - \mathbf{x}^*) - \frac{\rho}{2}\|\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^* - \mathbf{x}^*\|_2^2 \\
&= f(\mathbf{x}^*) + \mathbf{g}^T(\lambda\mathbf{x} - \lambda\mathbf{x}^*) - \frac{\rho}{2}\|\lambda\mathbf{x} - \lambda\mathbf{x}^*\|_2^2 \\
&= f(\mathbf{x}^*) + \lambda\mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) - \lambda^2\frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2.
\end{aligned}$$

Using convexity of $f$, we can strengthen this since

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}^*) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}^*),$$

therefore,

$$\lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}^*) \geq f(\mathbf{x}^*) + \lambda\mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) - \lambda^2\frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2,$$

that is,

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) - \lambda\frac{\rho}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2.$$

Taking the limit as $\lambda \to 0$ gives the claim. □

**Exercise 3.** For any $\mu$-strongly convex $f$ with global minimizer $\mathbf{x}^*$, show that $\forall \mathbf{x} \in \mathbb{R}^d$ has

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \frac{\mu}{2} \|\mathbf{x} - \mathbf{x}^*\|_2^2 .$$

*Proof.* Let $h(\mathbf{x}) := f(\mathbf{x}) - \frac{\mu}{2} \|\mathbf{x}\|_2^2$, which is convex by assumption. Then $\mathbf{x}^*$ minimizes $f(\mathbf{x}) = h(\mathbf{x}) + \frac{\mu}{2} \|\mathbf{x} - 0\|_2^2$. This is precisely the shape of the problem considered in Exercise 2. Letting $\rho = \mu$, $\bar{\mathbf{x}} = 0$, we then know for any $\mathbf{x} \in \mathbb{R}^d$,

$$\begin{aligned} h(\mathbf{x}) &\geq h(\mathbf{x}^*) + \rho(0 - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \\ &= h(\mathbf{x}^*) + \frac{\rho}{2} \|\mathbf{x}^*\|_2^2 + \frac{\rho}{2} \|\mathbf{x} - \mathbf{x}^*\|_2^2 - \frac{\rho}{2} \|\mathbf{x}\|_2^2 . \end{aligned}$$

Therefore,

$$\underbrace{h(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{x}\|_2^2}_{f(\mathbf{x})} \geq \underbrace{h(\mathbf{x}^*) + \frac{\rho}{2} \|\mathbf{x}^*\|_2^2}_{f(\mathbf{x}^*)} + \frac{\rho}{2} \|\mathbf{x} - \mathbf{x}^*\|_2^2 .$$

□

*Remark* 1. Exercise 2 and Exercise 3 did not require differentiability of $f$.