**COMM 616: Modern Optimization with Applications in ML and OR 2024-25 Fall**

# Lecture 3: Element of Convex Analysis II

Instructor: Jiajin Li  Scribe: Alex Chow

## 1 Convex Functions

**Definition 1.** *Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be an extended real-valued function that is not identically $+\infty$. We say that $f$ is convex if*

$$f(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) \le \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2),$$

*for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$ and $\alpha \in [0,1]$.*

**Proposition 2** (Connection Between Convex Sets and Convex Functions)**.** *Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be an extended real-valued function that is not identically $+\infty$. Then, $f$ is convex if and only if $\operatorname{epi}(f)$ is convex.*

**Proposition 3** (Jensen's Inequality)**.** *Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be as defined above. Then, $f$ is convex if and only if*

$$f\left(\sum_{i=1}^{k} \alpha_i \boldsymbol{x}_i\right) \le \sum_{i=1}^{k} \alpha_i f(\boldsymbol{x}_i),$$

*for any $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathbb{R}^d$ and $\alpha_1, \ldots, \alpha_k \in [0,1]$ such that $\sum_{i=1}^{k} \alpha_i = 1$.*

*Proof.*

$$\Leftarrow: \text{ By definition.}$$

$$\Rightarrow: f \text{ is convex } \Rightarrow \operatorname{epi}(f) \text{ is convex: } \sum_{i=1}^{k} \alpha_i (\boldsymbol{x}_i, f(\boldsymbol{x}_i)) = \left(\sum_{i=1}^{k} \alpha_i \boldsymbol{x}_i, \sum_{i=1}^{k} \alpha_i f(\boldsymbol{x}_i)\right) \in \operatorname{epi}(f).$$

■

**Theorem 4** (Geometric Characterization of Convex Functions)**.** *Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be a convex function such that $\operatorname{epi}(f)$ is closed. Then, $f$ can be represented as the pointwise supremum of all affine functions $h : \mathbb{R}^d \to \mathbb{R}$ satisfying $h \le f$.*

**Observation 5** (Geometric Characterization of Convex Functions and their Conjugate Functions)**.** *We consider a set:*

$$\mathcal{F} = \{(\boldsymbol{y}, c) \in \mathbb{R}^d \times \mathbb{R} : \boldsymbol{y}^T \boldsymbol{x} - c \le f(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{R}^d\}.$$

*which consists of the coefficients of affine functions $h : \mathbb{R}^d \to \mathbb{R}$ satisfying $h \le f$. Clearly, we have:*

$$\boldsymbol{y}^T \boldsymbol{x} - c \le f(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d \iff \sup_{\boldsymbol{x} \in \mathbb{R}^d} (\boldsymbol{y}^T \boldsymbol{x} - f(\boldsymbol{x})) \le c.$$

*This shows that $\mathcal{F}$ is the epigraph of the function $f^* : \mathbb{R}^d \to \overline{\mathbb{R}}$, given by:*

$$f^*(\boldsymbol{y}) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} \left(\boldsymbol{y}^T \boldsymbol{x} - f(\boldsymbol{x})\right),$$

*which is the **conjugate function**. Since $\mathcal{F}$ is closed and convex, $f^*$ is convex.*

## 2 Convexity-Preserving Transformations

As in the case of convex sets, it's sometimes difficult to check directly from the definition whether a given function is convex or not.

**Theorem 6.** *The following hold:*

*(i)* ***(Non-negative Combination)****: Let* $f_1, \ldots, f_m : \mathbb{R}^d \to \overline{\mathbb{R}}$ *be convex functions satisfying* $\bigcap_{i=1}^m \operatorname{dom}(f_i) \neq \emptyset$*. Then, for any* $\alpha_1, \alpha_2, \ldots, \alpha_m \geq 0$*, the function* $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ *defined by*

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \alpha_i f_i(\boldsymbol{x})$$

*is convex.*

*(ii)* ***(Pointwise Supremum)****: Let* $\mathcal{I}$ *be an index set for* $\{f_i\}_{i\in\mathcal{I}}$ *for all* $i \in \mathcal{I}$*. Define the pointwise supremum* $f_i : \mathbb{R}^d \to \mathbb{R}$ *of* $\{f_i\}_{i\in\mathcal{I}}$ *by*

$$f(\boldsymbol{x}) = \sup_{i\in\mathcal{I}} f_i(\boldsymbol{x})$$

*(note:* $\mathcal{I}$ *may not be a finite set). Suppose that* $\operatorname{dom}(f) \neq \emptyset$*. Then, the function* $f$ *is convex.*

**Proof Idea** *The epigraph of* $f$ *is the intersection of the epigraphs of* $f_i$*. The intersection of (possibly infinite) convex sets remains convex.* ■

*(iii)* ***(Affine Transformation)****: Let* $g : \mathbb{R}^d \to \overline{\mathbb{R}}$ *be a convex function, and let* $A : \mathbb{R}^m \to \mathbb{R}^d$ *be an affine mapping. Suppose that* $\operatorname{range}(A) \cap \operatorname{dom}(g) \neq \emptyset$*. Then,* $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ *defined by*

$$f(\boldsymbol{x}) = g(A(\boldsymbol{x}))$$

*is convex.*

*(iv)* ***(Composition with an Increasing Convex Function)****: Let* $g : \mathbb{R}^d \to \overline{\mathbb{R}}$ *and* $h : \mathbb{R} \to \overline{\mathbb{R}}$ *be convex functions but not identically* $+\infty$*. Suppose that* $h$ *is increasing on dom(*$h$*). Define the function* $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ *by* $f(\boldsymbol{x}) = h(g(\boldsymbol{x}))$ *with the convention* $h(+\infty) = +\infty$*. Suppose that dom(*$f$*)* $\neq \emptyset$*. Then* $f$ *is convex.*

**Example 7** (Counter-example)**.** *Consider*

$$h(x) = -\sqrt{x} \quad \text{when } x \geq 0$$

$$\text{and } g(x) = x^2.$$

*Then, we have*

$$f(x) = h(g(x)) = -\sqrt{|x|^2} = -|x|,$$

*which is nonconvex.*

*(v)* ***(Restriction on Lines)****: Given a function* $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ *that is not identically* $+\infty$*, a point* $\boldsymbol{x}_0 \in \mathbb{R}^d$*, and a direction* $\boldsymbol{h} \in \mathbb{R}^n$*, define the function* $\tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ *by*

$$\tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}} = f(\boldsymbol{x}_0 + t\boldsymbol{h}).$$

*Then, the function* $f$ *is convex if and only if the function* $\tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}}$ *is convex for all* $\boldsymbol{x}_0 \in \mathbb{R}^d$ *and* $t \in \mathbb{R}$*.*

**Remark 8.** *The convexity of a high-dimensional function can always be verified by examining a collection of one-dimensional functions.*

*Proof.* Let $\boldsymbol{x}_0, \boldsymbol{h}$ be arbitrary. Then, for any $t_1, t_2 \in \mathbb{R}$ and $\alpha \in [0,1]$, we have

$$\begin{aligned} \hat{f}_{\boldsymbol{x}_0,\boldsymbol{h}}(\alpha t_1 + (1-\alpha)t_2) &= f(\boldsymbol{x}_0 + (\alpha t_1 + (1-\alpha)t_2)\boldsymbol{h}) \\ &\leq \alpha f(\boldsymbol{x}_0 + t_1\boldsymbol{h}) + (1-\alpha) f(\boldsymbol{x}_0 + t_2\boldsymbol{h}) \\ &= \alpha \tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}}(t_1) + (1-\alpha)\tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}}(t_2), \end{aligned}$$

where the first inequality follows from the convexity of $f$.

Conversely, let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$ and $\alpha \in [0,1]$. Upon setting $\boldsymbol{x}_0 = \boldsymbol{x}_1$ and $\boldsymbol{h} = \boldsymbol{x}_2 - \boldsymbol{x}_1$, we have

$$\begin{aligned} f((1-\alpha)\boldsymbol{x}_1 + \alpha\boldsymbol{x}_2) &= \tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}}(\alpha) \\ &= \tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}}(\alpha \cdot 1 + (1-\alpha)\cdot 0) \\ &\leq \alpha \tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}}(1) + (1-\alpha)\tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}}(0) \\ &= \alpha f(\boldsymbol{x}_2) + (1-\alpha) f(\boldsymbol{x}_1), \end{aligned}$$

where the first inequality follows from the convexity of $\tilde{f}_{\boldsymbol{x}_0,\boldsymbol{h}}$.

■

# 3 Differentiable Convex Functions

When a given function is differentiable, it's possible to characterize its convexity via their derivatives.

**Theorem 9.** *Let $f : \Omega \to \mathbb{R}$ be a differentiable function on the open set $\Omega \subseteq \mathbb{R}^d$ and $\mathcal{S} \subseteq \mathbb{R}^d$ be a convex set. Then, $f$ is convex on $\mathcal{S}$ if and only if*

$$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}})$$

*for all $\boldsymbol{x}, \bar{\boldsymbol{x}} \in \mathcal{S}$.*

When $f$ is twice-differentiable, then we have

**Theorem 10.** *Let $f : \mathcal{S} \to \mathbb{R}$ be a twice differentiable function on the* ***open convex set*** *$\mathcal{S} \subseteq \mathbb{R}^d$. Then $f$ is convex if and only if $\nabla^2 f(\boldsymbol{x}) \succeq 0$ for all $\boldsymbol{x} \in \mathcal{S}$.*

**Example 11** (Counter-example)**.** *Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = x^2 - y^2$. This function is convex on the set $\mathcal{S} = \mathbb{R} \times \{0\}$. However, its Hessian is not postive semi-definite for any $x, y \in \mathbb{R}$, i.e.,*

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

# 4 Establishing Convexity of Functions

(i) Consider $f : \mathbb{R}^d \times \mathcal{S}^d_{++} \to \mathbb{R}$, $f(\boldsymbol{x}, \mathbf{Y}) = \boldsymbol{x}^T\mathbf{Y}^{-1}\boldsymbol{x}$.

*Proof.* By Proposition 2, we check the convexity of its epigraph, i.e.,

$$\begin{aligned} \text{epi}(f) &= \{(\boldsymbol{x}, \mathbf{Y}, t) \in \mathbb{R}^d \times \mathcal{S}^d_{++} \times \mathbb{R} : \mathbf{Y} \succ 0, \boldsymbol{x}^T\mathbf{Y}^{-1}\boldsymbol{x} \leq t\}, \\ &= \left\{(\boldsymbol{x}, \mathbf{Y}, t) \in \mathbb{R}^d \times \mathcal{S}^d_{++} \times \mathbb{R} : \mathbf{Y} \succ 0, \begin{bmatrix} \mathbf{Y} & \boldsymbol{x} \\ \boldsymbol{x}^T & t \end{bmatrix} \succeq 0 \right\}. \end{aligned}$$

where the last equality follows from the Schur complement. ■

(ii) Let $f : \mathbb{R}^{m\times n} \to \mathbb{R}_+$, $f(\mathbf{X}) = ||\mathbf{X}\|_2$, where $\|\cdot\|_2$ is the spectral norm.

It is well known that (see, e.g., [1])

$$\|\mathbf{X}\|_2 = \sup_{\|\boldsymbol{v}\|_2=1, \|\boldsymbol{u}\|_2=1} \boldsymbol{v}^T \mathbf{X} \boldsymbol{u}.$$

By Theorem 6 (ii), we finished the proof.

(iii) Let $f : \mathbb{R}^d \to \mathbb{R}$ be given by $f(\boldsymbol{x}) = \log\left(\sum_{i=1}^d \exp(x_i)\right)$.

*Proof.* By Theorem 10, we compute the Hessian of $f$:

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \begin{cases} \frac{\exp(x_i)}{\sum_{k=1}^d \exp(x_k)} - \frac{\exp(x_i)\exp(x_j)}{\left(\sum_{k=1}^d \exp(x_k)\right)^2} & \text{if } i = j, \\ -\frac{\exp(x_i)\exp(x_j)}{\left(\sum_{k=1}^d \exp(x_k)\right)^2} & \text{if } i \neq j. \end{cases}$$

This gives the compact form as

$$\nabla^2 f(\boldsymbol{x}) = \frac{1}{(\mathbf{1}^T \boldsymbol{z})^2} \left(\operatorname{diag}(\boldsymbol{z}) - \boldsymbol{z}\boldsymbol{z}^T\right),$$

where $\boldsymbol{z} = (\exp(z_1), \exp(z_2), \ldots, \exp(z_d))$. Next, we are trying to check the Hessian $\nabla^2 f(\boldsymbol{x})$ is positive semi-definite for any $\boldsymbol{x} \in \mathbb{R}^d$. That is, for any $\boldsymbol{v} \in \mathbb{R}^d$, we have

$$\begin{aligned} \boldsymbol{v}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{v} &= \frac{1}{(\mathbf{1}^T \boldsymbol{z})^2} \left[ \left(\sum_{i=1}^d z_i\right)\left(\sum_{i=1}^d z_i v_i^2\right) - \left(\sum_{i=1}^d z_i v_i\right)^2 \right] \\ &= \frac{1}{(\mathbf{1}^T \boldsymbol{z})^2} \left[ \left(\sum_{i=1}^d \sqrt{z_i}^2\right)\left(\sum_{i=1}^d (\sqrt{z_i} v_i)^2\right) - \left(\sum_{i=1}^d \sqrt{z_i}\sqrt{z_i} v_i\right)^2 \right] \\ &\geq 0, \end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality. ∎

(iv) Let $f : \mathbb{R}^d \times \mathbb{R}_+ \to \overline{\mathbb{R}}$ be given by $f(\boldsymbol{x}, t) = t \log\left(\sum_{i=1}^d \exp\left(\frac{x_i}{t}\right)\right)$.

**Sketch of Proof** By Proposition 2, we check the convexity of its epigraph, i.e.,

$$\operatorname{epi}(f) = \left\{ (\boldsymbol{x}, t, r) \in \mathbb{R}^d \times \mathbb{R}_{++} \times \mathbb{R} : \log\left(\sum_{i=1}^d \exp\left(\frac{x_i}{t}\right)\right) \leq \frac{r}{t} \right\}.$$

By the properties of the perspective function, we only have to check the set

$$\left\{ (\boldsymbol{x}, r) \in \mathbb{R}^d \times \mathbb{R} : \log\left(\sum_{i=1}^d \exp(x_i)\right) \leq r \right\}.$$

∎

(v) Let $f : \mathcal{S}^d_{++} \to \mathbb{R}$ be given by $f(\mathbf{X}) = -\ln\det\mathbf{X}$. For those readers who are well versed in matrix calculus (see, e.g., [3] for a comprehensive treatment), the following formulas should be familiar:

$$\nabla f(\mathbf{X}) = \mathbf{X}^{-1}, \quad \nabla^2 f(\mathbf{X}) = \mathbf{X}^{-1} \otimes \mathbf{X}^{-1}.$$

Here, $\otimes$ denotes the Kronecker product. Since $\mathbf{X}^{-1} \succ 0$, it can be shown that $\mathbf{X}^{-1} \otimes \mathbf{X}^{-1} \succ 0$.

*Proof.* Let $\mathbf{X}_0 \in \mathcal{S}^d_{++}$ and $\mathbf{H} \in \mathcal{S}^d$. Define the set

$$\mathcal{D} = \{t \in \mathbb{R} : \mathbf{X}_0 + t\mathbf{H} \succ 0\} = \{t \in \mathbb{R} : \lambda_{\min}(\mathbf{X}_0 + t\mathbf{H}) > 0\}.$$

Since $\lambda_{\min}(\cdot)$ is continuous (see, e.g., [2, Chapter IV, Theorem 4.11], we see that $\mathcal{D}$ is open and convex. Now, we consider $f_{\mathbf{X}_0,\mathbf{H}} : \mathcal{D} \to \mathbb{R}$ given by

$$\hat{f}_{\mathbf{X}_0,\mathbf{H}}(t) = f(\mathbf{X}_0 + t\mathbf{H}).$$

For any $t \in D$, we compute

$$\begin{aligned}
\tilde{f}_{\mathbf{X}_0,\mathbf{H}}(t) &= -\ln\det(\mathbf{X}_0 + t\mathbf{H}) \\
&= -\ln\det\left(\mathbf{X}_0^{\frac{1}{2}}\left(\mathbf{I} + t\mathbf{X}_0^{-\frac{1}{2}}\mathbf{H}\mathbf{X}_0^{-\frac{1}{2}}\right)\mathbf{X}_0^{\frac{1}{2}}\right) \\
&= -\left(\sum_{i=1}^{d} \ln(1 + t\lambda_i) + \ln\det\mathbf{X}_0\right),
\end{aligned}$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of $\mathbf{X}_0^{-\frac{1}{2}}\mathbf{H}\mathbf{X}_0^{-\frac{1}{2}}$.

Then, by Theorem 6 (v), we only have to prove the convexity of $\tilde{f}_{\mathbf{X}_0,\mathbf{H}}(\cdot)$. Since the domain $\mathcal{D}$ is open and convex, we can apply Theorem 10 and check its second derivative:

$$\tilde{f}''_{\mathbf{X}_0,\mathbf{H}}(t) = \sum_{i=1}^{d} \frac{\lambda_i^2}{(1 + t\lambda_i)^2} \geq 0.$$

■

## References


[1] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge university press, 2012. 4

[2] Ren-Cang Li. *Matrix Perturbation Theory*. Chapman and Hall/CRC, 2006. 5

[3] Jan R. Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 2019. 5



This lecture draws extensively from the material available at ENGG 5501 Handout 2 taught by Prof. Anthony Man-Cho So.