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

Lecture 2: Element of Convex Analysis I

Instructor: Jiajin Li September 3rd, 2024

First and foremost, the concept of convexity plays a crucial role in both the theoretical and algorithmic aspects of optimization.

1 Why Convexity is Special?

(i) They exhibit favorable geometric properties, such as the fact that any local minimum is also a global minimum.

(ii) There are excellent software tools (e.g., CVX, Mosek, Gurobi) that can efficiently solve a wide range of convex problems.

(iii) Although many machine learning tasks are still nonconvex, modern nonconvex theories and algorithm designs continue to heavily rely on fundamental convex analysis techniques.

Theorem 1 (Local Implies Global)**.** *Consider an optimization problem*

$$\min_{\boldsymbol{x}\in\mathcal{S}} f(\boldsymbol{x}),$$

where $f:\mathcal{S}\to\mathbb{R}$ is a convex function and $\mathcal{S}$ is a convex set. Then, any local minima is also a global minima.

Proof. Let $\bar{\boldsymbol{x}}$ be a local minima. By its definition (see Definition 5 in Lecture 1), we have: There exists $\epsilon > 0$ such that $f(\bar{\boldsymbol{x}}) \le f(\boldsymbol{x})$, for all $\boldsymbol{x} \in \mathcal{B}(\bar{\boldsymbol{x}}, \epsilon)$. Suppose for the sake of contradiction, there exists a point $\boldsymbol{z} \in \mathcal{S}$ with

$$f(\boldsymbol{z}) < f(\bar{\boldsymbol{x}}).$$

Moreover, due the convexity of the set $\mathcal{S}$, we have

$$\alpha\bar{\boldsymbol{x}} + (1-\alpha)\boldsymbol{z} \in \mathcal{S}, \forall \alpha \in [0,1].$$

By the convexity of f, we have

$$\begin{aligned} f(\alpha\bar{\boldsymbol{x}} + (1-\alpha)\boldsymbol{z}) &\le \alpha f(\bar{\boldsymbol{x}}) + (1-\alpha)f(\boldsymbol{z}) \\ &< \alpha f(\bar{\boldsymbol{x}}) + (1-\alpha)f(\bar{\boldsymbol{x}}) = f(\bar{\boldsymbol{x}}). \end{aligned} \tag{1}$$

However, as $\alpha \to 1$, we have $\alpha\bar{\boldsymbol{x}} + (1-\alpha)\boldsymbol{z} \to \bar{\boldsymbol{x}}$ and the inequality (1) contradicts the fact that $\bar{\boldsymbol{x}}$ is a local minima. ■

2 Convex Sets

Definition 2. *Let $\mathcal{S} \subseteq \mathbb{R}^d$. We say that $\mathcal{S}$ is convex if $\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in \mathcal{S}$ whenever $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$ and $\alpha \in [0,1]$.*

Proposition 3. *Let $\mathcal{S} \subseteq \mathbb{R}^d$ be non-empty. Then, the following are equivalent:*

(i) $\mathcal{S}$ is convex.

(ii) Any convex combination of points in $\mathcal{S}$ belongs to $\mathcal{S}$.

Observation 4. *A set $\mathcal{S}$ is convex if for any $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathcal{S}$, the $\boldsymbol{x}-\boldsymbol{y}$ line entirely in the set $\mathcal{S}$.*

Example 5 (Some Examples of Convex Sets). *(i) Non-Negative Orthant: $\mathbb{R}^d_+ = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \geq \mathbf{0}\}$.*

(ii) Hyperplane: $H(\boldsymbol{s}, c) = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{s}^T\boldsymbol{x} = c\}$.

(iii) Halfspaces: $H^-(\boldsymbol{s}, c) = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{s}^T\boldsymbol{x} \leq c\}, H^+(\boldsymbol{s}, c) = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{s}^T\boldsymbol{x} \geq c\}$.

(iv) Euclidean Ball: $\mathcal{B}(\bar{\boldsymbol{x}}, r) = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2 \leq r\}$.

Proof Idea *By Definition 2 and apply the triangle inequality.* ■

(v) Ellipsoid: $\mathcal{E}(\bar{\boldsymbol{x}}, \mathbf{Q}) = \{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x} - \bar{\boldsymbol{x}})^T\mathbf{Q}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \leq 1\}$, where $\mathbf{Q}$ is an $d \times d$ symmetric, positive definite matrix (i.e., $\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ and denoted by $\mathbf{Q} \in \mathcal{S}^d_{++}$).

Proof Idea *proving that $\|\boldsymbol{x}\|_{\mathbf{Q}} = \sqrt{\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x}}$ is a norm and repeat above.* ■

(vi) Simplex: $\Delta = \left\{\sum_{i=0}^d \alpha_i x_i : \sum_{i=0}^d \alpha_i = 1, \alpha_i \geq 0 \text{ for } i = 0, 1, \ldots, n\right\}$, where $x_0, x_1, \ldots, x_d$ are vectors in $\mathbb{R}^d$ such that the vectors $x_1 - x_0, x_2 - x_0, \ldots, x_d - x_0$ are linearly independent.

(vii) Convex Cone: A set $\mathcal{K} \subseteq \mathbb{R}^d$ is called a cone if $\{\alpha x : \alpha > 0\} \subseteq \mathcal{K}$ whenever $\boldsymbol{x} \in \mathcal{K}$. If $\mathcal{K}$ is also convex, then $\mathcal{K}$ is called a convex cone.

(viii) Positive Semidefinite Cone: $\mathcal{S}^d_+ = \{\mathbf{Q} \in \mathcal{S}^d : \boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} \geq 0 \text{ for all } \boldsymbol{x} \in \mathbb{R}^d\}$.

2.1 Convexity-preserving Operations

Establishing the convexity of a set directly from its definition can sometimes be challenging. Here, we will study certain operations such that if $\mathcal{S}$ is convex, then applying an operator A ensures that $A(\mathcal{S})$ is also convex.

(i) Intersection of two convex sets if convex. That is, if $\mathcal{S}_1$ and $\mathcal{S}_2$ are convex, then $\mathcal{S}_1 \cap \mathcal{S}_2$ is also convex.

Remark 6. *This is also true for infinite/finite intersections or Minkowski sums. However, the union of two convex sets is not necessarily convex.*

(ii) Convexity is preserved under affine mappings.

Definition 7 (Affine Functions). *We say that a map $A : \mathbb{R}^n \to \mathbb{R}^m$ is affine if*

$$A(\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) = \alpha A(\boldsymbol{x}_1) + (1-\alpha)A(\boldsymbol{x}_2)$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}$. It can be shown that A is affine iff there exist $A_0 \in \mathbb{R}^{m \times d}$ and $\boldsymbol{y}_0 \in \mathbb{R}^m$ such that $A(x) = A_0\boldsymbol{x} + \boldsymbol{y}_0$ for all $\boldsymbol{x} \in \mathbb{R}^d$.

Proposition 8. *Let $A : \mathbb{R}^d \to \mathbb{R}^m$ be an affine mapping and $\mathcal{S} \subseteq \mathbb{R}^d$ be a convex set. Then, the image $A(\mathcal{S}) = \{A(\boldsymbol{x}) \in \mathbb{R}^d : \boldsymbol{x} \in S\}$ is convex. Conversely, if $\mathcal{T} \subseteq \mathbb{R}^m$ is a convex set, then the inverse image $A^{-1}(\mathcal{T}) = \{\boldsymbol{x} \in \mathbb{R}^d : A(\boldsymbol{x}) \in \mathcal{T}\}$ is convex.*

Let us use Proposition 8 to revisit the convexity of Ellipsoid.

Proof. Consider the ball $\mathcal{B}(\mathbf{0}, r) = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x}^T\boldsymbol{x} \leq r^2\} \subseteq \mathbb{R}^d$, where $r > 0$. Clearly, $\mathcal{B}(\mathbf{0}, r)$ is convex. Now, let $\mathbf{Q}$ be a $d \times d$ symmetric positive definite matrix. Then, it is well-known that $\mathbf{Q}$ is invertible and the $d \times d$ symmetric matrix $\mathbf{Q}^{-1}$ is also positive definite. Moreover, there exists a $d \times d$ symmetric matrix $\mathbf{Q}^{-1/2}$ such that $\mathbf{Q}^{-1} = \mathbf{Q}^{-1/2}\mathbf{Q}^{-1/2}$. Thus, we may define an affine mapping $A : \mathbb{R}^d \to \mathbb{R}^d$ by $A(\boldsymbol{x}) = \mathbf{Q}^{-1/2}\boldsymbol{x} + \bar{\boldsymbol{x}}$. We claim that

$$A(\mathcal{B}(\mathbf{0}, r)) = \left\{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x} - \bar{\boldsymbol{x}})^T \mathbf{Q}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \leq r^2\right\} = \mathcal{E}(\bar{\boldsymbol{x}}, \mathbf{Q}/r^2).$$

Indeed, let $x \in \mathcal{B}(\mathbf{0}, r)$ and consider the point $A(\boldsymbol{x})$. We compute

$$(A(\boldsymbol{x}) - \bar{\boldsymbol{x}})^T \mathbf{Q}(A(\boldsymbol{x}) - \bar{\boldsymbol{x}}) = \boldsymbol{x}^T \mathbf{Q}^{-1/2}\mathbf{Q}\mathbf{Q}^{-1/2}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{x} \leq r^2$$

i.e., $A(\mathcal{B}(\mathbf{0}, r)) \subseteq \mathcal{E}(\bar{\boldsymbol{x}}, \mathbf{Q}/r^2)$. Conversely, let $\boldsymbol{x} \in \mathcal{E}\left(\bar{\boldsymbol{x}}, \mathbf{Q}/r^2\right)$. Consider the point $\boldsymbol{y} = \mathbf{Q}^{1/2}(\boldsymbol{x} - \bar{\boldsymbol{x}}) = A^{-1}(\boldsymbol{x})$. Then, we have $\boldsymbol{y}^T\boldsymbol{y} \leq r^2$, which implies that $\mathcal{E}(\bar{\boldsymbol{x}}, \mathbf{Q}/r^2) \subseteq A(\mathcal{B}(\mathbf{0}, r))$. Hence, we conclude from the above calculation and Proposition 8 that $\mathcal{E}(\bar{\boldsymbol{x}}, \mathbf{Q}/r^2)$ is convex. ∎

(iii) Convexity is preserved by perspective functions. Define the perspective function $P : \mathbb{R}^d \times \mathbb{R}_{++} \to \mathbb{R}^d$ by $P(\boldsymbol{x}, t) = \frac{\boldsymbol{x}}{t}$.

Proposition 9. *Let $P : \mathbb{R}^d \times \mathbb{R}_{++} \to \mathbb{R}^d$ be the perspective function and $\mathcal{S} \subseteq \mathbb{R}^d \times \mathbb{R}_{++}$ be a convex set. Then, the image $P(\mathcal{S}) = \left\{\boldsymbol{x}/t \in \mathbb{R}^d : (\boldsymbol{x}, t) \in S\right\}$ is convex. Conversely, if $\mathcal{T} \subseteq \mathbb{R}^d$ is a convex set, then the inverse image $P^{-1}(\mathcal{T}) = \left\{(\boldsymbol{x}, t) \in \mathbb{R}^d \times \mathbb{R}_{++} : \boldsymbol{x}/t \in \mathcal{T}\right\}$ is convex.*

2.2 Projection onto Closed Convex Sets

The projection operator plays a crucial role in optimization algorithms. Specifically, the efficiency of these algorithms depends in part on the efficient computation of the projection operator.

Theorem 10. *Let $\mathcal{S} \subseteq \mathbb{R}^d$ be non-empty, closed and convex. Then, for every $\boldsymbol{x} \in \mathbb{R}^d$, there exists a unique $\boldsymbol{z}^\star \in \mathcal{S}$ that is closest (in the Euclidean norm) to $\boldsymbol{x}$.*

Proof Idea A useful characterization of projection:

$$\text{proj}_{\mathcal{S}}(\boldsymbol{x}) = \text{argmin}_{\boldsymbol{z} \in \mathcal{S}} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2.$$

We apply the Weierstrass theorem (which states that a continuous function over a compact set attains its minimum and maximum) to prove existence. Moreover, we use the method of contradiction to prove uniqueness. ∎

Theorem 11. *Let $\mathcal{S} \subseteq \mathbb{R}^d$ be non-empty, closed and convex. Given any $\boldsymbol{x} \in \mathbb{R}^d$, we have $\boldsymbol{z}^\star = \text{proj}_{\mathcal{S}}(\boldsymbol{x})$ if and only $\boldsymbol{z}^\star \in \mathcal{S}$ and*

$$(\boldsymbol{z} - \boldsymbol{z}^\star)^T(\boldsymbol{x} - \boldsymbol{z}^\star) \leq 0, \quad \forall \boldsymbol{z} \in \mathcal{S}.$$

Proof. Let $\boldsymbol{z}^\star = \text{proj}_{\mathcal{S}}(\boldsymbol{x})$ and $\boldsymbol{z} \in \mathcal{S}$. Consider points of the form $\boldsymbol{z}(\alpha) = \alpha\boldsymbol{z} + (1 - \alpha)\boldsymbol{z}^\star$, where $\alpha \in [0, 1]$. By convexity of the set $\mathcal{S}$, we have $\boldsymbol{z}(\alpha) \in \mathcal{S}$. Moreover, we have

$$\|\boldsymbol{z}^\star - \boldsymbol{x}\|_2 \leq \|\boldsymbol{z}(\alpha) - \boldsymbol{x}\|_2$$

for all $\alpha \in [0, 1]$. On the other hand, note that

$$\begin{aligned}\|\boldsymbol{z}(\alpha) - \boldsymbol{x}\|_2^2 &= (\boldsymbol{z}^\star + \alpha(\boldsymbol{z} - \boldsymbol{z}^\star) - \boldsymbol{x})^T (\boldsymbol{z}^\star + \alpha(\boldsymbol{z} - \boldsymbol{z}^\star) - \boldsymbol{x}) \\ &= \|\boldsymbol{z}^\star - \boldsymbol{x}\|_2^2 + 2\alpha(\boldsymbol{z} - \boldsymbol{z}^\star)^T(\boldsymbol{z}^\star - \boldsymbol{x}) + \alpha^2\|\boldsymbol{z} - \boldsymbol{z}^\star\|_2^2.\end{aligned}$$

Thus, we see that $\|\boldsymbol{z}(\alpha) - \boldsymbol{x}\|_2^2 \geq \|\boldsymbol{z}^\star - \boldsymbol{x}\|_2^2$ for all $\alpha \in [0, 1]$ if and only if $(\boldsymbol{z} - \boldsymbol{z}^\star)^T(\boldsymbol{z}^\star - \boldsymbol{x}) \geq 0$. This is precisely the stated condition.

Conversely, suppose that for some $\boldsymbol{z}' \in \mathcal{S}$, we have $(\boldsymbol{z} - \boldsymbol{z}')^T (\boldsymbol{x} - \boldsymbol{z}') \leq 0$ for all $\boldsymbol{z} \in \mathcal{S}$. Upon setting $\boldsymbol{z} = \text{proj}_{\mathcal{S}}(\boldsymbol{x})$, we have

$$(\text{proj}_{\mathcal{S}}(\boldsymbol{x}) - \boldsymbol{z}')^T (\boldsymbol{x} - \boldsymbol{z}') \leq 0. \tag{2}$$

On the other hand, by our argument in the preceding paragraph, the point $\text{proj}_{\mathcal{S}}(\boldsymbol{x})$ satisfies

$$(\boldsymbol{z}' - \text{proj}_{\mathcal{S}}(\boldsymbol{x}))^T (\boldsymbol{x} - \text{proj}_{\mathcal{S}}(\boldsymbol{x})) \leq 0. \tag{3}$$

Upon adding (2) and (3), we obtain

$$(\text{proj}_{\mathcal{S}}(\boldsymbol{x}) - \boldsymbol{z}')^T (\text{proj}_{\mathcal{S}}(\boldsymbol{x}) - \boldsymbol{z}') = \|\text{proj}_{\mathcal{S}}(\boldsymbol{x}) - \boldsymbol{z}'\|_2^2 \leq 0$$

which is possible only when $\boldsymbol{z}' = \text{proj}_{\mathcal{S}}(\boldsymbol{x})$. ■

References

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