CSE6243: Advanced Machine Learning **Fall 2023**

Lecture 3: Optimization: Convex Sets and Functions

Lecturer: Bo Dai *Scribes: Lu Li, Santusht Sairam*

3.1 Recap

The previous lecture focused on the following facets of convexity:

(i) The motivation behind convex optimization

(ii) The equivalence of local and global minima in convex spaces

(iii) The first-order characterization of convexity

To this end we formulated the problem of convex optimization as:

$$\min_{x \in \Omega} f(x) \tag{3.1}$$

wherein

$$\begin{cases} \Omega\text{: Convex Set} \\ f(\cdot) \in \Omega \to \mathbb{R}\text{ : Convex Function} \end{cases} \tag{3.2}$$

3.2 New Content

The present lecture discusses:

(i) The determination of convexity (for sets and functions)

(ii) The quantification of said convexity

3.2.1 Convex Sets

We shall now discuss the determination and quantification of convexity for sets, beginning with a reiteration of the definition of convex sets as:

Definition 3.1 (convex set)

$$\Omega \textit{ is convex} \Leftrightarrow \begin{array}{c} \forall x, y \in \Omega, t \in [0, 1], \\ tx + (1-t)y \in \Omega \end{array} \tag{3.3}$$

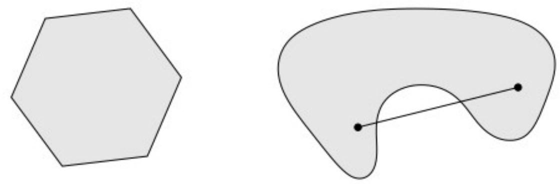

Figure 3.1: Example of convex (left) and non-convex (right) sets. Note the main visual distinction, wherein the line between any two points on a convex set also falls within its domain.

The definition above can be leveraged towards a proof of convexity for the following examples:

Example 1 (balls)

$$\begin{gathered}\{x: \ \|x\|_2 \leq \ c\}, \\ \text{where } \|x\|_2 = \sqrt{x^T x}\end{gathered} \tag{3.4}$$

Proof: Now let $x, y \in \Omega$, then:

$$\begin{gathered}||tx + (1-t)y|| \leq c \\ t||x|| + (1-t)||y|| \leq c\end{gathered} \tag{3.5}$$

■

Hence by definition Ω is a convex set. However, proofs of this nature from first principles can be rather involved, as in the case of affine planes:

Example 2 (affine plane)

$$\begin{gathered}\{x: \ Ax = b\}, \\ \text{where } A \in \mathbb{R}^{k \times d}, \ x \in \mathbb{R}^d, \ b \in \mathbb{R}^k\end{gathered} \tag{3.6}$$

Example 3 (simplex)

$$\{x \in \mathbb{R}^d | x \geq 0, \sum_{i=1}^{d} x_i = 1\} \tag{3.7}$$

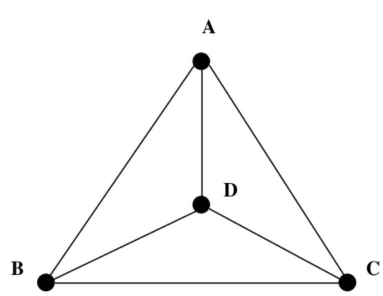

Figure 3.2: Example of a simplex in $\mathbb{R}^3$: $\{x \in \mathbb{R}^3 : \ x \geq 0, \sum_{i=1}^{d} x_i = 1\}$

To this end, the process of identification can be expedited through the consideration of **operations preserving convexity** given convex inputs:

(i) Affine transformations

$$A\Omega + b : \{y = Ax + b,\ x \in \Omega\} \tag{3.8}$$

(ii) Linear fractions

$$f(x) = \frac{A^2x + b}{C^2x + d} \quad \text{where } C^2x + d > 0 \tag{3.9}$$

In this case, $f(x)$ and $f^{-1}(x)$ both preserve convexity given $x \in \Omega$.

(iii) Conditional probability functions

$$\mathbb{P}(x = i|y = j) = \frac{\mathbb{P}(x = i, y = j)}{\mathbb{P}(y = j)} \tag{3.10}$$

Where the overarching joint probability function $\mathbb{P}(x = i|y = j) \in S$.

The full proof of this statement is omitted, but begins with the recognition of the linear fractional nature of $\mathbb{P}(x = i|y = j)$.

3.2.2 Convex Functions

Definition 3.2 *A function $f : X \to \mathbb{R}$ defined on a nonempty set $X \subset \mathbb{R}^n$ is called convex if*

(i) X is convex, and

(ii) $\forall x, y \in X, \forall \lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \tag{3.11}$$

A function $f : X \to \mathbb{R}$ is called concave if

$$-f : X \to \mathbb{R} \text{is a convex function.} \tag{3.12}$$

Below are some common **examples of convex functions**.

(i) linear form: $f(x) = Ax + b$

(ii) quadratic form: $f(x) = x^TQx$, note here $Q = A^TA$, Q is a symmetric positive definite matrix

(iii) norm of a vector: $f(x) = \|x\|_p = (\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}}$, it's easy to tell all norms are convex

(iv) operator norm of a matrix: $f(X) = \|X\|_{OP} = \sigma_i(X)$, σ_i is the maximum eigenvalue.

(v) trace norm: $f(X) = \sqrt{tr(X^TX)}$

Next we'll introduce some **operations preserving convexity of sets**, or allow us to construct convex sets from others.

(i) non-negative linear combinations:

$$g(x) = \sum_{i=1}^{k} w_i f_i(x), \quad \forall i, f_i(\cdot) \text{ is a convex function and } w_i \geq 0 \tag{3.13}$$

Note that this still holds when $k \to \infty$. In fact, when $k \to \infty$, the linear combinations become expectation: $\mathbb{E}_p(z)f(x,z) = g(x)$.

(ii) pointwise maximum

$$\begin{aligned} g(x) &= \max_{i\in(1,\ldots,k)} f_i(x) \\ g(x) &= \max_{z} f(x,z), \quad k \to \infty \end{aligned} \tag{3.14}$$

(iii) partial minimization: if $f(x,y)$ is convex and the function

$$g(x) = \inf_{y\in\Omega} f(x,y) \tag{3.15}$$

is proper, i.e., is $> -\infty$ everywhere and is finite at least at one point, then g is convex.

(iv) Affine substitutions: the superposition $f(Ax+b)$ of a convex function $f(\cdot)$ on $\mathbb{R}^n$ and affine mapping $x \to Ax+b$ is convex.

(v) General composition rule:

$$g(x) = f(h(x)) \tag{3.16}$$

$g(\cdot)$ is convex if:

a. $f(\cdot)$ is convex (non-decreasing) and $h(\cdot)$ is convex.
b. $f(\cdot)$ is convex (non-increasing) and $h(\cdot)$ is concave.

This can be validated using first-order and second-order conditions for optimality:

$$\begin{cases} g'(x) = f'(h(x)) \cdot h'(x) & \geq 0 \\ g''(x) = f''(h(x)) \cdot (h'(x))^2 + f'(h(x)) \cdot h''(x) & \geq 0 \end{cases} \tag{3.17}$$

Note that the general composition rule conditions are not sufficient and necessary conditions for convexity. Here's an example:

$$g(x) = \log \sum_{i=1}^{k} e^{a_i^T x + b_i} \tag{3.18}$$

3.2.3 Elementary property of convex functions: Jensen's Inequality

Here we'll introduce one of the most useful observations in the world: Jensen's inequality.

Definition 3.3 (Jensen's Inequality) *$x \in dom(f)$ with probability one, and f is convex, then we have:*

$$f(\mathbb{E}_p(x)[x]) \leq \mathbb{E}_p(x)[f(x)] \tag{3.19}$$

Proof: there're finite many elements in set x then the above inequality can be re-written as:

$$f(\sum_{i=1}^{N} \lambda_i x_i) \leq \sum_{i=1}^{N} \lambda_i f(x_i)$$

The points $(f(x_i), x_i)$ belong to the epigraph of f; since f is convex, its epigraph is a convex set, so that the convex combination

$$\sum_{i=1}^{N} \lambda_i (f(x_i), x_i) = (\sum_{i=1}^{N} \lambda_i f(x_i), \sum_{i=1}^{N} \lambda_i x_i) \tag{3.20}$$

of the points also belongs to the epigraph of f. By definition of the epigraph: Set $\lambda = 1$ in equation (3.10), the latter means exactly that $f(\sum_{i=1}^{N} \lambda_i x_i) \leq \sum_{i=1}^{N} \lambda_i f(x_i)$.

The number of points N can get as much as possible, when $N \to \infty$, it's equation (3.12) ■