Lecture-03: Review of convexity

1 Convexity

Let $\mathcal{X} \subseteq \mathbb{R}^d$ for $d \geqslant 1$ and $f : \mathcal{X} \to \mathbb{R}$ be a smooth function.

Definition 1.1 (Gradient). The gradient of function f at point $x \in \mathcal{X}$ is defined as the column vector $\nabla f(x) \in \mathbb{R}^d$, where the entry $i \in [d]$ is defined as $(\nabla f(x))_i \triangleq \frac{\partial f}{\partial x_i}(x)$.

Definition 1.2 (Hessian). The Hessian of function f at point $x \in \mathcal{X}$ is denoted by the matrix $\nabla^2 f(x) \in \mathbb{R}^{N\times N}$, where the entry $(i,j) \in [d] \times [d]$ is defined as $\nabla^2 f_{i,j}(x) \triangleq \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$.

Remark 1. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth function over d-dimensional reals. Then, we can write its Taylor series expansion around the neighborhood of $x \in \mathbb{R}^d$, in terms of the gradient vector $\nabla f(x) \in \mathbb{R}^d$ and the Hessian matrix $\nabla^2 f(x) \in \mathbb{R}^{N\times N}$, as

$$f(y) = f(x) + \langle \nabla f(x), (y-x) \rangle + \frac{1}{2} \left\langle (y-x), \nabla^2 f(x)(y-x) \right\rangle + o(\|y-x\|_2^2). \tag{1}$$

Definition 1.3 (Stationary Point). A point $x \in \mathcal{X}$ is called a stationary point of $f : \mathcal{X} \to \mathbb{R}$, if f attains a local extremum at x.

Remark 2. If $f : \mathcal{X} \to \mathbb{R}$ is smooth, then $\nabla f(x) = 0$ at a stationary point $x \in \mathcal{X}$.

1.1 Convexity

Definition 1.4 (Convex Set). A set $\mathcal{X}$ is called convex if for all $x, y \in \mathcal{X}$ and $\alpha \in [0,1]$, the convex combination $\alpha x + \bar{\alpha} y \in \mathcal{X}$ where $\bar{\alpha} \triangleq (1-\alpha)$.

Definition 1.5 (Convex Hull). A convex hull of a set A is the smallest convex set including A, i.e. $\text{conv}(A) \triangleq \left\{\sum_{x\in A} \alpha_x x : 0 \leqslant \alpha_x \leqslant 1, \sum_{x\in A} \alpha_x = 1\right\}$.

Definition 1.6. Let $\mathcal{X} \subseteq \mathbb{R}^d$. For a function $f : \mathcal{X} \to \mathbb{R}$, we define its epigraph as

$$\text{Epi}(f) \triangleq \{(x,y) \in \mathcal{X} \times \mathbb{R} : y \geqslant f(x)\}.$$

Definition 1.7. A function $f : \mathcal{X} \to \mathbb{R}$ is convex if the associated domain $\mathcal{X}$ and epigraph $\text{Epi}(f)$ are convex sets.

Theorem 1.8. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a convex set. Then the following are equivalent statements.*
(a) $f : \mathcal{X} \to \mathbb{R}$ is a convex function.
(b) For all $\alpha \in [0,1]$ and $x_1, x_2 \in \mathcal{X}$, we have $f(\alpha x_1 + (1-\alpha)x_2) \leqslant \alpha f(x_1) + (1-\alpha) f(x_2)$.
(c) For differentiable f, we have $f(x_2) - f(x_1) \geqslant \langle \nabla f(x_1), x_2 - x_1 \rangle$ for all $x_1, x_2 \in \mathcal{X}$.
(d) For twice differentiable f, we have $\nabla^2 f \succeq 0$, i.e. $\nabla^2 f$ is a positive semi-definite matrix.

Proof. For convex set $\mathcal{X} \subseteq \mathbb{R}^d$ and a function $f : \mathcal{X} \to \mathbb{R}$, we will show that statement a implies part (b), which implies part (c), which implies part (d), which implies part (a).

a $\implies$ b: Let $(x_1, f(x_2)), (x_2, f(x_2)) \in \text{Epi}(f)$ for $x_1, x_2 \in \mathcal{X}$. Let $\alpha \in [0,1]$ and $x_1, x_2 \in \mathcal{X}$, then from the convexity of $\mathcal{X}$, we have $\alpha x_1 + \bar{\alpha} x_2 \in \mathcal{X}$. Further from the convexity of Epi(f), we have $(\alpha x_1 + \bar{\alpha} x_2, \alpha f(x_1) + \bar{\alpha} f(x_2)) \in \text{Epi}(f)$. That is, $\alpha f(x_1) + \bar{\alpha} f(x_2) \geqslant f(\alpha x_1 + \bar{\alpha} x_2)$.

b $\implies$ c: Recall that $\alpha x_1 + \bar{\alpha} x_2 = x_1 + \bar{\alpha}(x_2 - x_1)$. From part (a), we have $f(x_2) - f(x_1) \geqslant \frac{f(\alpha x_1 + \bar{\alpha} x_2) - f(x_1)}{\bar{\alpha}}$. Taking $\bar{\alpha} \to 0$, we observe that the right hand side is equal to $\langle \nabla f(x_1), x_2 - x_1 \rangle$.

c $\implies$ d: Let $x_1, x_2 \in \mathcal{X}$. From (1) and part (b), it follows that $f(x_2) - f(x_1) - \langle \nabla f(x_1), x_2 - x_1 \rangle = \frac{1}{2}(x_2 - x_1)^T \nabla^2 f(x_1)(x_2 - x_1) + o(\|x_2 - x_1\|_2^2) \geqslant 0$.

d $\implies$ a: Let $\alpha \in [0,1]$ and $x_1, x_2 \in \mathcal{X}$. Then, it suffices to show that $\alpha f(x_1) + \bar{\alpha} f(x_2) \geqslant f(\alpha x_1 + \bar{\alpha} x_2)$. From the Taylor expansion of f in the neighborhood of x_2, we get

$$\alpha(f(x_1) - f(x_2)) = \alpha \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{\alpha}{2} \left\langle (x_1 - x_2), \nabla^2 f(x_2)(x_1 - x_2) \right\rangle + o(\|x_1 - x_2\|_2^2).$$

Similarly, we write the Taylor expansion of f in the neighborhood of x_2, to get

$$f(\alpha x_1 + \bar{\alpha} x_2) - f(x_2) = \alpha \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{\alpha^2}{2} \left\langle (x_1 - x_2), \nabla^2 f(x_2)(x_1 - x_2) \right\rangle + o(\|x_1 - x_2\|_2^2).$$

Taking the difference, we get $\alpha(f(x_1) - f(x_2)) \geqslant f(\alpha x_1 + \bar{\alpha} x_2) - f(x_2)$.

□

Example 1.9 (Convex Function). Following functions $f : \mathbb{R}^d \to \mathbb{R}$ are convex.
(a) Linear Function: $f(x) = \langle w, x \rangle$ for each $w \in \mathbb{R}^d$.
(b) Quadratic Function: $f(x) = x^T A x$ for a positive semi definite matrix $A \in \mathbb{R}^{N \times N}$.
(c) Abs Maximum: $f(x) = \max\{|x_i| : i \in [N]\} = \|x\|_\infty$.

Definition 1.10. We define a composition function $f \triangleq (h \circ g) : \mathcal{X} \to \mathcal{Z}$ for functions $h : \mathcal{Y} \to \mathcal{Z}$ and $g : \mathcal{X} \to \mathcal{Y}$ by defining $f(x) \triangleq h(g(x))$ for all $x \in \mathcal{X}$.

Lemma 1.11 (Composition of functions). *Let $h : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$. Then, the following statements are true for the composition function $f \triangleq h \circ g$.*
(a) If h is convex and nondecreasing and g is convex, then f is convex.
(b) If h is convex and nonincreasing and g is concave, then f is convex.
(c) If h is concave and nondecreasing and g is concave, then f is concave.
(d) If h is concave and nonincreasing and g is convex, then f is concave.

Proof. We will use the property that a function f is convex iff (i) $\text{dom}(f)$ is convex and (ii) $f(\alpha x_1 + \bar{\alpha} x_2) \leqslant \alpha f(x_1) + \bar{\alpha} f(x_2)$ for all $\alpha \in [0,1]$ and $x_1, x_2 \in \text{dom}(f)$.

Recall that $\mathbb{R}^d$ is convex for all $d \geqslant 1$. We will only show the first statement, and rest follow similar steps. Let $x_1, x_2 \in \mathbb{R}^d$ and $\alpha \in [0,1]$. From the convexity of g, we get $g(\alpha x_1 + \bar{\alpha} x_2) \leqslant \alpha g(x_1) + \bar{\alpha} g(x_2)$. From the nondecreasing property of h, we get $h(g(\alpha x_1 + \bar{\alpha} x_2)) \leqslant h(\alpha g(x_1) + \bar{\alpha} g(x_2))$. From the convexity of h, we get $h(\alpha g(x_1) + \bar{\alpha} g(x_2)) \leqslant \alpha h(g(x_1)) + \bar{\alpha} h(g(x_2))$. □

Theorem 1.12 (Jensen's Inequality). *Let $X : \Omega \to \mathcal{X} \subseteq \mathbb{R}^d$ be a random vector with finite marginal means, and $f : \mathcal{X} \to \mathbb{R}$ be a convex function. Then the mean $\mathbb{E}[X] \in \mathcal{X}$, the mean $\mathbb{E}[f(X)]$ is fnite, and $f(\mathbb{E}[X]) \leqslant \mathbb{E}[f(X)]$.*

Proof. We will show this for simple random vector $X : \Omega \to \{x_1, \dots, x_m\} \subseteq \mathcal{X}$, such that $\alpha_i \triangleq P\{X = x_i\}$ for all $i \in [m]$. Then, the mean $\mathbb{E}X = \sum_{i=1}^m \alpha_i x_i \in \mathcal{X}$ from the convexity of $\mathcal{X}$, and $\mathbb{E}f(X) = \sum_{i=1}^m \alpha_i f(x_i)$ is finite. Further, from the convexity of f, we get $f\left(\sum_{i=1}^m \alpha_i x_i\right) \leqslant \sum_{i=1}^m \alpha_i f(x_i)$. □

Corollary 1.13 (Young's inequality). *Let $p, q \geqslant 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then, $ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}$ for all $a, b \in \mathbb{R}_+$.*

Proof. Consider a positive random variable $X : \Omega \to \{a^p, b^q\}$ with probability mass function $P_X(a^p) = \frac{1}{p}$ and $P_X(b^q) = \frac{1}{q}$. Applying Jensen's inequality to the concave log function, we obtain

$$\ln\left(\frac{1}{p}a^p + \frac{1}{q}b^q\right) = \ln \mathbb{E}X \geqslant \mathbb{E}\ln X = \frac{1}{p}\ln a^p + \frac{1}{q}\ln b^q = \ln ab.$$

Since $\ln(\cdot)$ is an increasing function, the above inequality implies the result. □