DSCC 435 Optimization for Machine Learning	Lecture 2
Convexity and Complexity	
Lecturer: Jiaming Liang	*August 29, 2024*

1 Convexity

1.1 Convex set

Definition 1. *A set $S \subseteq \mathbb{R}^n$ is called convex if for any $x, y \in S$ and $\lambda \in [0,1]$ it holds that $\lambda x + (1-\lambda)y \in S$. For a set $S \subseteq \mathbb{R}^n$, the convex hull of S, denoted by $\mathrm{Conv}(S)$, is the intersection of all convex sets containing S.*

Clearly, $\mathrm{Conv}(S)$ is by itself a convex set, and it is the smallest convex set containing S (w.r.t. inclusion).

Affine sets can be defined similarly if we replace $\lambda \in [0,1]$ by $\lambda \in \mathbb{R}$. A hyperplane is a subset of $\mathbb{R}^n$ given by

$$H_{a,b} = \{x \in \mathbb{R}^n : \langle a, x \rangle = b\},$$

where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. It is an easy exercise to show that hyperplanes are affine sets.

Examples: Evidently, affine sets are always convex. Open and closed balls are always convex regardless of the choice of norm. For given $x, y \in \mathbb{R}^n$, the closed line segment between x and y is a subset of $\mathbb{R}^n$ denoted by $[x, y]$ and defined as

$$[x, y] = \{\alpha x + (1-\alpha)y : \alpha \in [0,1]\}$$

The open line segment (x, y) is similarly defined as

$$(x, y) = \{\alpha x + (1-\alpha)y : \alpha \in (0,1)\}$$

when $x \neq y$ and is the empty set $\emptyset$ when $x = y$. Closed and open line segments are convex sets. Another example of convex sets are half-spaces, which are sets of the form

$$H^-_{a,b} = \{x \in \mathbb{R}^n : \langle a, x \rangle \leq b\},$$

where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

1.2 Convex function

Definition 2. *A proper extended real-valued function f is convex if and only if*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \quad \textit{for all} \quad x, y \in \mathbb{R}^n, \lambda \in [0,1].$$

The epigraph of an extended real-valued function $f : \mathbb{R}^n \to [-\infty, \infty]$ is defined by $\mathrm{epi}(f) = \{(x, t) : f(x) \leq t, x \in \mathbb{R}^n, t \in \mathbb{R}\}$

Theorem 1. *An extended real-valued function $f : \mathbb{R}^n \to [-\infty, \infty]$ is convex if and only if* $\mathrm{epi}(f)$ *is a convex set.*

Proof. ($\Rightarrow$) Consider any two pairs (x_1, t_1) and (x_2, t_2) in $\mathrm{epi}(f)$, by definition, we know

$$f(x_1) \leq t_1, \quad f(x_2) \leq t_2.$$

If f is convex, then we know for every $\lambda \in [0, 1]$,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \leq \lambda t_1 + (1 - \lambda)t_2.$$

By the definition of epigraph again, we know $(\lambda x_1 + (1 - \lambda)x_2, \lambda t_1 + (1 - \lambda)t_2) \in \mathrm{epi}(f)$. Hence, $\mathrm{epi}(f)$ is convex.

($\Leftarrow$) Assume that $\mathrm{epi}(f)$ is convex, then for any $(x_1, t_1), (x_2, t_2) \in \mathrm{epi}(f)$ such that $f(x_1) = t_1$ and $f(x_2) = t_2$, we also have for every $\lambda \in [0, 1]$,

$$(\lambda x_1 + (1 - \lambda)x_2, \lambda t_1 + (1 - \lambda)t_2) \in \mathrm{epi}(f).$$

Therefore, for every $\lambda \in [0, 1]$,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda t_1 + (1 - \lambda)t_2 = \lambda f(x_1) + (1 - \lambda)f(x_2).$$

We prove that f is convex. □

Theorem 2. *A continuously differentiable function f is convex on $\mathbb{R}^n$ if and only if for any $x, y \in \mathbb{R}^n$, we have*

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle. \tag{1}$$

Proof. Let us consider a function $g : [0, 1] \mapsto \mathbb{R}$ defined as

$$g(t) = tf(x) + (1 - t)f(y) - f(y + t(x - y)).$$

We note that $g(0) = g(1) = 0$ and f being convex is equivalent to $g(t) \geq 0$ for every $t \in [0, 1]$. By calculation, we know

$$g'(t) = f(x) - f(y) - \langle \nabla f(y + t(x - y)), x - y \rangle$$

and (1) is equivalent to $g'(0) \geq 0$.

($\Rightarrow$) If f is convex, i.e., $g(t) \geq 0$ for every $t \in [0, 1]$, then

$$g'(0) = \lim_{t \to 0} \frac{g(t) - g(0)}{t} \geq 0.$$

($\Leftarrow$) If (1) holds, then we have

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle, \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

Summing the above inequalities, we have

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.$$

This is called the gradient monotone condition. Thus, for $0 < t_1 < t_2$, we have

$$\begin{aligned} g'(t_2) - g'(t_1) &= -\langle \nabla f(y + t_2(x-y)) - \nabla f(y + t_1(x-y)), x - y \rangle \\ &= -\frac{1}{t_2 - t_1} \langle \nabla f(y + t_2(x-y)) - \nabla f(y + t_1(x-y)), y + t_2(x-y) - y - t_1(x-y) \rangle \leq 0. \end{aligned}$$

Now, suppose that there exists $t \in (0,1)$ sucht that $g(t) < 0$. By the mean value theorem, we know there exists $0 < t_1 < t < t_2 < 1$ such that

$$g'(t_1) = \frac{g(t) - g(0)}{t} < 0, \quad g'(t_2) = \frac{g(1) - g(t)}{1 - t} > 0.$$

However, it contradicts with the monotonicity of g' derived previously. So, $g(t) \geq 0$ for every $t \in [0,1]$, i.e., f is convex. □

Theorem 3. *A twice continuously differentiable function f is convex on $\mathbb{R}^n$ if and only if for any $x \in \mathbb{R}^n$ we have*

$$\nabla^2 f(x) \succeq 0.$$

Proof. We first prove a lemma: A twice continuously differentiable function g is concave on $\mathbb{R}$ if and only if for any $t \in \mathbb{R}$ we have $g''(t) \leq 0$.

($\Rightarrow$) If g is concave in t, then by Taylor's expansion, we have

$$g(t+h) = g(t) + g'(t)h + \frac{1}{2} g''(s_1) h^2, \quad g(t-h) = g(t) - g'(t)h + \frac{1}{2} g''(s_2) h^2,$$

where $h > 0$, $s_1 \in (t, t+h)$ and $s_2 \in (t-h, t)$. Summing the two equalities, we have

$$\frac{1}{2}(g''(s_1) + g''(s_2)) = \frac{g(t+h) + g(t-h) - 2g(t)}{h^2}.$$

Hence, using g is concave, we have for every $t \in \mathbb{R}$,

$$g''(t) = \lim_{h \to 0} \frac{g(t+h) + g(t-h) - 2g(t)}{h^2} \leq 0.$$

($\Leftarrow$) Since $g''(t) \leq 0$ for every t, by Taylor's expansion, we have for every $t_1, t_2 \in \mathbb{R}$ and $\lambda \in [0,1]$,

$$\begin{aligned} g(t_1) &= g(\lambda t_1 + (1-\lambda)t_2) + g'(\lambda t_1 + (1-\lambda)t_2)(1-\lambda)(t_1 - t_2) + \frac{1}{2}g''(s_1)(1-\lambda)^2(t_1-t_2)^2 \\ &\leq g(\lambda t_1 + (1-\lambda)t_2) + g'(\lambda t_1 + (1-\lambda)t_2)(1-\lambda)(t_1 - t_2), \qquad (2) \\ g(t_2) &= g(\lambda t_1 + (1-\lambda)t_2) + g'(\lambda t_1 + (1-\lambda)t_2)\lambda(t_2 - t_1) + \frac{1}{2}g''(s_1)\lambda^2(t_2-t_1)^2 \\ &\leq g(\lambda t_1 + (1-\lambda)t_2) + g'(\lambda t_1 + (1-\lambda)t_2)\lambda(t_2 - t_1). \qquad (3) \end{aligned}$$

Multiplying (2) and (3) by λ and $1-\lambda$, respectively, and summing the resulting inequalities, we have

$$\lambda g(t_1) + (1-\lambda)g(t_2) \leq g(\lambda t_1 + (1-\lambda)t_2) \quad \forall t_1, t_2 \in \mathbb{R}, \lambda \in [0,1].$$

Therefore, we prove that g is concave.

Now, we are ready to prove the theorem. Consider again the function g on $[0,1]$,

$$g(t) = tf(x) + (1-t)f(y) - f(y + t(x-y)).$$

We have

$$g'(t) = f(x) - f(y) - \langle \nabla f(y + t(x-y)), x - y\rangle$$

and

$$g''(t) = -(x-y)^\top \nabla^2 f(y + t(x-y))(x-y). \qquad (4)$$

($\Rightarrow$) If f is convex, then g is concave in t. By the lemma, we know $g''(t) \leq 0$ for every $t \in [0,1]$. In view of (4), for any $x, y \in \mathbb{R}^n$, $g''(t) \leq 0$, so we must have

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in \mathbb{R}^n.$$

($\Leftarrow$) If $\nabla^2 f(x) \succeq 0$ for every $x \in \mathbb{R}^n$, then $g''(t) \leq 0$ for every $t \in [0,1]$. By the lemma, we know $g(t)$ is concave. Hence, we have for every $t \in [0,1]$,

$$0 = tg(1) + (1-t)g(0) \leq g(t),$$

which is equivalent to f being convex. $\square$

Example

- Every linear function $f(x) = \alpha + \langle a, x\rangle$ is convex.
- Let matrix A be symmetric and positive semidefinite. Then the quadratic function

  $$f(x) = \alpha + \langle a, x\rangle + \frac{1}{2}\langle Ax, x\rangle$$

  is convex (since $\nabla^2 f(x) = A \succeq 0$).

Operations preserving convexity

- $A : \mathbb{R}^n \to \mathbb{R}^m$, $b \in \mathbb{R}^m$, $f : \mathbb{R}^n \to (-\infty, \infty]$ is a convex function, then

$$g(x) = f(Ax + b)$$

  is convex.

- Let $f_1, f_2, \ldots, f_m : \mathbb{R}^n \to (-\infty, \infty]$ be extended real-valued convex functions, and let $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}_+$. Then the function $\sum_{i=1}^m \alpha_i f_i$ is convex.

- Let $f_i : \mathbb{R}^n \to (-\infty, \infty], i \in I$, be extended real-valued convex functions, where I is a given index set. Then the function

$$f(x) = \max_{i \in I} f_i(x)$$

  is convex.

Example Let $C \subseteq \mathbb{R}^n$ be a nonempty closed set, and consider the function

$$\varphi_C(x) = \frac{1}{2}\left(\|x\|^2 - d_C^2(x)\right),$$

where

$$d_C(x) = \min_{y \in C} \|x - y\|.$$

We want to show that $\varphi_C(x)$ is convex. Note that

$$d_C^2(x) = \min_{y \in C} \|x - y\|^2 = \|x\|^2 - \max_{y \in C}\left[2\langle y, x\rangle - \|y\|^2\right].$$

Hence,

$$\varphi_C(x) = \max_{y \in C}\left[\langle y, x\rangle - \frac{1}{2}\|y\|^2\right].$$

Therefore, since φ_C is a maximization of affine—and hence convex—functions, by the maximization rule above, it is necessarily convex. Also note that φ_C is convex regarless of whether C is convex or not.

2 Complexity

Definition 3. *We say $f(x) = \mathcal{O}(g(x))$ if there exists scalars $M > 0$ and $x_0 \in \mathbb{R}$ such that*

$$|f(x)| \leq Mg(x) \quad \textit{for all} \quad x \geq x_0.$$

We say $f(x) = \Omega(g(x))$ if there exists scalars $M > 0$ and $x_0 \in \mathbb{R}$ such that

$$f(x) \geq Mg(x) \quad \textit{for all} \quad x \geq x_0.$$

We say $f(x) = \Theta(g(x))$ if $f(x) = \mathcal{O}(g(x))$ and $f(x) = \Omega(g(x))$.

Convergence rate

Sublinear rate. This rate is described in terms of a power function of the iteration counter. For example, suppose that for some method we can prove the rate of convergence $r_k \le \frac{c}{\sqrt{k}}$. In this case, the upper complexity bound justified by this scheme for the corresponding problem class is $\left(\frac{c}{\epsilon}\right)^2$.

Example Gradient method:

$$f(x_k) - f_* \le \frac{L d_0^2}{k}$$

Accelerated gradient method:

$$f(x_k) - f_* \le \frac{L d_0^2}{k^2}$$

Subgradient method:

$$f(x_k) - f_* \le \frac{M d_0}{\sqrt{k}}$$

Linear rate. This rate is given in terms of an exponential function of the iteration counter. For example,

$$r_{k+1} \le (1-q) r_k$$

$$r_k \le c(1-q)^k \le c e^{-qk}, \quad 0 < q \le 1.$$

Note that the corresponding complexity bound is $\frac{1}{q}\left(\ln c + \ln \frac{1}{\epsilon}\right)$.

Eample Gradient method:

$$\mathcal{O}\left(\frac{L}{\mu} \log\left(\frac{\mu d_0^2}{\varepsilon}\right)\right)$$

Accelerated gradient method:

$$\mathcal{O}\left(\sqrt{\frac{L}{\mu}} \log\left(\frac{\mu d_0^2}{\varepsilon}\right)\right)$$

Quadratic rate. This rate has a double exponential dependence in the iteration counter. For example,

$$r_{k+1} \le c r_k^2$$

The corresponding complexity estimate depends on the double logarithm of the desired accuracy: $\ln \ln \frac{1}{\epsilon}$.

Example Suppose that the initial starting point x_0 is close enough to x^* :

$$\|x_0 - x^*\| \le \bar{r} = \frac{2\mu}{3M}.$$

Then $\|x_k - x^*\| \le \bar{r}$ for all k and the Newton's Method converges quadratically:

$$\|x_{k+1} - x^*\| \le \frac{M \|x_k - x^*\|^2}{2\left(\mu - M\|x_k - x^*\|\right)}.$$

Note that following the definition of $\bar{r}$ and $\|x_k - x^*\| \leq \bar{r}$, we have

$$\|x_{k+1} - x^*\| \leq \frac{M \|x_k - x^*\|^2}{2(\mu - M\bar{r})} = \frac{3M \|x_k - x^*\|^2}{2\mu} = \frac{\|x_k - x^*\|^2}{\bar{r}}.$$