Convex Optimization

윤 정 훈

optimization problem

- Mathematical optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(\mathbb{x}) \\ \text{subject to} & f_i(\mathbb{x}) \le b_i \ , \ i = 1, \dots, m \end{array}$$

○ optimization variable $\mathbb{x} = (x_1, \dots, x_n)$

○ objective function $f_0 : \mathbb{R}^n \to \mathbb{R}$

○ (inequality) constraint functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$

○ optimal or solution $\mathbb{x}^\star$

① 제약조건을 만족시키는 벡터 $\mathbb{x}$중에서 objective의 값을 가장 작게 하는 벡터

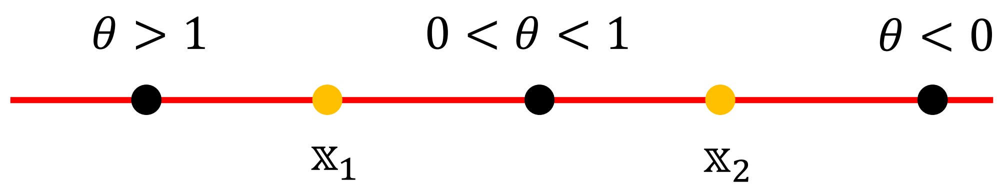
② for any $\mathbb{x} \in \mathbb{R}^n$ with $f_1(\mathbb{x}) \le b_1$, … , $f_m(\mathbb{x}) \le b_m$, we have $f_0(\mathbb{x}^\star) \le f_0(\mathbb{x})$

line

- line
  - $\mathbb{x}_1 \neq \mathbb{x}_2$ and $\mathbb{x}_1, \mathbb{x}_2 \in \mathbb{R}^n$
  - $\theta \in \mathbb{R}$

$$\mathbb{y} = \theta \mathbb{x}_1 + (1-\theta)\mathbb{x}_2$$

  - $\mathbb{y}$는 $\mathbb{x}_1$과 $\mathbb{x}_2$사이를 지나는 line을 만든다.



affine

- Affine
  - $C \subseteq \mathbb{R}^n$ and C is affine,

    ① If the line through any two distinct points in C lies in C.

    ② (수학적 표현) For any $\mathbb{x}_1, \mathbb{x}_2 \in C$ and $\theta \in \mathbb{R} \Rightarrow \theta\mathbb{x}_1 + (1-\theta)\mathbb{x}_2 \in C$
  - If C is an affine set and $\mathbb{x}_1, \ldots, \mathbb{x}_k \in C$ and $\theta_1 + \cdots + \theta_k = 1$, then $\theta_1\mathbb{x}_1 + \cdots + \theta_k\mathbb{x}_k \in C$

- Affine hull
  - The set of all affine combinations of points in some set $C \in \mathbb{R}^n$

$$\textbf{aff}\ C = \{\underbrace{\theta_1\mathbb{x}_1 + \cdots + \theta_k\mathbb{x}_k}_{\text{affine combination}} \mid \mathbb{x}_1, \ldots, \mathbb{x}_k \in C, \theta_1 + \cdots + \theta_k = 1\}$$

convex (set)

- Convex
  - $C \subseteq \mathbb{R}^n$ and C is convex,

    ① If the line segment between any two points in C lies in C.

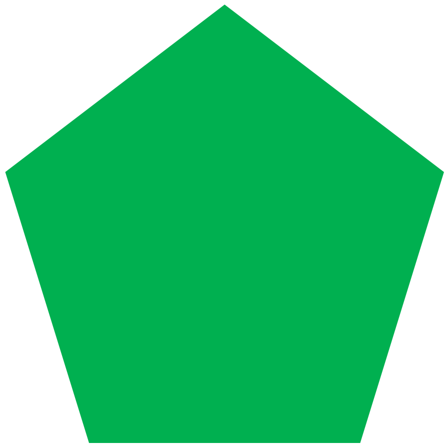
    ② (수학적) For any $\mathbb{x}_1, \mathbb{x}_2 \in C$ and any θ with $0 \leq \theta \leq 1 \Rightarrow \theta\mathbb{x}_1 + (1-\theta)\mathbb{x}_2 \in C$

    ③ if and only if C contains every convex combination of its points

- Convex hull
  - the set of all convex combinations of points in C $(i = 1, \dots, k)$

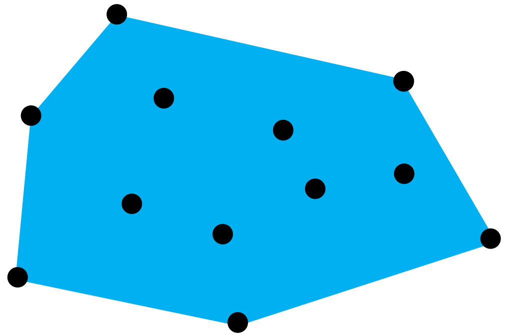
$$\mathbf{conv}\, C = \{\underbrace{\theta_1\mathbb{x}_1 + \cdots + \theta_k\mathbb{x}_k}_{\text{convex combination}} | \mathbb{x}_i \in C, \theta_i \geq 0, \theta_1 + \cdots + \theta_k = 1\}$$

example



convex set

non convex set



12개 points의 convex hull

위의 non convex set의 convex hull

affine (function)

- Affine function
  - a function composed of a linear function and constant (translation)
  - in 1-dim

$$y = Ax + C$$

  - in 2-dim

$$f(x, y) = Ax + By + C$$

  - in 3-dim

$$f(x, y, z) = Ax + By + Cz + D$$

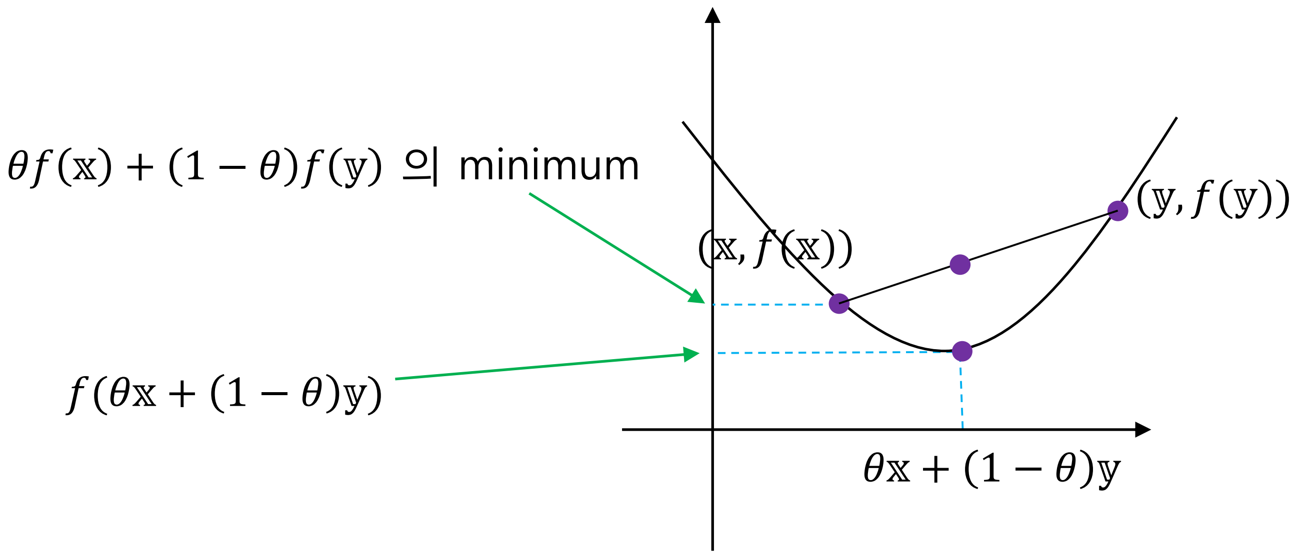
translation : a transformation consisting of a constant offset with no rotation or distortion

convex (function)

- $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if $\mathbf{dom}\, f$ is convex and

$$f(\theta \mathbb{x} + (1-\theta)\mathbb{y}) \leq \theta f(\mathbb{x}) + (1-\theta) f(\mathbb{y})$$

for $\forall \mathbb{x}, \mathbb{y} \in \mathbf{dom}\, f$, $0 \leq \theta \leq 1$.



    - f is concave if $-f$ is convex.

$\mathbf{dom}\, f$: 함수 f의 유효한 입력 값의 집합. 이 영역을 f가 define되는 영역이라고도 표현 한다.

Optimization problem in standard form

- optimization problem

$$
\begin{array}{ll}
\text{minimize} & f_0(\mathbb{x}) \\
\text{subject to} & f_i(\mathbb{x}) \le 0 \ , \ i = 1, \dots, m \\
 & h_i(\mathbb{x}) = 0 \ , \ i = 1, \dots, p
\end{array}
$$

  - ○ $\mathbb{x} \in \mathbb{R}^n$ is the optimization variable.
  - ○ $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the objective function or cost function.
  - ○ $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, are inequality constraint functions.
  - ○ $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, p$, are equality constraint functions.
  - ○ optimal value $p^\star$

$$p^\star = \inf\{f_o(\mathbb{x}) \mid f_i(\mathbb{x}) \le 0, i = 1, \dots, m, h_i(\mathbb{x}) = 0, i = 1, \dots, p\}$$

Optimal and locally optimal points

- optimal point

  - $\mathcal{D} = (\bigcap_{i=1}^{m} \textbf{dom}\, f_i) \cap (\bigcap_{i=1}^{p} \textbf{dom}\, h_i)$ is domain of the optimization problem.

  - A point $\mathbb{x} \in \mathcal{D}$ is feasible, if it satisfies $f_i(\mathbb{x}) \leq 0$, $i = 1, \dots, m$ and $h_i(\mathbb{x}) = 0$, $i = 1, \dots, p$.

  - A feasible $\mathbb{x}$ is optimal if $f_0(\mathbb{x}) = p^\star$ ($p^\star$ is the optimal value.)

  - optimal set X_{opt}

$$X_{opt} = \{\mathbb{x} \mid f_i(\mathbb{x}) \leq 0, i = 1, \dots, m, h_i(\mathbb{x}) = 0, i = 1, \dots, p, f_0(\mathbb{x}) = p^\star\}$$

- locally optimal point

  - A feasible $\mathbb{x}$ is locally optimal if there is an $R > 0$ such that

$$f_0(\mathbb{x}) = \inf\{f_0(\mathbb{z}) \mid f_i(\mathbb{z}) \leq 0, i = 1, \dots, m, h_i(\mathbb{z}) = 0, i = 1, \dots, p, \|\mathbb{z} - \mathbb{x}\|_2 \leq R\}$$

Convex optimization problem in standard form

- convex optimization problem

$$
\begin{array}{ll}
\text{minimize} & f_0(\mathbb{x}) \\
\text{subject to} & f_i(\mathbb{x}) \le 0 \ , \ i = 1, \dots, m \\
 & \mathbb{a}_i^T \mathbb{x} = b_i \ , \ i = 1, \dots, p
\end{array}
$$

  - ○ $f_0, \dots, f_m$ are convex.
  - ○ $h_i(\mathbb{x}) = \mathbb{a}_i^T \mathbb{x} - b_i \ , \ i = 1, \dots, p$ are affine.
  - ○ The feasible set of a convex optimization problem is convex.
    - ① Since $f_1, \dots, f_m$ are convex, $\cap_{i=1}^{m} \mathbf{dom}\, f_i$ is convex.
    - ② Since $\{\mathbb{x} \mid \mathbb{a}_i \mathbb{x} = b_i\}$ is hyperplane, $\cap_{i=1}^{p} \{\mathbb{x} \mid \mathbb{a}_{\boldsymbol{i}} \mathbb{x} = b_i\}$ is affine(i.e convex).
  - ○ In a convex optimization, we minimize a convex objective function over a convex set.

Local and global optima

- Any locally optimal point of convex optimization problem is globally optimal.
  - 증명 아이디어 : local optimal value보다 더 작은 값을 가지는 point가 존재한다고 가정할 때, local optimal의 feasible 영역에 local optimal보다 더 작은 값이 존재하는 것을 보인다.

proof)

Recall, $\mathbb{x}$ is local optimal, then there is an $R > 0$ such that

$$f_0(\mathbb{x}) = \inf\{f_0(\mathbb{w}) \mid \mathbb{w} \text{ is feasible}, \|\mathbb{w} - \mathbb{x}\|_2 \leq R\}$$

Suppose $\mathbb{x}$ is not globally optimal, then there is a feasible $\mathbb{y}$ with $f_0(\mathbb{y}) < f_0(\mathbb{x})$ and $\|\mathbb{y} - \mathbb{x}\| > R$.

Consider $\mathbb{z} = (1 - \theta)\mathbb{x} + \theta\mathbb{y}$ with $\theta = \frac{R}{2\|\mathbb{y}-\mathbb{x}\|_2}$.

Then, $\|\mathbb{z} - \mathbb{x}\|_2 = \frac{R}{2} < R$ and $\theta < \frac{1}{2}$ and $\mathbb{z}$ is feasible. ($\because$ convexity)

So, $f_0(\mathbb{z}) = (1 - \theta)f_0(\mathbb{x}) + \theta f_0(\mathbb{y}) < f_0(\mathbb{x})$

It contradicts $\mathbb{x}$ is local optimal.

Hence, $\mathbb{x}$ is globally optimal.

Lagrangian

- optimization problem in standard form (not necessarily convex)

$$\begin{array}{ll} \text{minimize} & f_0(\mathbb{x}) \\ \text{subject to} & f_i(\mathbb{x}) \le 0 \ , \ i = 1, \dots, m \\ & h_i(\mathbb{x}) = 0 \ , \ i = 1, \dots, p \end{array}$$

  - variable $\mathbb{x} \in \mathbb{R}^n$, domain $\mathcal{D} = (\cap_{i=1}^{m} \mathbf{dom}\, f_i) \cap \left(\cap_{i=1}^{p} \mathbf{dom}\, h_i\right)$, optimal value $p^\star$

- Lagrangian $\mathcal{L}: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\mathbf{dom}\, \mathcal{L} = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$

$$\mathcal{L}(\mathbb{x}, \Lambda, V) = f_o(\mathbb{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbb{x}) + \sum_{i=1}^{p} v_i h_i(\mathbb{x})$$

  - weighted sum of objective and constraint functions
  - λ_i is Lagrange multiplier associated with $f_i(\mathbb{x}) \le 0$ and $\Lambda = (\lambda_1, \dots, \lambda_m)$
  - v_i is Lagrange multiplier associated with $h_i(\mathbb{x}) = 0$ and $V = (v_1, \dots, v_p)$

Lagrange dual Function

- Lagrange dual function $g: \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$g(\Lambda, V) = \inf_{\mathbb{x} \in \mathcal{D}} \mathcal{L}(\mathbb{x}, \Lambda, V) = \inf_{\mathbb{x} \in \mathcal{D}} \left(f_0(\mathbb{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbb{x}) + \sum_{i=1}^{p} v_i h_i(\mathbb{x}) \right)$$

  - g is concave. ($\because$ g is pointwise infimum of a family of affine functions of (Λ, V)) (?)
  - $\mathbb{x}^\star$가 g를 최소화 시키는가?

- Lower bounds on optimal value
  - If $\Lambda \succcurlyeq 0$, then $g(\Lambda, V) \leq p^\star$
  - dual feasible is a pair of (Λ, V) with $\Lambda \succcurlyeq 0$ and $(\Lambda, V) \in \mathbf{dom}\, g$

Lagrange dual problem

- Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\Lambda, V) \\ \text{subject to} & \Lambda \succcurlyeq 0 \end{array}$$

  - Find lower bound on $p^\star$, obtained from Largrange dual function.
  - Λ, V are dual feasible if $\Lambda \succcurlyeq 0$ and $(\Lambda, V) \in \textbf{dom}\ g$
  - A convex optimization problem : objective function to be maximized is concave and constraint is convex.

    ① Concave maximization problem is readily solved by minimizing the convex objective function $-f_0$.
  - $d^\star$is the optimal value of the Lagrange dual function.

Weak and strong duality

- Weak duality : $d^\star \leq p^\star$
  - always holds (Whether original problem is convex or not.)
  - can be used to find a lower bound on the optimal value of a problem that is difficult to solve.
  - duality gap is $p^\star - d^\star$.
- Strong duality : $d^\star = p^\star$
  - (usually) holds for convex problem
    - ① objective function is convex.
    - ② inequality constraint is convex.
    - ③ equality constraint is affine.

Complementary slackness

- Suppose that strong duality holds, $\mathbb{x}^\star$ is primal optimal, $(\Lambda^\star, V^\star)$ is dual optimal.

$$
\begin{aligned}
f_0(\mathbb{x}^\star) &= g(\Lambda^\star, V^\star) \\
&= \inf_{\mathbb{x}\in\mathcal{D}} \left(f_0(\mathbb{x}) + \sum_{i=1}^{m} \lambda_i^\star f_i(\mathbb{x}) + \sum_{i=1}^{p} v_i^\star h_i(\mathbb{x}) \right) \\
&\le \left(f_0(\mathbb{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(\mathbb{x}^\star) + \sum_{i=1}^{p} v_i^\star h_i(\mathbb{x}^\star) \right) \\
&\le f_0(\mathbb{x}^\star)
\end{aligned}
$$

모두 "="

★ ○ $\mathbb{x}^\star$ minimizes $\mathcal{L}(\mathbb{x}, \Lambda^\star, V^\star)$

★ ○ complementary slackness : $\lambda_i^\star f_i(\mathbb{x}^\star) = 0$, $i = 1, \dots, m$

① $\lambda_i^\star > 0 \Longrightarrow f_i(\mathbb{x}^\star) = 0$

② $f_i(\mathbb{x}^\star) < 0 \Longrightarrow \lambda_i^\star = 0$

Karush-Kuhn-Tucker (KKT) condition

non-convex

- $\mathbb{x}^\star$ is a primal optimal, $(\Lambda^\star, V^\star)$ is a dual optimal, $d^\star = p^\star$(strong duality)
  - In strong duality, $\mathbb{x}^\star$ minimize $\mathcal{L}(\mathbb{x}, \Lambda^\star, V^\star)$

- Then, the KKT conditions must be satisfied. (f_i, h_i : differentiable)
  - primal constraints
    ① $f_i(\mathbb{x}^\star) \leq 0,\ i = 1, \ldots, m$
    ② $h_i(\mathbb{x}^\star) = 0,\ i = 1, \ldots, p$
  - dual constaints
    ① $\lambda_i^\star \geq 0,\ i = 1, \ldots, m$
  - complementary slackness
    ① $\lambda_i^\star f_i(\mathbb{x}^\star) = 0,\ i = 1, \ldots, m$
  - gradient of Lagrangian with respect to $\mathbb{x}$ vanishes at $\mathbb{x}^\star$
    ① $\nabla f_0(\mathbb{x}^*) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\mathbb{x}^\star) + \sum_{i=1}^{p} v_i^\star \nabla h_i(\mathbb{x}^\star) = 0$

Karush-Kuhn-Tucker (KKT) condition

non-convex

- For any optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions.

$$\mathbb{x}^\star$$

$$(\Lambda^\star, V^\star) \longrightarrow \text{KKT conditions are satisfied.}$$

$$d^\star = p^\star$$

KKT conditions for convex problems

- If f_i are convex and h_i are affine and $\mathbb{x}'$, Λ', V' are any points that satisfy the KKT conditions, then $\mathbb{x}'$ is primal optimal and (Λ', V') is dual optimal with zero duality gap.

$$f_i(\mathbb{x}') \leq 0,\ i = 1, \dots, m$$

$$h_i(\mathbb{x}') = 0,\ i = 1, \dots, p$$

$$\lambda_i' \geq 0,\ i = 1, \dots, m$$

$$\lambda_i' f_i(\mathbb{x}') = 0,\ i = 1, \dots, m$$

$$\nabla f_0(\mathbb{x}') + \sum_{i=1}^{m} \lambda_i' \nabla f_i(\mathbb{x}') + \sum_{i=1}^{p} v_i' \nabla h_i(\mathbb{x}') = 0$$

$\Rightarrow$

$\mathbb{x}'$ is primal optimal

(Λ', V') is dual optimal

zero duality gap

Since $\mathcal{L}(\mathbb{x}, \Lambda', V')$ is convex, $g(\Lambda', V') = \inf_{\mathbb{x} \in \mathbb{R}^n} \mathcal{L}(\mathbb{x}', \Lambda', V') = \mathcal{L}(\mathbb{x}', \Lambda', V')$

KKT conditions for convex problems

convex

$\mathbb{x}^\star$

$(\Lambda^\star, V^\star)$

$d^\star = p^\star$

$\Longleftrightarrow$

KKT conditions are satisfied.