IE 521 Convex Optimization

Lecture 9: Optimality Conditions, Saddle Point

Niao He

16th April 2019

Outline

Recap: Convex Theorem on Alternative

Theorem. Let X be convex and $f, g_1, \ldots, g_m$ be convex. Assume $g_1, ..., g_m$ satisfy the relaxed Slater condition:

$$\exists \bar{x} \in \mathrm{rint}(X), \text{s.t. } g_i(\bar{x}) < 0, \forall \text{ non-affine } g_i.$$

Exactly one of the following two systems must be empty:

(I) $\{x \in X : f(x) < 0, g_i(x) \leq 0, i = 1, ..., m\}$

(II) $\{\lambda \in \mathbb{R}^m : \lambda \geq 0, \inf_{x \in X}\{f(x) + \sum_{i=1}^m \lambda_i g_i(x)\} \geq 0\}$

Recap: Lagrange Duality

General convex program:

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, ..., m \end{aligned} \tag{P}$$

Lagrange dual program:

$$\begin{aligned} \max_{\lambda} \quad & \underline{L}(\lambda) := \inf_{x \in X} L(x, \lambda) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned} \tag{D}$$

where $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$.

Convex Duality Theorem

- (Weak duality) $\text{Opt}(D) \leq \text{Opt}(P)$
- (Strong duality) If (P) is convex, solvable, and satisfies relaxed Slater condition, then $\text{Opt}(D) = \text{Opt}(P)$.

Recap: Detecting Optimality

Let (P) be convex and satisfies relaxed Slater condition.
Let x^* be a feasible solution.

x^* is optimal to (P)

$\Leftrightarrow$ $\{x \in X : f(x) < f(x^*), g_i(x) \leq 0, \forall i\}$ is **infeasible**

$\Leftrightarrow$ $\{\lambda : \lambda \geq 0, \inf_{x \in X} \{f(x) + \sum \lambda_i g_i(x)\} \geq f(x^*)\}$ is **feasible**

$$\Leftrightarrow \exists \lambda^* \geq 0, \text{s.t.} \inf_{x \in X} \left\{ f(x) + \sum \lambda_i^* g_i(x) \right\} = f(x^*).$$

$$\Leftrightarrow \exists \lambda^* \geq 0, \text{s.t.} \begin{cases} \sum \lambda_i^* g_i(x^*) = 0, \textit{i.e.},\ \lambda_i^* g_i(x^*) = 0, \forall i \\ x^* \in \operatorname{argmin}_{x \in X} \{f(x) + \sum \lambda_i^* g_i(x)\} \end{cases}$$

$$\Leftrightarrow \exists \lambda^* \geq 0, \text{s.t.} \begin{cases} f(x^*) + \sum \lambda_i g_i(x^*) \leq f(x^*) + \sum \lambda_i^* g_i(x^*), \forall \lambda \geq 0 \\ f(x^*) + \sum \lambda_i^* g_i(x^*) \leq f(x) + \sum \lambda_i^* g_i(x), \forall x \in X \end{cases}$$

$$\Leftrightarrow \exists \lambda^* \geq 0, \text{s.t.} \begin{cases} L(x^*, \lambda) \leq L(x^*, \lambda^*), \forall \lambda \geq 0 \\ L(x^*, \lambda^*) \leq L(x, \lambda^*), \forall x \in X \end{cases}$$

Optimality Condition

Theorem. Let f be convex and differentiable. Then

$$x^* \in \operatorname*{argmin}_x f(x) \iff \nabla f(x^*) = 0.$$

Theorem. Let f be convex and differentiable, $X \subset \mathrm{dom}(f)$ be a convex set. Then

$$\begin{aligned} x^* \in \operatorname*{argmin}_{x \in X} f(x) &\iff \langle \nabla f(x^*), x - x^* \rangle \geq 0, \forall x \in X. \\ &\iff \nabla f(x^*) \in N_X(x^*). \end{aligned}$$

Definition. The *normal cone* of X at x is defined as the set

$$N_X(x) = \left\{ h \in \mathbb{R}^n : h^T(y - x) \geq 0, \forall y \in X \right\}.$$

Normal Cone

$$N_X(x) = \left\{ h \in \mathbb{R}^n : h^T(y - x) \geq 0, \forall y \in X \right\}$$

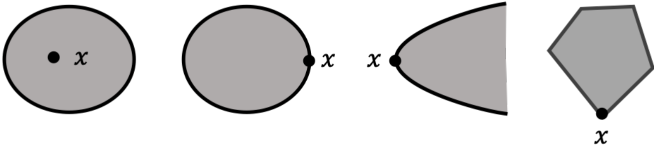


Figure: Find the normal cones

Normal Cone

$$N_X(x) = \left\{h \in \mathbb{R}^n : h^T(y - x) \geq 0, \forall y \in X\right\}$$

- $N_X(x)$ is always a closed convex cone.
- If $x \in \text{int}(X)$, $N_X(x) = \{0\}$.
- If $X = \left\{x : a_i^T x \geq b_i, i = 1, ..., m\right\}$, $x \notin \text{int}(X)$,

$$N_X(x) = \text{Cone}(\{a_i | a_i^T x = b_i\}).$$

Remark. Let $f(x)$ be convex and differentiable.

- If $x^* \in \text{int}(X)$, then x^* is optimal iff $\nabla f(x^*) = 0$.
- If $X = \mathbb{R}^n$, then x^* is optimal iff $\nabla f(x^*) = 0$.

Optimality Conditions for Convex Programs

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, ..., m \end{aligned} \qquad (P)$$

Theorem. Let (P) be a convex program and let x^* be feasible. Assume $f, g_1, ..., g_m$ are differentiable at x^*. Then

$$x_* \in X \text{ is optimal for (P)}$$

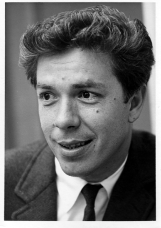
$$\overset{\text{(slater)}}{\Longleftrightarrow} \exists \lambda^* \geq 0, \text{ s.t.} \begin{cases} \nabla f(x_*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x_*) \in N_X(x_*) \\ \lambda_i^* g_i(x_*) = 0, \forall i = 1, \ldots, m \end{cases}$$

Karush-Kuhn-Tucker (KKT) Conditions (1951)

(x^*, λ^*) is an optimal primal-dual pair if it satisfies:

- *Primal feasibility:* $x^* \in X, g_i(x^*) \leq 0$
- *Dual feasibility:* $\lambda^* \geq 0$
- *Lagrange optimality:* $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) \in N_X(x^*)$
- *Complementary slackness:* $\lambda_i^* g_i(x^*) = 0, \forall i = 1, ..., m$

Albert W. Tucker
(1905-1995)



Harold W. Kuhn
(1925-2014)

Example

Given $a_i > 0, i = 1, ..., n$, solve the problem

$$\begin{aligned} \min_{x>0} \quad & \sum_{i=1}^{n} \frac{a_i}{x_i} \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i \leq 1 \end{aligned}$$

- The Lagrange function

$$L(x, \lambda) = \sum_{i=1}^{n} \frac{a_i}{x_i} + \lambda\Big(\sum_{i=1}^{n} x_i - 1\Big).$$

- The KKT optimality conditions yield

$$\begin{cases} x_i^* > 0, \sum_{i=1}^n x_i^* \leq 1 \\ \lambda^* \geq 0 \\ -\frac{a_i}{(x_i^*)^2} + \lambda^* = 0 \\ \lambda^*(\sum_{i=1}^n x_i^* - 1) = 0 \end{cases} \Rightarrow \begin{cases} \lambda^* = (\sum_{i=1}^n \sqrt{a_i})^2 \\ x_i^* = \frac{\sqrt{a_i}}{\sum_{i=1}^n \sqrt{a_i}}, i = 1, \ldots, n \end{cases}$$

Remarks

For general nonlinear programs,

- The KKT conditions are only the necessary (first-order) conditions for a solution to be local optimal, provided that some regularity conditions (e.g., linear independence constraint qualification) are satisfied.

- The KKT conditions are not sufficient for global or even local optimality.

Minimax Problems

Recall that

$$L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

$$\underline{L}(\lambda) := \inf_{x\in X} L(x,\lambda)$$

$$\overline{L}(\lambda) := \sup_{\lambda\geq 0} L(x,\lambda)$$

Observe that

$$\min_{x\in X, g_i(x)\leq 0, \forall i} f(x) = \min_{x\in X} \overline{L}(x) =: \min_{x\in X} \max_{\lambda\geq 0} L(x,\lambda) \qquad (P)$$

$$\max_{\lambda\geq 0} \underline{L}(x) := \max_{\lambda\geq 0} \min_{x\in X} L(x,\lambda) \qquad (D)$$

Saddle Point

Let $x^* \in X, \lambda^* \geq 0$.

Definition. (x^*, λ^*) is a saddle point of $L(x, \lambda)$ if

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*), \forall x \in X, \lambda \geq 0$$

i.e.,

$$\sup_{\lambda \geq 0} L(x^*, \lambda) = L(x^*, \lambda^*) = \inf_{x \in X} L(x, \lambda^*)$$

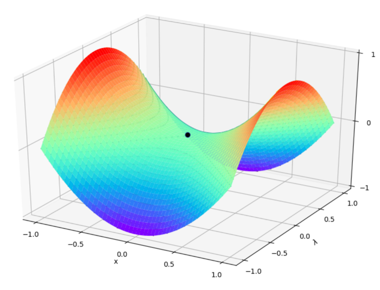


Figure: Saddle point

Saddle Point vs Optimality

$$\min_{x\in X} \overline{L}(x) =: \min_{x\in X} \max_{\lambda\geq 0} L(x,\lambda) \qquad (P)$$

$$\max_{\lambda\geq 0} \underline{L}(\lambda) := \max_{\lambda\geq 0} \min_{x\in X} L(x,\lambda) \qquad (D)$$

Theorem. (x^*, λ^*) is a saddle point of $L(x,\lambda)$ if and only if x^* is an optimal solution to (P), λ^* is an optimal solution to (D) and $\mathrm{Opt}(P) = \mathrm{Opt}(D)$.

Remark.

- Apply to any saddle function $L(x,\lambda)$, not limited to the Lagrange function.
- Saddle point exists for the Lagrange function of a solvable convex program satisfying the Slater condition.

Proof of Theorem

- (if part) By optimality of (x^*, λ^*):

$$\mathrm{Opt}(P) = \overline{L}(x^*) = \sup_{\lambda\geq 0} L(x^*, \lambda) \geq L(x^*, \lambda^*)$$
$$\mathrm{Opt}(D) = \underline{L}(\lambda^*) = \inf_{x\in X} L(x, \lambda^*) \leq L(x^*, \lambda^*)$$

$$\mathrm{Opt}(D) = \mathrm{Opt}(P) \Rightarrow \sup_{\lambda\geq 0} L(x^*, \lambda) = L(x^*, \lambda^*) = \inf_{x\in X} L(x^*, \lambda)$$

- (only if part) Assume (x^*, λ^*) is a saddle point,

$$L(x, \lambda^*) \geq L(x^*, \lambda^*) \geq L(x^*, \lambda), \forall x \in X, \lambda \geq 0$$

$$\mathrm{Opt}(P) = \inf_{x\in X} \overline{L}(x) \leq \overline{L}(x^*) = \sup_{\lambda\geq 0} L(x^*, \lambda) = L(x^*, \lambda^*)$$
$$\mathrm{Opt}(D) = \sup_{\lambda\geq 0} \underline{L}(\lambda) \geq \underline{L}(\lambda^*) = \inf_{x\in X} L(x, \lambda^*) = L(x^*, \lambda^*)$$

Hence, $\mathrm{Opt}(P) \leq L(x^*, \lambda^*) \leq \mathrm{Opt}(D)$
Combined with weak duality, we have $\mathrm{Opt}(P) = \mathrm{Opt}(D)$.
Hence, $\mathrm{Opt}(P) = \overline{L}(x^*) = L(x^*, \lambda^*) = \underline{L}(\lambda^*) = \mathrm{Opt}(D)$.
Thus, x^* solves (P), λ^* solves (D), and $\mathrm{Opt}(P) = \mathrm{Opt}(D)$.

Saddle Point vs Optimality

Corollary. Let (P) be a convex program and $x^* \in X$. Then

$$x_* \in X \text{ is optimal for (P)}$$
$$\xleftarrow[]{\text{(slater)}}\!\!\Longrightarrow \exists \lambda^* \geq 0, \text{ s.t. } (x^*, \lambda^*) \text{ is a saddle point of } L(x, \lambda)$$

Remark.

- For general non-convex problems, saddle point may not always exist.
- The existence of saddle point is far from being necessary for primal optimality.

General Saddle Point Problem

Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$, $L(x, y) : X \times Y \to \mathbb{R}$.

Definition. $(x^*, y^*) \in X \times Y$ is a saddle point of $L(x, y)$ if

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*), \forall x \in X, y \in Y$$

Two Induced Problems:

$$(P): \quad \min_{x \in X} \max_{y \in Y} L(x, y) := \min_{x \in X} \overline{L}(x)$$

$$(D): \quad \max_{y \in Y} \min_{x \in X} L(x, y) := \max_{y \in Y} \underline{L}(y)$$

Proposition. It holds true that

$$\sup_{y \in Y} \inf_{x \in X} L(x, y) \leq \inf_{x \in X} \sup_{y \in Y} L(x, y).$$

If a saddle point exists, then

$$\sup_{y \in Y} \inf_{x \in X} L(x, y) = \inf_{x \in X} \sup_{y \in Y} L(x, y).$$

Example

Consider the following problem

$$L(x,y) = (x-y)^2, X = [0,1], Y = [0,1].$$

Does there exist a saddle point?

- $\bar{L}(x) = \sup_{y\in[0,1]} L(x,y) = \max\left\{x^2, (x-1)^2\right\}$

$$\inf_{x\in X} \sup_{y\in Y} L(x,y) = \frac{1}{4}$$

- $\underline{L}(y) = \inf_{x\in[0,1]} L(x,y) = 0$

$$\sup_{y\in Y} \inf_{x\in X} L(x,y) = 0$$

Minimax Lemma

Lemma Let $f_i(x), i = 1, ..., m$ be convex and continuous on a convex compact set X. Then

$$\min_{x \in X} \max_{1 \le i \le m} f_i(x) = \min_{x \in X} \sum_{i=1}^{m} \lambda_i^* f_i(x)$$

for some $\lambda^* \in \Delta_m := \{\lambda \in \mathbb{R}^m : \lambda \ge 0, \sum_{i=1}^{m} \lambda_i = 1\}$.

- The above result implies that

$$\max_{\lambda \in \Delta_m} \min_{x \in X} \sum \lambda_i f_i(x) = \min_{x \in X} \max_{\lambda \in \Delta_m} \sum \lambda_i f_i(x).$$

von Neumann's Minimax Theorem (1928)

Theorem. Assume

- X and Y be convex and compact,
- $L(x, y) : X \times Y \to \mathbb{R}$ is continuous, **convex-concave**, i.e., convex in $x \in X$ for fixed $y \in Y$ and concave in $y \in Y$ for fixed $x \in X$.

Then $L(x, y)$ has a saddle point on $X \times Y$, and

$$\min_{x \in X} \max_{y \in Y} L(x, y) = \max_{y \in Y} \min_{x \in X} L(x, y)$$

Remark. The compactness is sufficient but not necessary:

(i) $\min_x \max_y (x + y) = \infty \neq -\infty = \max_y \min_x (x + y)$

(ii) $\min_x \max_{0 \leq y \leq 1} (x + y) = -\infty = \max_{0 \leq y \leq 1} \min_x (x + y)$

(iii) $\min_x \max_{y \leq 1} (x + y) = -\infty = \max_{y \leq 1} \min_x (x + y)$

Sion's Minimax Theorem (1958)

Theorem. Assume

- X and Y be convex, and at least one of them is compact,
- $L(x, y) : X \times Y \to \mathbb{R}$ is **lower semi-continuous** and **quasi-convex** on $x \in X$,
- $L(x, y) : X \times Y \to \mathbb{R}$ is **upper semi-continuous** and **quasi-concave** on $y \in Y$.

Then $L(x, y)$ has a saddle point on $X \times Y$, and

$$\min_{x \in X} \max_{y \in Y} L(x, y) = \max_{y \in Y} \min_{x \in X} L(x, y)$$

References

- Ben-Tal & Nemirovski, Chapter 3.2-3.4