IE 521 Convex Optimization

Lecture 8: Convex Program and Duality

Niao He

16th April 2019

Outline

Convex Program

The standard form of an optimization problem is

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, ..., m \qquad (P) \\ & h_j(x) = 0, j = 1, ..., k \end{aligned}$$

Definition. An optimization problem (P) is *convex* if

1. the objective function f is convex.
2. the inequality constraint functions $g_1, ..., g_m$ are convex.
3. there is either no equality constraint or only linear equality constraint.

Feasibility and Optimality

Definition. The feasible set of (P) is

$$C = \{x \in \mathrm{dom}(f) : g_i(x) \leq 0, \forall i, h_j(x) = 0, \forall j\}.$$

The optimal value of (P) is

$$p^* = \inf\{f(x) : g_i(x) \leq 0, \forall i, h_j(x) = 0, \forall j\}.$$

- (P) is *feasible* if $C \neq \emptyset$.
- (P) is *infeasible* if $C = \emptyset$ and we set $p^* = +\infty$.
- (P) is *unbounded below* is $p^* = -\infty$.
- (P) is *solvable* if $\exists$ a feasible solution $x^* \in C$, s.t. $p^* = f(x^*)$. We call such x^*, an optimal solution.
- (P) is *unattainable* if $|p^*| < \infty$ but $\nexists$ a feasible solution $x^* \in C$, s.t. $p^* = f(x^*)$.

Example

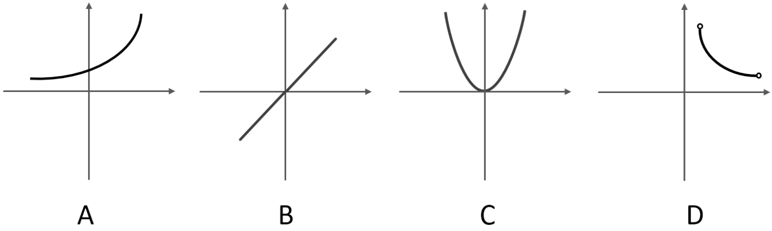

Figure: Convex functions

Local vs Global Optimum

Definition. Let $x^* \in C$ be a feasible solution.

- x^* is a global optimum for (P) if
$$f(x^*) \leq f(x), \forall x \in C.$$
- x^* is a local optimum for (P) if
$$\exists r > 0, \text{ s.t. } f(x^*) \leq f(x), \forall x \in B(x^*, r) \cap C.$$

Proposition. For convex programs, a local optimum is always a global optimum.

Proof. Let x^* be a local optimum and $z = \epsilon x^* + (1-\epsilon)x$. Then $z \in C \cap B(x^*, r)$ for small enough ϵ.

$$f(x^*) \leq f(z) \leq \epsilon f(x^*) + (1-\epsilon) f(x) \Rightarrow f(x^*) \leq f(x), \forall x \in C.$$

Epigraph Form

The standard problem (P) is equivalent as the problem

$$\begin{aligned} \min_{x\in\mathbb{R}^n, t\in\mathbb{R}} \quad & t \\ s.t. \quad & f(x) - t \leq 0 \qquad\qquad (P') \\ & g_i(x) \leq 0, i = 1, ..., m \\ & h_j(x) = 0, j =, 1, ..., k \end{aligned}$$

Note that

1. (P') is still convex if (P) is convex
2. (x^*, t^*) is optimal to (P') if and only if x^* is optimal to (P) and $t^* = f(x^*)$

Example: Piecewise Linear Minimization

- Consider the convex problem

$$\begin{aligned} \min_{x} \quad & f(x) := \max_{j=1,\ldots,m} (a_1^T x + b_1, \ldots, a_m^T x + b_m) \\ \text{s.t.} \quad & Cx = d. \end{aligned}$$

- This is equivalent to the following linear program

$$\begin{aligned} \min_{x,t} \quad & t \\ \text{s.t.} \quad & a_i^T x + b_i - t \leq 0, i = 1, \ldots, m \\ & Cx = d. \end{aligned}$$

Example: Robust Linear Program

- Consider the robust linear program

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & a_i^T x \leq b_i, \forall a_i \in \mathcal{E}_i, i = 1, \ldots, m \end{aligned}$$

where $\mathcal{E}_i = \{\bar{a}_i + P_i u : \|u\|_2 \leq 1\}$.

- This is equivalent to the convex program

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & \bar{a}_i^T x + \|P_i^T x\|_2 \leq b_i, i = 1, \ldots, m. \end{aligned}$$

General Form of Convex Program

We will focus on the general form of convex program

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

where

- $X \subseteq \text{dom}(f) \cap (\cap_{i=1}^m \text{dom}(g_i))$ is convex;
- $f, g_1, \dots, g_m$ are convex.

Q. How to verify whether a solution x^* is optimal?

Detecting Optimality for Linear Programs

$$(P) \quad \begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad\qquad (D) \quad \begin{aligned} \max \quad & b^T y \\ \text{s.t.} \quad & A^T y \leq c \end{aligned}$$

- We know that

$$\begin{aligned} x^* \text{ is optimal } \Leftrightarrow & Ax^* = b, x \geq 0 \text{ (primal feasibility)} \\ & \exists y^*, A^T y^* \leq c \text{ (dual feasibility)} \\ & c^T x^* = b^T y^* \text{ (zero duality gap)} \end{aligned}$$

Theorem on Alternative

Farkas' Lemma Let $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$. Exactly one of the following sets must be empty:

(i) $\{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$;

(ii) $\left\{y \in \mathbb{R}^m : A^T y \leq 0, b^T y > 0\right\}$.

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 Slater condition:

$$\exists \bar{x} \in X, \text{s.t. } g_i(\bar{x}) < 0, \forall i = \{1, ..., m\}.$$

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

Proof

(I) feasible $\Rightarrow$ (II) infeasible. (easy to check)
(I) infeasible $\Rightarrow$ (II) feasible.

- Denote $u = (u_0, u_1, ..., u_m)$ and consider the two sets
$$S = \left\{u \in \mathbb{R}^{m+1} : \exists x \in X, f(x) \le u_0, g_i(x) \le u_i, i = 1, \dots, m\right\}$$
$$T = \left\{u \in \mathbb{R}^{m+1} : u_0 < 0, u_1 \le 0, ..., u_m \le 0\right\}$$
- Note that S, T are convex, nonempty, and $S \cap T = \emptyset$.
- By separation theorem, $\exists a = (a_0, a_1, ..., a_m) \neq 0$ that
$$\sup_{u_0<0, u_i \le 0, \forall i} \sum_{i=0}^m a_i u_i \le \inf_{x \in X, u: u_0 \ge f(x), u_i \ge g_i(x), \forall i} \sum_{i=0}^m a_i u_i$$
- Observe $a \ge 0$, hence:
$$0 \le \inf_{x \in X} \{a_0 f(x) + a_1 g_1(x) + ... + a_m g_m(x)\}.$$
- Note that $a_0 > 0$, otherwise, by Slater condition
$$\inf_{x \in X} \{a_1 g_1(x) + ... + a_m g_m(x)\} < 0, \text{ contradiction!}$$
- Setting $\lambda_i = \frac{a_i}{a_0}, i = 1, ..., m$, we obtain a solution to (II).

Relaxed Slater Condition

Slater condition: $\exists \bar{x} \in X, \text{s.t. } g_i(\bar{x}) < 0, \forall i = \{1, ..., m\}$.

Relaxed Slater condition: $\exists x \in \text{rint}(X)$ s.t. $g_i(x) < 0$ for all $i = \{1, ..., m\}$ such that $g_i(x)$ is not affine.

- Informally speaking, the feasible region must have a relative interior point.

Example . Does Slater condition hold true?

$$X = \{(x_1, x_2) : x_2 > 0\} \text{ and } g(x_1, x_2) = \frac{x_1^2}{x_2}.$$

Detecting Optimality of Convex Program

Consider the general convex program

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

- X is convex, $f, g_1, \ldots, g_m$ are convex
- Slater condition holds.

Q. How to verify whether a solution x^* is optimal?

x^* is optimal to (P)

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

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

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

Lagrange Dual

Definition. The Lagrange function:

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

and the Lagrange dual function:

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

where $\lambda = (\lambda_1, \ldots, \lambda_m)$ is called the Lagrange multiplier.

Definition. The *Lagrange dual* of the problem (P) is

$$\begin{aligned} \max_{\lambda} \quad & \underline{L}(\lambda) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned} \qquad (D)$$

Duality Theorem

Theorem. Denote $\text{Opt}(P)$ and $\text{Opt}(D)$ as the optimal values to (P) and (D), we have

(a) **(Weak Duality)** $\forall \lambda \geq 0, \underline{L}(\lambda) \leq \text{Opt}(P)$. Moreover,

$$\text{Opt}(D) \leq \text{Opt}(P).$$

(b) **(Strong Duality)** If (P) is convex and below bounded, and satisfies the relaxed Slater condition, then (D) is solvable, and

$$\text{Opt}(D) = \text{Opt}(P)$$

Illustration: Linear Program Duality

$$(P)\quad \begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad\qquad (D)\quad \begin{aligned} \max \quad & b^T y \\ \text{s.t.} \quad & A^T y \leq c \end{aligned}$$

$$\Downarrow \qquad\qquad\qquad\qquad \Uparrow$$

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax - b \leq 0 \quad (\lambda_1) \\ & b - Ax \leq 0 \quad (\lambda_2) \\ & x \geq 0 \end{aligned} \qquad \Rightarrow \quad \begin{aligned} \max \quad & b^T(\lambda_2 - \lambda_1) \\ \text{s.t.} \quad & c + A^T(\lambda_1 - \lambda_2) \geq 0 \\ & \lambda \geq 0 \end{aligned}$$

The Lagrange dual function is

$$\begin{aligned} \underline{L}(\lambda) &= \inf_{x \geq 0} (c + A^T\lambda_1 - A^T\lambda_2)^T x + b^T(\lambda_2 - \lambda_1) \\ &= \begin{cases} b^T(\lambda_2 - \lambda_1), & c + A^T\lambda_1 - A^T\lambda_2 \geq 0 \\ -\infty, & o.w. \end{cases} \end{aligned}$$

Illustration: Quadratic Program Duality

Let $Q \succ 0$.

$$(P) \quad \begin{aligned} \min_x \quad & \frac{1}{2}x^T Qx + q^T x \\ \text{s.t.} \quad & Ax \geq b \end{aligned}$$

$$(D) \quad \begin{aligned} \max_{y,\lambda} \quad & -\frac{1}{2}y^T Qy + b^T \lambda \\ \text{s.t.} \quad & A^T \lambda - Qy = q \\ & \lambda \geq 0 \end{aligned}$$

The Lagrange dual function is

$$\begin{aligned} \underline{L}(\lambda) &= \inf_x \{\frac{1}{2}x^T Qx + (q - A^T \lambda)x + b^T \lambda\} \\ &= -\frac{1}{2}(A^T \lambda - q)^T Q^{-1}(A^T \lambda - q) + b^T \lambda \end{aligned}$$

More Examples of Lagrange Dual

Example . Consider the problem

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \sum_{i=1}^n f_i(x_i) \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = 1. \end{aligned}$$

The Lagrange dual is

$$\min_{\lambda} \quad \lambda + \sum_{i=1}^n f_i^*(-\lambda).$$

Representation Affects Duality

$$\begin{aligned} \min_{x_1, x_2} \quad & e^{-x_2} \\ \text{s.t.} \quad & \|x\|_2 \leq x_1, \\ & x_2 \geq 0. \end{aligned}$$

- Find the feasible set and optimal value.
- Represent the problem with

$$g(x) = \|x\|_2 - x_1,\ X = \{(x_1, x_2) : x_2 \geq 0\}.$$

  Does the slater condition hold? Is there a duality gap?
- Now represent the problem with

$$g(x) = -x_2\ X = \{(x_1, x_2) : \|x\|_2 \leq x_1\}.$$

  Does the slater condition hold? Is there a duality gap?

Remarks on Nonconvex Problems

- Even for general nonconvex problems, the dual problem is always convex.
- Weak duality always holds, i.e., the optimal value to the dual problem always provides a lower bound to the optimal value to the primal problem.

Example .

$$\begin{aligned} \min \quad & x^T W x \\ \text{s.t.} \quad & x_i^2 = 1, i = 1, \ldots, n \end{aligned}$$

The Lagrange dual problem is given by

$$\begin{aligned} \max_{\lambda} \quad & -1^T \lambda \\ \text{s.t.} \quad & W + \text{diag}(\lambda) \succeq 0 \end{aligned}$$

References

- Ben-Tal & Nemirovski, Chapter 3.1-3.3