Outline

Basics of Convex Program

Convex Theorem on Alternatives

Lagrange Duality
Convex Program

The standard form of an optimization problem is

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad g_i(x) \leq 0, \ i = 1, \ldots, m \quad (P) \\
& \quad h_j(x) = 0, \ j = 1, \ldots, k
\end{align*}
\]

Definition. An optimization problem \((P)\) is \textit{convex} if

1. the objective function \(f\) is convex.
2. the inequality constraint functions \(g_1, \ldots, g_m\) are convex.
3. there is either no equality constraint or only linear equality constraint.
Feasibility and Optimality

Definition. The feasible set of \((P)\) is

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

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 \textit{feasible}\ if \(C \neq \emptyset\).
- \((P)\) is \textit{infeasible}\ if \(C = \emptyset\) and we set \(p^* = +\infty\).
- \((P)\) is \textit{unbounded below}\ is \(p^* = -\infty\).
- \((P)\) is \textit{solvable}\ if \(\exists\) a feasible solution \(x^* \in C\), s.t. \(p^* = f(x^*)\). We call such \(x^*\), an optimal solution.
- \((P)\) is \textit{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 \( \epsilon \).

\[
 f(x^*) \leq f(z) \leq \epsilon f(x^*) + (1 - \epsilon)f(x) \implies f(x^*) \leq f(x), \forall x \in C.
\]
Epigraph Form

The standard problem $(P)$ is equivalent as the problem

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

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

\[
\min_x f(x) := \max_{j=1,...,m} (a_1^T x + b_1, ..., a_m^T x + b_m)
\]

s.t. \( Cx = d \).

This is equivalent to the following linear program

\[
\min_{x,t} t
\]

s.t. \( a_i^T x + b_i - t \leq 0, i = 1, ..., m \)

\( Cx = d \).
Example: Robust Linear Program

Consider the robust linear program

$$\min_{x} \ c^T x$$
$$\text{s.t.} \ \bar{a}_i^T x \leq b_i, \forall a_i \in \mathcal{E}_i, i = 1, \ldots, m$$

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

This is equivalent to the convex program

$$\min_{x} \ c^T x$$
$$\text{s.t.} \ \bar{a}_i^T x + \|P_i^T x\|_2 \leq b_i, i = 1, \ldots, m.$$
General Form of Convex Program

We will focus on the general form of convex program

$$\min_{x \in X} f(x)$$

s.t. \( g_i(x) \leq 0, \ i = 1, \ldots, m \)  

where

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

Q. How to verify whether a solution \( x^* \) is optimal?
Detecting Optimality for Linear Programs

\[
\begin{align*}
\min_{(P)} & \quad c^T x \\
\text{s.t.} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]

\[
\begin{align*}
\max_{(D)} & \quad b^T y \\
\text{s.t.} & \quad A^T y \leq c
\end{align*}
\]

We know that

- \(x^*\) is optimal \(\iff Ax^* = b, x \geq 0\) (primal feasibility)
- \(\exists y^*, A^T y^* \leq c\) (dual feasibility)
- \(c^T x^* = b^T y^*\) (zero duality gap)
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) $\{y \in \mathbb{R}^m : A^T y \leq 0, b^T y > 0\}$. 
Convex Theorem on Alternative

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

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

Exactly one of the following two systems must be empty:

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

(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 ⇒ (II) infeasible. (easy to check)

(I) infeasible ⇒ (II) feasible.

▶ Denote \( u = (u_0, u_1, ..., u_m) \) and consider the two sets

\[
S = \{ u \in \mathbb{R}^{m+1} : \exists x \in X, f(x) \leq u_0, g_i(x) \leq u_i, i = 1, \ldots, m \} \\
T = \{ u \in \mathbb{R}^{m+1} : u_0 < 0, u_1 \leq 0, \ldots, u_m \leq 0 \}
\]

▶ Note that \( S, T \) are convex, nonempty, and \( S \cap T = \emptyset \).

▶ By separation theorem, \( \exists a = (a_0, a_1, \ldots, a_m) \neq 0 \) that

\[
\sup_{u_0 < 0, u_i \leq 0, \forall i} \sum_{i=0}^{m} a_i u_i \leq \inf_{x \in X, u_0 \geq f(x), u_i \geq g_i(x), \forall i} \sum_{i=0}^{m} a_i u_i
\]

▶ Observe \( a \geq 0 \), hence:

\[
0 \leq \inf_{x \in X} \{ a_0 f(x) + a_1 g_1(x) + \ldots + a_m g_m(x) \}.
\]

▶ Note that \( a_0 > 0 \), otherwise, by Slater condition

\[
\inf_{x \in X} \{ a_1 g_1(x) + \ldots + a_m g_m(x) \} < 0, \text{ contradiction!}
\]

▶ Setting \( \lambda_i = \frac{a_i}{a_0}, i = 1, \ldots, m \), we obtain a solution to (II).
Relaxed Slater Condition

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

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

- 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

$$\min_{x \in X} f(x)$$

s.t. $g_i(x) \leq 0, \; i = 1, \ldots, m$

- $X$ 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:

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

and the Lagrange dual function:

\[ L(\lambda) = \inf_{x \in X} L(x, \lambda) \]

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

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

\[
\max_{\lambda} \quad L(\lambda) \\
\text{s.t.} \quad \lambda \geq 0
\]  

\((D)\)
Duality Theorem

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

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

\[ \begin{align*}
\min \quad & c^T x \\
(P) \quad \text{s.t.} \quad & Ax = b \\
& x \geq 0
\end{align*} \]

\[ \downarrow \]

\[ \begin{align*}
\min \quad & c^T x \\
\text{s.t.} \quad & Ax - b \leq 0 \quad (\lambda_1) \\
& b - Ax \leq 0 \quad (\lambda_2) \\
& x \geq 0
\end{align*} \]

\[ \Rightarrow \]

\[ \begin{align*}
\max \quad & b^T \lambda_2 - \lambda_1 \\
(D) \quad \text{s.t.} \quad & c + A^T (\lambda_1 - \lambda_2) \geq 0 \\
& \lambda \geq 0
\end{align*} \]

The Lagrange dual function is

\[ L(\lambda) = \inf_{x \geq 0} (c + A^T \lambda_1 - A^T \lambda_2)^T x + b^T (\lambda_2 - \lambda_1) \]

\[ = \begin{cases} 
  b^T (\lambda_2 - \lambda_1), & c + A^T \lambda_1 - A^T \lambda_2 \geq 0 \\
  -\infty, & \text{o.w.}
\end{cases} \]
Illustration: Quadratic Program Duality

Let $Q > 0$.

\[
\begin{align*}
\min_x & \quad \frac{1}{2}x^T Q x + q^T x \\
\text{s.t.} & \quad Ax \geq b
\end{align*}
\]

\[\text{(P)}\]

\[
\begin{align*}
\max_{y,\lambda} & \quad -\frac{1}{2}y^T Q y + b^T \lambda \\
\text{s.t.} & \quad A^T \lambda - Q y = q \\
& \quad \lambda \geq 0
\end{align*}
\]

\[\text{(D)}\]

The Lagrange dual function is

\[
L(\lambda) = \inf_x \left\{ \frac{1}{2}x^T Q x + (q - A^T \lambda)x + b^T \lambda \right\}
\]

\[
= -\frac{1}{2} (A^T \lambda - q)^T Q^{-1} (A^T \lambda - q) + b^T \lambda
\]
Example. Consider the problem

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

The Lagrange dual is

\[
\min_{\lambda} \quad \lambda + \sum_{i=1}^{n} f_i^*(-\lambda).
\]
Representation Affects Duality

\[
\min_{x_1, x_2} e^{-x_2} \\
\text{s.t.} \quad \|x\|_2 \leq x_1, \quad x_2 \geq 0.
\]

- Find the feasible set and optimal value.
- Represent the problem with
  \[
g(x) = \|x\|_2 - x_1, \quad X = \{(x_1, x_2) : x_2 \geq 0\}.
\]
  Does the slater condition holds? Is there a duality gap?
- Now represent the problem with
  \[
g(x) = -x_2 \quad X = \{(x_1, x_2) : \|x\|_2 \leq x_1\}.
\]
  Does the slater condition holds? 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.

\[
\begin{align*}
\min & \quad x^T W x \\
\text{s.t.} & \quad x_i^2 = 1, \; i = 1, \ldots, n
\end{align*}
\]

The Lagrange dual problem is given by

\[
\begin{align*}
\max_{\lambda} & \quad -1^T \lambda \\
\text{s.t.} & \quad W + \text{diag}(\lambda) \succeq 0
\end{align*}
\]
References

- Ben-Tal & Nemirovski, Chapter 3.1-3.3