In this lecture, we cover the following topics

- Illustrations of Lagrange Duality
- Saddle Point Formulation
- Optimality Conditions (KKT conditions)

References: Bental & Nemirovski Chapter 3.2

9.1 Recall

- Convex program:

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

where \( f(x), g_1, \ldots, g_m \) are convex and \( X \) is convex.

- Lagrange function:

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

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

- Lagrange dual function:

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

- Note that \( \forall \lambda \geq 0, L(\lambda) \leq \text{Opt}(P) \)

- Lagrange dual program:

\[
\max_{\lambda \geq 0} \quad L(\lambda)
\]

We show that \( \text{Opt}(D) = \text{Opt}(P) \) when (relaxed) Slater condition holds
9.2 Illustrations of Lagrange Duality

- Linear Program Duality

\[
\begin{align*}
\text{min} & \quad c^T x \\
\text{s.t.} & \quad Ax = b \\
\text{max} & \quad b^T y \\
\text{s.t.} & \quad A^T y \leq c \\
\end{align*}
\]

First, rewrite the original problem as:

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

Introducing the multipliers \( \lambda = (\lambda_1, \lambda_2) \geq 0 \), the Lagrange function is

\[
L(x, \lambda) = c^T x + \lambda_1^T (Ax - b) + \lambda_2^T (b - Ax) = (c + A^T \lambda_1 - A^T \lambda_2)^T x + b^T (\lambda_2 - \lambda_1)
\]

The Lagrange dual function is

\[
L(\lambda) = \inf_{x \in \mathbb{R}^n} (c + A^T \lambda_1 - A^T \lambda_2)^T x + b^T (\lambda_2 - \lambda_1) = \begin{cases} 
  b^T (\lambda_2 - \lambda_1), & c + A^T \lambda_1 - A^T \lambda_2 \geq 0 \\
  -\infty, & \text{o.w.}
\end{cases}
\]

The Lagrange dual is \( \max_{\lambda \geq 0} L(\lambda) \), which is equivalent to

\[
\max_{\lambda \geq 0} \quad b^T (\lambda_2 - \lambda_1) \\
\quad c + A^T (\lambda_1 - \lambda_2) \geq 0
\]

Substituting \( y = \lambda_2 - \lambda_1 \), the formulation above is also equivalent to:

\[
\max_{y} \quad b^T y \\
\text{s.t.} \quad A^T y \leq c
\]

- Quadratic Program Duality:

\[
\begin{align*}
\text{min} & \quad \frac{1}{2} x^T Q x + q^T x \\
\text{s.t.} & \quad Ax \leq b \\
& \quad x \geq 0
\end{align*}
\]

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

where \( Q \succ 0 \).
The Lagrange function is \( L(x, \lambda) = \frac{1}{2} x^T Q x + q^T x + \lambda^T (b - Ax) \).

The Lagrange dual function is \( \bar{L}(\lambda) = \inf_x \{ \frac{1}{2} x^T Q x + (q - A^T x) + b^T\lambda \} \).

The infimum is achieved when \( Q x = A^T \lambda - q, x = Q^{-1} (A^T \lambda - q) \).

The Lagrange dual is

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

In both cases discussed above, strong duality holds true.

### 9.3 Saddle Point Formulation

Recall

\[
(P): \quad \min_{x \in X} \{ f(x) : g_i(x) \geq 0, i = 1, \ldots, m \} = \min_{x \in X} \bar{L}(x) = \min_{x \in X} \max_{\lambda \geq 0} L(x, \lambda)
\]

\[
(D): \quad \max_{\lambda \geq 0} \bar{L}(\lambda) = \max_{\lambda \geq 0} \min_{x \in X} L(x, \lambda)
\]

**Definition 9.1** (*Saddle point*) We call \((x^*, \lambda^*)\), where \(x^* \in X, \lambda^* \geq 0\), a saddle point of \(L(x, \lambda)\) if \(L(x, \lambda^*) \geq L(x^*, \lambda^*) \geq L(x^*, \lambda), \forall x \in X, \lambda \geq 0\).

**Theorem 9.2** \((x^*, \lambda^*)\) is saddle point of \(L(x, \lambda)\) if and only if \(x^*\) is an optimal solution \((P)\), \(\lambda^*\) is an optimal solution to \((D)\) and \(\text{Opt}(P) = \text{Opt}(D)\).

**Proof:**

\((\Rightarrow)\) Assume \((x^*, \lambda^*)\) is a saddle point,

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

\[
\text{Opt}(P) = \inf_{x \in X} \bar{L}(x) \leq \bar{L}(x^*) = \sup_{\lambda \geq 0} L(x^*, \lambda) = L(x^*, \lambda^*)
\]

\[
\text{Opt}(D) = \sup_{\lambda \geq 0} L(\lambda) \geq L(\lambda^*) = \inf_{x \in X} L(x, \lambda^*) = L(x^*, \lambda^*)
\]

Hence, \(\text{Opt}(P) \leq L(x^*, \lambda^*) \leq \text{Opt}(D)\)

Combined with weak duality, we have \(\text{Opt}(P) = \text{Opt}(D)\). Hence, \(\text{Opt}(P) = \bar{L}(x^*) = L(x^*, \lambda^*) = L(\lambda^*) = \text{Opt}(D)\). Thus, \(x^*\) solves \((P)\), \(\lambda^*\) solves \((D)\), and \(\text{Opt}(P) = \text{Opt}(D)\)
Assume \((x^*, \lambda^*)\) are optimal solutions to \((P)\) and \((D)\), and \(\text{Opt}(P) = \text{Opt}(D)\). By optimality,

\[
\text{Opt}(P) = \bar{L}(x^*) = \sup_{\lambda \geq 0} L(x^*, \lambda) \geq L(x^*, \lambda^*)
\]

\[
\text{Opt}(D) = L(\lambda^*) = \inf_{x \in X} L(x, \lambda^*) \leq L(x^*, \lambda^*)
\]

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

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

i.e. \((x^*, \lambda^*)\) is a saddle point of \(L(x, \lambda)\).

**Remark:**

- The above theorem holds true for any saddle function \(L(x, \lambda)\) and its induced primal and dual problems, not limited to the Lagrange function.
- Saddle point always exists for the Lagrange function of a solvable convex program satisfying the Slater condition. More generally, the existence of saddle points is guaranteed for convex-concave saddle functions over convex compact domains (Minimax Theorem).

### 9.4 Optimality Conditions

**Theorem 9.3:** Let \(x^* \in X\)

(i) *(sufficient condition)* If there exists \(\lambda^* \geq 0\), such that \((x^*, \lambda^*)\) is a saddle point of \(L(x, \lambda)\), then \(x^*\) is an optimal solution to \((P)\).

(ii) *(necessary condition)* Assume \((P)\) is convex and satisfies the Slater condition. If \(x^*\) is an optimal solution to \((P)\) then \(\exists \lambda^* \geq 0\), s.t. \((x^*, \lambda^*)\) is a saddle point of \(L(x, \lambda)\).

**Proof:**

(i) *(sufficient part)* Follows from previous theorem.

(ii) *(necessary part)* By strong duality theorem, \(\exists\) optimal dual solution \(\lambda^* \geq 0\), such that \(\text{Opt}(P) = \text{Opt}(D)\). Hence, following from the previous theorem, \((x^*, \lambda^*)\) is a saddle point of \(L(x, \lambda)\).

**Remark** Note that the sufficient condition holds for general constrained program, not necessarily convex ones. However, they are far from being necessary and hardly satisfied.

**Definition 9.4 (Normal Cone)** Let \(X \subset \mathbb{R}^n\) and \(x \in X\). The normal cone of \(X\), denoted as \(N_X(x)\), is the set

\[
N_X(x) = \{h \in \mathbb{R}^n : h^T(y - x) \geq 0, \forall y \in X\}
\]
Note that $N_X(x)$ is a closed convex cone.

Examples:

- $x \in \text{int}(X), N_X(x) = \emptyset$
- $x \in \text{rint}(X), N_X(x) = L^\perp$ where $L = \text{Linear subspace parallel to } \text{Aff}(X)$
- $X = \{ x : a_i^T x \geq b_i, i = 1, ..., m \}, x \notin \text{int}(X)$. 
  
  $$N_X(x) = \text{Cone}(\{ a_i | a_i^T x = b_i \})$$

**Theorem 9.5** Let $(P)$ be a convex program and let $x^*$ be a feasible solution. Assume $f, g_1, ..., g_m$ are differentiable at $x^*$.

(a) *(sufficient condition)* If there exists $\lambda^* \geq 0$ satisfying

1) $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i g_i(x^*) \in N_X(x^*)$

2) $\lambda^* g_i(x^*) = 0, \forall i = 1, ..., m$

then $x^*$ is optimal solution to $(P)$ *(Karush-Kuhn-Tucker, KKT condition (1951))*

(b) *(necessary condition)* If $(P)$ also satisfies the slater condition. then the above condition is also necessary for $x^*$ to be optimal.

**Proof:**

(a) *(sufficient part)* Under the KKT condition

1) implies that $L(x, \lambda^*) \geq L(x^*, \lambda^*), \forall x \in X$ because

$$L(x, \lambda^*) \geq L(x^*, \lambda^*) + \nabla_x L(x^*, \lambda^*)(x - x^*)$$

as the last part is non-negative

2) + feasibility of $x^*$ implies $L(x^*, \lambda^*) \geq L(x^*, \lambda), \forall \lambda \geq 0$ because

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

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

(b) *(necessary part)* From previous theorem, there exists $\lambda \geq 0$ such that $(x^*, \lambda^*)$ is a saddle point of $L(x, \lambda)$. We have

$$L(x, \lambda^*) \geq L(x^*, \lambda^*), \forall x \in X \implies (y - x^*)^T \nabla L(x^*, \lambda^*) = \lim_{\epsilon \to 0} \frac{L(x^* + \epsilon(y - x^*), \lambda^*) - L(x^*, \lambda^*)}{\epsilon} \geq 0$$

$$\implies \nabla_x L(x^*, \lambda^*) \in N_X(x^*)$$
L(x^*, \lambda^*) \geq L(x, \lambda), \forall \lambda \geq 0 \Rightarrow \sum_{i=1}^{m} \lambda^*_ig_i(x^*) \geq \sum_{i=1}^{m} \lambda_ig_i(x^*), \forall \lambda \geq 0 \Rightarrow \sum_{i=1}^{m} \lambda^*_ig_i(x^*) \geq 0 \quad \text{(note we also have } \lambda^*_ig_i(x^*) \leq 0)\Rightarrow \lambda^*_ig_i(x^*) = 0, \forall i

This leads to the KKT condition.

**Remark** To summarize, \((x^*, \lambda^*)\) 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**: \(\lambda^*_ig_i(x^*) = 0, \forall i = 1, ..., m\)

**Example**: Given \(a_i > 0, i = 1, ..., n\), solve the problem

\[
\begin{align*}
\min_x & \quad \sum_{i=1}^{n} \frac{a_i}{x_i} \\
\text{s.t.} & \quad x > 0 \\
& \quad \sum_{i=1}^{n} x_i \leq 1
\end{align*}
\]

The Lagrange function \(L(x, \lambda) = \sum_{i=1}^{n} \frac{a_i}{x_i} + \lambda(\sum_{i=1}^{n} x_i - 1)\). The KKT optimality conditions yield

\[
\begin{cases}
x^*_i > 0, \sum_{i=1}^{n} x^*_i \leq 1 \\
\lambda^* \geq 0 \\
-\frac{a_i}{(x^*_i)^2} + \lambda^* = 0 \\
\lambda^*(\sum_{i=1}^{n} x^*_i - 1) = 0
\end{cases}
\Rightarrow x_i = \sqrt{\frac{a_i}{\lambda^*}} \quad \text{and} \quad \sum_{i=1}^{m} x_i = 1 \Rightarrow \begin{cases}
\lambda^* = \sum_{i=1}^{n} \sqrt{\frac{a_i}{\lambda}} \\
x^*_i = \frac{\sqrt{a_i}}{\sum_{i=1}^{m} \sqrt{a_i}}, i = 1, \ldots, m
\end{cases}
\]