LECTURE 12: CONSTRAINED OPTIMIZATION – LAGRANGIAN DUAL PROBLEM

1. Lagrangian dual problem
2. Duality gap
3. Saddle point solution
Lagrangian dual problem

Primal Problem:

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{s.t.} & \quad h(x) = 0 \quad \leftarrow \quad \lambda \in E^m \\
& \quad g(x) \leq 0 \quad \leftarrow \quad \mu \in E^p_+ \\
& \quad x \in X
\end{align*}
\]

Lagrangian Dual Problem:

\[
\begin{align*}
(\text{LD}) & \quad \text{maximize} \quad \phi(\lambda, \mu) \\
\text{s.t.} & \quad \mu \geq 0 \\
& \quad \text{where} \quad \phi(\lambda, \mu) = \inf_{x \in X} \left[ f(x) + \lambda^T h(x) + \mu^T g(x) \right]
\end{align*}
\]
Property 1 – weak duality

Let $\bar{x}$ be a primal feasible solution and $(\bar{\lambda}, \bar{\mu})$ be a dual feasible solution. Then

$$\phi(\bar{\lambda}, \bar{\mu}) = \inf_{x \in X} \{ f(x) + \bar{\lambda}^T h(x) + \bar{\mu}^T g(x) \}$$

$$\leq f(\bar{x}) + \bar{\lambda}^T h(\bar{x}) + \bar{\mu}^T g(\bar{x})$$

$$= 0 \quad \leq 0$$

$$\leq f(\bar{x}).$$
Weak duality theorem

**Theorem (Weak Duality Theorem):**

Let \( \bar{x} \) be primal feasible and \((\bar{\lambda}, \bar{\mu})\) be dual feasible. Then,

\[
\phi(\bar{\lambda}, \bar{\mu}) \leq f(\bar{x}).
\]

**Corollary 1:**

\[
\inf_{x \in \mathcal{R}} f(x) \geq \sup_{(\lambda, \mu) \in \mathcal{D}} \phi(\lambda, \mu)
\]

where \( \mathcal{R} = \{x \in X \mid g(x) \leq 0, \text{ and } h(x) = 0\} \),
\( \mathcal{D} = \{(\lambda, \mu) \mid \lambda \in E^m, \mu \in E_+^p\} \).

**Corollary 2:**

Let \( \bar{x} \) be primal feasible and \((\bar{\lambda}, \bar{\mu})\) be dual feasible. If \( f(\bar{x}) = \phi(\bar{\lambda}, \bar{\mu}) \), then \( \bar{x} \) solves (P) and \((\bar{\lambda}, \bar{\mu})\) solves (LD).

**Corollary 3:**

If \( \sup_{(\lambda, \mu) \in \mathcal{D}} \phi(\lambda, \mu) = +\infty \), then (P) is infeasible.

**Corollary 4:**

If \( \inf_{x \in \mathcal{R}} f(x) = -\infty \), then \( \phi(\lambda, \mu) = -\infty \) for any \( \mu \geq 0 \).
Property 2 – concavity and subgradient

Let \( X \subseteq E^m \) be nonempty and compact, \( f, g, h \) be continuous. Then,

(a) \( \phi(\lambda, \mu) = \inf_{x \in X} \{ f(x) + \lambda^T h(x) + \mu^T g(x) \} \)
    is well defined on \( E^m \times E^p_+ \).
(b) \( \phi(\lambda, \mu) \) is concave over \( E^m \times E^p_+ \).

Proof: Given any \( \omega \in (0, 1) \),

\[
\phi(\omega \bar{\lambda} + (1 - \omega) \bar{\lambda}, \omega \bar{\mu} + (1 - \omega) \bar{\mu}) \\
\geq \omega \phi(\bar{\lambda}, \bar{\mu}) + (1 - \omega) \phi(\bar{\lambda}, \bar{\mu}).
\]

(c) Given any \((\bar{\lambda}, \bar{\mu}) \in E^m \times E^p_+, \) define

\[ X(\bar{\lambda}, \bar{\mu}) \triangleq \{ \bar{x} \in X \mid \bar{x} \text{ minimizes} \] \[ f(x) + \bar{\lambda}^T h(x) + \bar{\mu}^T g(x) \text{ over } X \}. \]

Then \( X(\bar{\lambda}, \bar{\mu}) \neq \phi \) in our setting.

(d) For any \( \bar{x} \in X(\bar{\lambda}, \bar{\mu}) \),

\[
\phi(\lambda, \mu) = \inf_{x \in X} \{ f(x) + \lambda^T h(x) + \mu^T g(x) \} \\
\leq f(\bar{x}) + \lambda^T h(\bar{x}) + \mu^T g(\bar{x}) \\
= \frac{f(\bar{x}) + (\lambda - \bar{\lambda})^T h(\bar{x}) + (\mu - \bar{\mu})^T g(\bar{x})}{\lambda^T h(\bar{x}) + \mu^T g(\bar{x})} \\
= \phi(\bar{\lambda}, \bar{\mu}) + (\lambda - \bar{\lambda})^T h(\bar{x}) + (\mu - \bar{\mu})^T g(\bar{x}).
\]

\[
\Rightarrow \begin{pmatrix} h(\bar{x}) \\ g(\bar{x}) \end{pmatrix} \text{ is a subgradient of } \phi \text{ at } (\bar{\lambda}, \bar{\mu}).
\]

(e) If \( X(\bar{\lambda}, \bar{\mu}) \) is singleton, and \( \bar{x} \in X(\bar{\lambda}, \bar{\mu}) \),
then \( \phi \) is differentiable at \((\bar{\lambda}, \bar{\mu})\) and

\[
\nabla \phi(\bar{\lambda}, \bar{\mu}) = \begin{pmatrix} h(\bar{x}) \\ g(\bar{x}) \end{pmatrix}.
\]
Property 3 – duality gap

• Duality gap may exist

Example 1:

Minimize \( f(x) = x^3 \)

s. t. \( h(x) = x - 1 = 0 \)

\( x \in E^1 \).

(a) \( f \) is not convex.

(b) \( x^* = 1 \) and \( v^* = f(x^*) = 1 \).

(c) \( \phi(\lambda) = \inf_{x \in R} \{ x^3 + \lambda(x - 1) \} \)

\[ = \inf_{x \in R} \{ x^3 + \lambda x - \lambda \} \]

\[ = \begin{cases} 
-\infty, & \lambda > 0 \\
-\infty, & \lambda = 0 \\
-\infty, & \lambda < 0 
\end{cases} \]

(d) \( \phi(\lambda^*) = -\infty \neq f(x^*) = 1 \).

(e) Can you check the local behavior of \( \phi(\lambda) \) around \( x^* = 1 \) and \( \lambda^* = -3 \) ?
Example of duality gap

Example 2 (Bazaraa/Sherali/Shetty p. 205-206)

Minimize $f(x) = -2x_1 + x_2$

s. t. $h(x) = x_1 + x_2 - 3 = 0$

$x \in X = \left\{ \begin{pmatrix} 0 \\ 0 \\ \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \\ \end{pmatrix}, \begin{pmatrix} 4 \\ 0 \\ \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ \end{pmatrix} \right\}$

(d) $\phi(\lambda) = \min_{x \in X} \{-2x_1 + x_2 + \lambda(x_1 + x_2 - 3)\} = \begin{cases} -4 + 5\lambda, & \text{if } \lambda \leq -1 \\ -8 + \lambda, & \text{if } -1 \leq \lambda \leq 2 \\ -3\lambda, & \text{if } \lambda \geq 2. \end{cases}$

(a) $X$ is compact, but not convex.

(b) Only $\begin{pmatrix} 1 \\ 2 \\ \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \\ \end{pmatrix}$ are feasible.

(c) $x^* = \begin{pmatrix} 2 \\ 1 \\ \end{pmatrix}$ with $v^* = f(x^*) = -3$.

(e) $\lambda^* = 2$ with $\phi(\lambda^*) = -6 \neq -3 = f(x^*)$!!
Property 4 – strong duality

Duality gap vanishes only under proper conditions — Strong Duality Theorem

- **Theorem:** (Bazaraa/Sherali/Shetty p.208)

  Assume that
  
  (i) \( X \neq \emptyset \) and is convex;

  (ii) \( f, g \) are convex and \( h \) is affine;

  (iii) (CQ) There exists \( x \in X \) such that
    
    (a) \( g(\bar{x}) < 0 \),
    (b) \( h(\bar{x}) = 0 \),
    (c) \( 0 \in \text{int}[h(X) \triangleq \{h(x) | x \in X\}] \).

  Then,

  \[
  \inf_{x \in \mathcal{F}} f(x) = \sup_{(\lambda, \mu) \in \mathcal{D}} \phi(\lambda, \mu).
  \]

  Moreover, if the inf is finite, then \( \sup_{(\lambda, \mu) \in \mathcal{D}} \phi(\lambda, \mu) \) is achieved at an \( (\bar{\lambda}, \bar{\mu}) \) with \( \bar{\mu} \geq 0 \).

  If the inf is achieved at \( \bar{x} \), then \( \bar{\mu}^T g(\bar{x}) = 0 \).
Geometric interpretation of LD

Consider a case with only one inequality constraint:

\[
\begin{align*}
(P) \quad & \min \ f(x) \quad \max \ \phi(\mu) \\
& \text{s.t.} \quad g_1(x) \leq 0 \quad \text{s.t.} \quad \mu \geq 0 \\
& x \in X \quad (LD) \quad \phi(\mu) = \inf_{x \in X} \{f(x) + \mu g_1(x)\}
\end{align*}
\]

Let

\[G \triangleq \{(y, z)|y = g_1(x), z = f(x) \text{ for some } x \in X\}.
\]

1. (P) says that “on the \((y, z)\) plane, we are looking for a point in \(G\) with \(y \leq 0\) and a minimum ordinate.”

2. \(\phi(\mu) = \inf_{x \in X} \left\{ f(x) + \mu g_1(x) \right\} \)

Note that the contour of

\[\alpha = z + \mu y\]

is a line in the \((y, z)\) plane with slope = \(-\mu\) (\(\leq 0\)) and intercept = \(\alpha\) on the \(z\) axis.

3. (LD) says that we should find the slope of the supporting hyperplane such that its intercept on the \(z\) axis is maximum.

4. When \(X\) is convex and \(f, g\) are convex, \(G\) must be convex. Its supporting hyperplane satisfies that

\[
\phi(\mu^*) = z^* + \mu^* y^*_0 = z^* = f(x^*).
\]
Picture of duality gap
Lagrangian dual of LP

**Example 1 (Linear Programming)**

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

(P)

Let \( X = \{ x \in \mathbb{R}^n \mid x \geq 0 \} \).

\[
\phi(\lambda) \triangleq \inf_{x \geq 0} \{ c^T x + \lambda^T (b - Ax) \}
\]

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

\[
= \begin{cases} 
\lambda^T b, & \text{if } c^T - \lambda^T A \geq 0, \\
-\infty, & \text{otherwise.}
\end{cases}
\]

maximize \( \phi(\lambda) = b^T \lambda \)

(LD) \quad \text{s.t.} \quad A^T \lambda \leq c

\lambda : \text{unrestricted}
Lagrangian dual of QP

Example 2 (Quadratic Programming)

minimize \[ \frac{1}{2} x^T Q x + c^T x \]

(QP) s.t. \[ Ax \leq b \]

where \( Q \) is positive semi-definite.

Let \( X = E^n \).

\[ \phi(\mu) \triangleq \inf_{x \in E^n} \left\{ \frac{1}{2} x^T Q x + c^T x + \mu^T (Ax - b) \right\} \]

convex for any given \( \mu \)

The necessary and sufficient conditions for a minimum is that

\[ Qx + A^T \mu + c = 0. \]

maximize \[ \frac{1}{2} x^T Q x + c^T x + \mu^T (Ax - b) \]

(LD) s.t. \[ Qx + A^T \mu + c = 0 \]

\[ \mu \geq 0. \]
Lagrangian dual of QP

Since \(c^T x + \mu^T A x = -x^T Q x\), we have

\[
\begin{align*}
\text{maximize} & \quad \frac{1}{2} x^T Q x - b^T \mu \\
\text{s.t.} & \quad Q x + A^T \mu = -c \\
& \quad \mu \geq 0.
\end{align*}
\]  

(Dorn’s Dual)

When \(Q\) is positive definite, then

\[
x^* = -Q^{-1}(c + A^T \mu)
\]

and

\[
\phi(\mu) = \frac{1}{2} [Q^{-1}(c + A^T \mu)]^T Q [Q^{-1}(c + A^T \mu)] \\
- c^T Q^{-1}(c + A^T \mu) \\
+ \mu^T (-AQ^{-1}(c + A^T \mu) - b) \\
= \frac{1}{2} \mu^T (-AQ^{-1}A^T) \mu + \mu^T (-b - AQ^{-1}c) \\
D: \text{negative definite} \\
- \frac{1}{2} c^T Q^{-1} c
\]

maximize \(\frac{1}{2} \mu^T D \mu + \mu^T d - \frac{1}{2} c^T Q^{-1} c\)

(LD) \quad \text{s.t.} \quad \mu \geq 0.
Saddle point solution

\[ \begin{align*}
\text{minimize} & \quad f(x) \\
\text{s.t.} & \quad g(x) \leq 0, \\
& \quad h(x) = 0, \\
& \quad x \in X \\
\end{align*} \]

\( \mathcal{F} \)

Lagrangian function

\[ \ell(x, \mu, \lambda) \triangleq f(x) + \mu^T g(x) + \lambda^T h(x). \]

- Definition

\((\bar{x}, \bar{\mu}, \bar{\lambda}) \in E^{n+m+p}\) is called a saddle point
(solution) of \( \ell(x, \mu, \lambda) \) if

(i) \( \bar{x} \in X \),

(ii) \( \bar{\mu} \geq 0 \),

(iii) \( \ell(\bar{x}, \lambda, \mu) \leq \ell(\bar{x}, \bar{\mu}, \bar{\lambda}) \leq \ell(x, \bar{\mu}, \bar{\lambda}) \),

\( \forall x \in X, \mu \in E_+^p, \lambda \in E^m. \)
Saddle point and duality gap

• Basic idea: The existence of a saddle point solution to the Lagrangian function is a necessary and sufficient condition for the absence of a duality gap!

\textbf{Theorem 1:}

Let $\bar{x} \in X$ and $\bar{\mu} \geq 0$. Then, $(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point solution to $\ell(x, \mu, \lambda)$ if and only if

(a) $\ell(\bar{x}, \bar{\mu}, \bar{\lambda}) = \min_{x \in X} \ell(x, \bar{\mu}, \bar{\lambda})$,

(b) $g(\bar{x}) \leq 0$ and $h(\bar{x}) = 0$,

(c) $\bar{\mu}^T g(\bar{x}) = 0$. 
Proof

Proof: (Part 1)
Let \((\bar{x}, \bar{\mu}, \bar{\lambda})\) be a saddle point solution.

By definition, we know (a) holds.

Moreover,
\[
f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) \geq f(\bar{x}) + \mu^T g(\bar{x}) + \lambda^T h(\bar{x}),
\]
\[
\forall \mu \in E_+^p, \lambda \in E^m.
\]

This implies that \(g(\bar{x}) \leq 0\) and \(h(\bar{x}) = 0\), otherwise the right-hand-side may go unbounded above. This proves (b).

Now, let \(\mu = 0\), the above inequality becomes
\[
\bar{\mu}^T g(\bar{x}) \geq 0.
\]

However, \(\bar{\mu} \geq 0\) and \(g(\bar{x}) \leq 0\) imply that
\[
\bar{\mu}^T g(\bar{x}) \leq 0.
\]

Hence \(\bar{\mu}^T g(\bar{x}) = 0\). This proves (c).

(Part 2)
Suppose that \((\bar{x}, \bar{\mu}, \bar{\lambda})\) with \(\bar{x} \in X\) and \(\bar{\mu} \geq 0\) such that (a),(b),(c) hold. Then, by (a)

\[
\ell(\bar{x}, \bar{\mu}, \bar{\lambda}) \leq \ell(x, \bar{\mu}, \bar{\lambda}), \forall x \in X.
\]

By (b)and (c)

\[
\ell(\bar{x}, \bar{\mu}, \bar{\lambda}) = f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) = f(\bar{x})
\]

\[
\geq f(\bar{x}) + \mu^T g(\bar{x}) + \lambda^T h(\bar{x})
\]

\[
= \ell(\bar{x}, \mu, \lambda)
\]

with \(\mu \in E_+^p\) and \(\lambda \in E^m\).

Hence \((\bar{x}, \bar{\mu}, \bar{\lambda})\) is a saddle point solution.
Saddle point theorem

**Theorem 2:**

$(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point solution of $\ell(x, \mu, \lambda)$ if and only if $\bar{x}$ is a primal optimal solution, $(\bar{\mu}, \bar{\lambda})$ is a dual optimal solution and $f(\bar{x}) = \phi(\bar{\mu}, \bar{\lambda})$.

**Proof:** (Part 1)

Let $(\bar{x}, \bar{\mu}, \bar{\lambda})$ be a saddle point solution of $\ell(x, \mu, \lambda)$.

By (b) of Theorem 1, $\bar{x}$ is primal feasible. Since $\bar{\mu} \geq 0$, $(\bar{\mu}, \bar{\lambda})$ is dual feasible. Combining (a), (b), and (c), we have

$$\phi(\bar{\mu}, \bar{\lambda}) = \ell(\bar{x}, \bar{\mu}, \bar{\lambda}) = f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) = f(\bar{x}).$$

By the Weak Duality Theorem, we know that $\bar{x}$ is primal optimal and $(\bar{\mu}, \bar{\lambda})$ is dual optimal.

(Part 2)

Let $\bar{x}$ and $(\bar{\mu}, \bar{\lambda})$ be optimal solutions to (P) and (D), respectively, with

$$f(\bar{x}) = \phi(\bar{\mu}, \bar{\lambda}).$$

Hence, we have $\bar{x} \in X$, $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{\mu} \geq 0$. Moreover,

$$\phi(\bar{\mu}, \bar{\lambda}) \triangleq \inf_{x \in X} \{ f(x) + \bar{\mu}^T g(x) + \bar{\lambda}^T h(x) \} \leq f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) = f(\bar{x}) + \bar{\mu}^T g(\bar{x}) \leq f(\bar{x})$$

But $\phi(\bar{\mu}, \bar{\lambda}) = f(\bar{x})$ is given, the inequalities become equalities. Hence $\mu^T g(\bar{x}) = 0$ and

$$\ell(\bar{x}, \bar{\mu}, \bar{\lambda}) = f(\bar{x}) = \phi(\bar{\mu}, \bar{\lambda}) = \text{minimum} \ell(x, \bar{\mu}, \bar{\lambda}).$$

By Theorem 1, we know $(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point solution to $\ell(x, \mu, \lambda)$. 
Saddle point and KKT conditions

Question:
How does saddle point optimality relate to the K-K-T conditions?

Theorem 3:
Let $\bar{x} \in \mathcal{F}$ satisfies the K-K-T conditions with $\bar{\mu} \in E^p_+$ and $\bar{\lambda} \in E^m$.

Suppose that $f, g_i \ (i \in I(\bar{x}))$ are convex at $\bar{x}$, and that $h_j$ is affine for those with $\bar{\lambda}_j \neq 0$.

Then, $(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point of $\ell(x, \mu, \lambda)$.

Conversely, let $(\bar{x}, \bar{\mu}, \bar{\lambda})$ be a saddle point solution of $\ell(x, \mu, \lambda)$ with $\bar{x} \in \text{int } X$. Then $\bar{x}$ is primal feasible and $(\bar{x}, \bar{\mu}, \bar{\lambda})$ satisfies the K-K-T conditions.
Proof

(Part 1)

Let $\bar{x} \in \mathcal{F}$, $\bar{\mu} \in E^p_+$, $\bar{\lambda} \in E^m$ and $(\bar{x}, \bar{\mu}, \bar{\lambda})$ satisfies the K-K-T conditions, i.e.,

$$\begin{cases}
\nabla f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) = 0 \\
\bar{\mu}^T g(\bar{x}) = 0.
\end{cases} \tag{\ast}$$

By convexity and linearity of $f$, $g_i$ and $h_j$, we have

$$f(x) \geq f(x) + \nabla f(x)(x - x),
$$

$$g_i(x) \geq g_i(x) + \nabla g_i(x)(x - x), \quad i \in I(x),$$

$$h_j(x) = h_j(x) + \nabla h_j(x)(x - x), \quad j = 1, \ldots, m, \quad \lambda_j \neq 0,$$

for $x \in X$.

Multiplying the second inequality by $\mu_i$ and the third inequality by $\lambda_j$, adding to the first inequality, and noting $(\ast)$, it follows from the definition of $\ell$ that

$$\ell(x, \bar{\mu}, \bar{\lambda}) \geq \ell(\bar{x}, \bar{\mu}, \bar{\lambda}), \quad \forall x \in X.$$

Moreover, since $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{\mu}^T g(\bar{x}) = 0$, we have

$$\ell(\bar{x}, \mu, \lambda) \leq \ell(\bar{x}, \bar{\mu}, \bar{\lambda})$$

for $\mu \in E^p_+$ and $\lambda \in E^m$.

Hence $(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point solution.

(Part 2)

Suppose that $(\bar{x}, \bar{\mu}, \bar{\lambda})$ with $\bar{x} \in \text{int } X$ and $\bar{\mu} \geq 0$ is a saddle point solution. Since $\ell(\bar{x}, \mu, \lambda) \leq \ell(\bar{x}, \bar{\mu}, \bar{\lambda})$ for $\mu \in E^p_+$ and $\lambda \in E^m$.

Like in Theorem 1 (Part 1), we have $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{\mu}^T g(\bar{x}) = 0$.

Hence $\bar{x}$ is primal feasible. Moreover $\bar{x}$ is a primal optimal solution because

$$\ell(\bar{x}, \mu, \lambda) \leq \ell(x, \mu, \lambda)$$

for $x \in X$.

Since $\bar{x} \in \text{int } X$, we have $\nabla_x \ell(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0$, i.e.,

$$\nabla f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) = 0.$$

This completes the proof.