Duality: Pick arbitrary vectors \( \lambda \) and \( \mu \geq 0 \).

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad f(x) \\
\text{s.t.} & \quad h(x) = 0 \\
& \quad g(x) \leq 0
\end{align*}
\]

\( (1) \) 

Constrained optimization

\[
\min_{x \in \mathbb{R}^n} \quad f(x) + \sum_{i=1}^{m} \lambda_i \cdot h_i(x) + \sum_{i=1}^{k} \mu_i \cdot g_i(x)
\]

\( (2) \) 

Unconstrained optimization

Theorem: If (1) is convex, (2) is convex too.

Proof: If \( f(x) \) is convex, \( h_i(x) \)'s are linear, and \( g_i(x) \)'s are convex, then

\[
f(x) + \sum_{i=1}^{m} \lambda_i \cdot h_i(x) + \sum_{i=1}^{k} \mu_i \cdot g_i(x) = \text{Convex}
\]

Convex \quad Linear \quad Convex

Theorem: Assume that (1) is convex and Slater's condition holds. Let \( (x_*, \lambda_*, \mu_*) \) be a KKT solution. Then,

\[
(1) = (2) \quad \text{for } (\lambda, \mu) = (\lambda_*, \mu_*)
\]

In other words, \( x_* \) is not only a solution of (1), but also a solution of (2) for \( \lambda = \lambda_* \) and \( \mu = \mu_* \).
Lagrange multipliers: \( \lambda \) and \( \mu \geq 0 \)

Lagrangian: \( l(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{K} \mu_i g_i(x) \)

\[
\min_{x \in \mathbb{R}^n} f(x) \\
\text{s.t.} \quad h(x) = 0 \\
g(x) \leq 0
\]

(1)

\[
\min_{x \in \mathbb{R}^n} l(x, \lambda, \mu)
\]

(2)

- Let \( x_* \) denote a solution of (1) and \( \bar{x}(\lambda, \mu) \) denote a solution of (2).
- Note that \( x = \bar{x}(\lambda, \mu) \) may not be feasible for (1).
- As proved before, we have:

\[
(1) \geq (2) \quad \text{i.e.} \quad f(x_*) \geq l(\bar{x}(\lambda, \mu), \lambda, \mu)
\]

- In other words, the optimal objective value of (2) provides a lower bound on that of (1).
- To find the best lower bound (tightest one) on \( f(x_*) \), we can maximize (2) over \( (\lambda, \mu) \).
Theorem:

- **Weak duality:** Assume \((p)\) is an arbitrary optimization. We have: \((p) \geq (d)\)
i.e., the optimal objective of \((d)\) is a lower bound of the optimal objective of \((p)\).

- **Strong duality:** Assume \((p)\) is convex and Slater's condition holds. We have: \((p) = (d)\)
i.e., the optimal objectives of \((p)\) and \((d)\) are equal. Also, if \((x^*, \lambda^*, \mu^*)\) satisfies KKT, then we have:
1. \( f(x^*_\star) = d(\lambda^*_\star, \mu^*_\star) \)

2. \( x^*_\star \) is a solution of (p)

3. \( (\lambda^*_\star, \mu^*_\star) \) is a solution of (d)

**Convexity:** (d) is a convex problem, i.e.,
\( d(\lambda, \mu) \) is concave and if the problem is recast as
\[
\min_{\lambda, \mu} -d(\lambda, \mu), \quad \text{s.t. } \mu > 0
\]

**Note:** (p) could be very hard to solve due to non-convexity in general, but (d) is always convex.

**Example:**
\[
\min_{x \in \mathbb{R}^2} x_1^2 + x_2^2 \\
\text{s.t. } x_1 + x_2 \leq 1 \quad \text{\(-\mu\)}
\]
\[
\min_{x \in \mathbb{R}^2} l(x, \mu) = \min_{x \in \mathbb{R}^2} (x_1^2 + x_2^2) + \mu (x_1 + x_2 - 1)
\]
This is convex and satisfies Slater's.
To minimize $l(x, \mu)$ with respect to $x$, we need to calculate its gradient with respect to $x$:

$$0 = \nabla_x l(x, \mu) \Rightarrow \begin{cases} 2x_1 + \mu = 0 \\ 2x_2 + \mu = 0 \end{cases}$$

$$\Rightarrow x(\mu) = \begin{bmatrix} -\frac{\mu}{2} \\ -\frac{\mu}{2} \end{bmatrix} \text{ is a stationary point.}$$

Since $l(x, \mu)$ is convex in $x$, $x(\mu)$ is a global min.

$$\Rightarrow d(\mu) = l(x(\mu), \mu) = \left(-\frac{\mu}{2}\right)^2 + \left(-\frac{\mu}{2}\right)^2 + \mu \left(\frac{-\mu}{2} - \frac{-\mu}{2} - 1\right)$$

$$= -\frac{\mu^2}{2} - \mu$$

Note that $d(\mu)$ is concave, as expected.

Weak duality: For every $\mu \geq 0$, $d(\mu)$ is a lower bound on the optimal objective of

$$\min x_1^2 + x_2^2 \quad \text{s.t. } x_1 + x_2 \leq 1$$

Strong duality: If $\mu^*$ is a KKT solution, then $d(\mu^*)$ is equal to the optimal objective of the original problem, i.e.,
Example:

\[
\begin{align*}
\min_{\mathbf{x} \in \mathbb{R}^3} & \quad \mathbf{x}^T \begin{bmatrix} 1 & -4 & -5 \\ -4 & 2 & -6 \\ -5 & -6 & 3 \end{bmatrix} \mathbf{x} \\
\text{s.t.} & \quad x_1^2 = 1 \leftrightarrow \lambda_1 \\
& \quad x_2^2 = 1 \leftrightarrow \lambda_2 \\
& \quad x_3^2 = 1 \leftrightarrow \lambda_3 
\end{align*}
\]

\[\Rightarrow \quad \text{Feasible set:} \quad \mathbf{x} = \begin{bmatrix} \pm 1 \\ \pm 1 \\ \pm 1 \end{bmatrix} \]

- Feasible set has 8 points that are the corners of a cube.

- Every feasible set is a local minimum that satisfies KKT:

\[\Rightarrow \quad \text{because you can draw a ball around each point such that it doesn't contain any (better) feasible point.}\]
- Let's use duality:

$$L(x, \lambda) = x^T \begin{bmatrix} 1 & -4 & -5 \\ -4 & 2 & -6 \\ -5 & -6 & 3 \end{bmatrix} x + \lambda_1 (x_1^2 - 1) + \lambda_2 (x_2^2 - 1) + \lambda_3 (x_3^2 - 1)$$

$$= x^T \begin{bmatrix} 1 + \lambda_1 & -4 & -5 \\ -4 & 2 + \lambda_2 & -6 \\ -5 & -6 & 3 + \lambda_3 \end{bmatrix} x + (-\lambda_1 - \lambda_2 - \lambda_3)$$

- What is the minimum of $x^T m x$?

- Answer: 1. If $m \succeq 0$ then $x^T m x \geq 0$ for all $x$ and $x^T m x = 0$ for $x = 0$

2. If $m$ has a negative eigenvalue, then $x^T m x$ goes toward $-\infty$ if $x$ is chosen as a big number multiplied by an eigenvector of $m$ corresponding to the negative eigenvalue.

$$\Rightarrow \min_{x \in \mathbb{R}^n} x^T m x = \begin{cases} 0 & \text{if } m \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$
\[ d(\lambda) = \min_{x} l(x, \lambda) = \begin{cases} -\lambda_1 - \lambda_2 - \lambda_3 & \text{if } \begin{bmatrix} 1 + \lambda_1 & -4 & -5 \\ -4 & 2 + \lambda_2 & -6 \\ -5 & -6 & 3 + \lambda_3 \end{bmatrix} \geq 0 \\ -\infty & \text{otherwise} \end{cases} \]

- Weak duality: pick a vector \( d \) such that
\[
\begin{bmatrix} 1 + \lambda_1 & -4 & -5 \\ -4 & 2 + \lambda_2 & -6 \\ -5 & -6 & 3 + \lambda_3 \end{bmatrix} \geq 0.
\]
Then, \( -\lambda_1 - \lambda_2 - \lambda_3 \) is a lower bound on the optimal objective value of the original discrete optimization.

- Since (P) is non-convex in this case, strong duality may not hold.

- Example:
\[
\begin{align*}
\min_{x \in \mathbb{R}^3} & \quad x_1 + x_2 \\
\text{s.t.} & \quad x_1 - x_2 + x_3 = 1 - \lambda \\
& \quad x_1 + 2x_2 \leq 1 \quad \rightarrow \mu_1 \geq 0 \\
& \quad x_2 + 3x_3 \leq 2 \quad \rightarrow \mu_2 \geq 0
\end{align*}
\]
Lagrangian:

\[ l(x, \lambda, \mu) = \min_{x \in \mathbb{R}^3} x_1 + x_2 + \lambda (x_1 - x_2 + x_3 - 1) 
+ \mu_1 (x_1 - 2x_2 - 1) 
+ \mu_2 (x_2 + 3x_3 - 2) \]

\[ = \min_{x \in \mathbb{R}^3} (1 + \lambda + \mu_1) x_1 + (1 - \lambda + 2\mu_1 + \mu_2) x_2 
+ (\lambda + 3\mu_2) x_3 + (-\lambda - \mu_1 - 2\mu_2) \]

Note: \[ \min_{x_1} (1 + \lambda + \mu_1) x_1 = \begin{cases} \rightarrow 0 & \text{if } 1 + \lambda + \mu_1 = 0 \\ \rightarrow -\infty & \text{otherwise} \end{cases} \]

\[ \Rightarrow \quad d(\lambda, \mu) = \min_{x \in \mathbb{R}^3} l(x, \lambda, \mu) \]

\[ = \begin{cases} (-\lambda - \mu_1 - 2\mu_2) & \text{if } 1 - \lambda + 2\mu_1 + \mu_2 = 0 \\
\text{otherwise} & \lambda + 3\mu_2 = 0 \end{cases} \]

**Problem (p):**

\[ \min_{x \in \mathbb{R}^3} x_1 + x_2 
\text{s.t.} \quad x_1 - x_2 + x_3 = 1 
\quad x_1 + 2x_2 \leq 1 
\quad x_2 + 3x_3 \leq 2 \]

**Problem (d):**

\[ \max_{\lambda, \mu} -\lambda - \mu_1 - 2\mu_2 
\text{s.t.} \quad 1 + \lambda + \mu_1 = 0 
\quad 1 - \lambda + 2\mu_1 + \mu_2 = 0 
\quad \lambda + 3\mu_2 = 0 
\quad \mu_1, \mu_2 \geq 0 \]