A. Lagrange Duality

The primary and dual functions are defined as following:

**Primary:**

\[
\begin{align*}
\text{minimize } & f_o(x) \\
\text{subject to } & f_i(x) \leq 0 \quad i = 1, 2, \ldots, m \\
& h_j(x) = 0 \quad p = 1, 2, \ldots, p
\end{align*}
\]

**Dual:**

\[
\begin{align*}
\text{maximize } & g(\lambda, v) \\
\text{subject to } & \lambda \geq 0
\end{align*}
\]

The **Lagrangian** is defined as following:

\[
L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x)
\]

The **Lagrangian dual function** is defined as following:

\[
g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v) = \inf_{x \in D} \left[ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x) \right]
\]

B. Properties

1) \( L(x, \lambda, v) \) is convex in \( x \), linear in \( \lambda \) and in \( v \).

2) \( g(\lambda, v) \) is concave in \( \lambda, v \) as an infimum over linear functions of \( \lambda, v \), regardless of the convexity of the primary problem!

3) for all \( \lambda \geq 0 \), we have \( g(\lambda, v) \leq p^* \), where \( p^* \) is the optimal value of the prime problem.
**Proof:** Let $x^*$ be the argument that achieves the optimal value of $p^*$, namely $f_0(x^*) = p^*$, and $f_i(x^*) \leq 0, h_i(x^*) = 0$. Since $\lambda \geq 0$ and $f_i(x^*) \leq 0$ we have

$$\sum_{i=1}^{m} \lambda_i f_i(x^*) \leq 0. \quad (5)$$

Since $h_i(x^*) = 0$ we have

$$\sum_{i=1}^{p} v_i h_i(x^*). \quad (6)$$

So by Eq. (5) and (6) we have

$$\sum_{i=1}^{m} \lambda_i f_i(x^*) + \sum_{i=1}^{p} v_i h_i(x^*) \leq 0. \quad (7)$$

Then,

$$g(\lambda, v) = \inf_{x \in D} \left[ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x) \right] \quad (8)$$

$$\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i f_i(x^*) + \sum_{i=1}^{p} v_i h_i(x^*) \overset{(a)}{\leq} f_0(x^*) \overset{(b)}{=} p^*.$$  

where $(a)$ follows from (5), and $(b)$ is by definition.

4) Since the above is true for all $\lambda \geq 0$, we have

$$\max_{\lambda \geq 0} g(\lambda, v) \leq p^* \quad (9)$$

5) For convex primary problems, under the condition that there exists some $x \in \text{dom} f$ such that $f_i(x) < 0$ for all $i$ (strong inequality), then:

$$\max_{\lambda \geq 0} g(\lambda, v) = p^* \quad (10)$$

6) Any solution of the primary problem (not necessarily the minimal value) is an upper bound for the solution. Any solution of the dual problem is a lower bound.
C. Examples

1) Primary:

\begin{align*}
\text{minimize } & \; x^T x \\
\text{subject to } & \; Ax = b
\end{align*}

So we have

\begin{align*}
L (x, \lambda, \nu) &= x^T x + \nu^T (Ax - b) \\
\frac{\partial}{\partial x} L (x, \lambda, \nu) &= 0 \Rightarrow x_{\min} = -\frac{1}{2} A \nu \\
g (\lambda, \nu) &= L (x_{\min}, \lambda, \nu) = -\frac{1}{4} \nu^T A A^T \nu - \nu^T b
\end{align*}

and the dual problem is:

\begin{align*}
\max_{\nu} \left[ -\frac{1}{4} \nu^T A A^T \nu - \nu^T b \right]
\end{align*}

Since the primary problem is convex (and there are no inequality constraints), we have that \( \max_{\nu} g (\nu) = \min_{x} f_0 (x) \) with the constraints satisfied.

2) Primary:

\begin{align*}
\text{minimize } & \; c^T x \\
\text{subject to } & \; Ax = b \\
& \; -x \leq 0
\end{align*}

So we have

\begin{align*}
L (x, \lambda, \nu) &= c^T x - \lambda^T x + \nu^T (Ax - b) \\
g (\lambda, \nu) &= \inf_{x} L (x_{\min}, \lambda, \nu) = \begin{cases} -\nu^T b, & c^T - \lambda^T + \nu^T A = 0 \\ -\infty, & \text{otherwise} \end{cases}
\end{align*}

So,

\begin{align*}
g (\lambda, \nu) &= \max_{\nu} \left[ -\nu^T b \right]
\end{align*}
subject to \[
\begin{cases}
  e^T - \lambda^T + v^T A = 0 \\
  \lambda \geq 0
\end{cases}
\]

Since \( \lambda \) is not a part of the maximization, we can write

\[g (\lambda, v) = \max_v [-v^T b]\]  

subject to \( c^T + v^T A = 0 \)

3) **Primary:**

\[
\begin{aligned}
\text{minimize } & \quad f_0 (x) \\
\text{subject to } & \quad Ax \leq b \\
& \quad Cx = d
\end{aligned}
\]

So we have

\[L (x, \lambda, v) = f_0 (x) + \lambda^T (Ax - b) + v^T (Cx - d)\]  

\[g (\lambda, v) = \inf_{x \in \text{dom} f_0} L (x_{\text{min}}, \lambda, v)\]

We simplify \( g (\lambda, v) \):

\[g (\lambda, v) = \inf_{x \in \text{dom} f_0} \left[ f_0 (x) + \lambda^T (Ax - b) + v^T (Cx - d) \right]\]

\[= - \sup_{x \in \text{dom} f_0} \left[ - (\lambda^T A + v^T C) x - f_0 (x) + \lambda^T b + v^T d \right]\]

\[= -f^* (-\lambda^T A + v^T C) - \lambda^T b - v^T d\]

where \( f^* (y) = \max_x [yx - f (x)] \) is the conjugate function.

4) Entropy maximization **Primary:**

\[
\begin{aligned}
\text{minimize } & \quad \sum_{i=1}^n x_i \log x_i \\
\text{subject to } & \quad Ax \leq b \\
& \quad 1^T x = 1
\end{aligned}
\]
So we have

\[ L(x, \lambda, \nu) = \sum_{i=1}^{n} x_i \log x_i + \lambda^T (Ax - b) + \nu \left( 1^T x - 1 \right) \quad (22) \]

Taking the derivative in order to find the minimum, we have

\[ \frac{\partial}{\partial x_i} L(x, \lambda, \nu) = \log x_i + 1 + \lambda^T a_i + \nu = 0 \quad (23) \]

so that the minimal value is obtained for

\[ x_i = e^{(-1 - \lambda^T a_i - \nu)} \quad (24) \]

. substituting this into (22) we have

\[ g(\lambda, \nu) = L(x_{\text{min}}, \lambda, \nu) = \sum_{i=1}^{n} e^{(-1 - \lambda^T a_i - \nu)} \log e^{(-1 - \lambda^T a_i - \nu)} \]

\[ + \lambda^T \left( \sum_{i=1}^{n} a_i e^{(-1 - \lambda^T a_i - \nu)} - b \right) \]

\[ + \nu \left( \sum_{i=1}^{n} e^{(-1 - \lambda^T a_i - \nu)} - 1 \right) \]

\[ = - \sum_{i=1}^{n} e^{(-1 - \lambda^T a_i - \nu)} - \lambda^T b - \nu \]

So the dual problem is

\[ \max_{\lambda, \nu} \left[ - \sum_{i=1}^{n} e^{(-1 - \lambda^T a_i - \nu)} - \lambda^T b - \nu \right] \quad (26) \]

subject to \( \lambda \geq 0 \)

We can find the optimal \( \nu \) analytically:

\[ \frac{\partial}{\partial \nu} g(\lambda, \nu) = \sum_{i=1}^{n} e^{(-1 - \lambda^T a_i - \nu)} - 1 = 0 \quad (27) \]

\[ \Rightarrow \nu_{\text{opt}} = \log \sum_{i=1}^{n} e^{(-1 - \lambda^T a_i)} \]

Substituting this into the maximization problem, and recalling that \( \sum_{i=1}^{n} e^{(-1 - \lambda^T a_i - \nu)} = 1 \) from (27), the dual problem can be written as

\[ \max_{\lambda} \left[ -1 - \lambda^T b - \nu_{\text{opt}} \right] = \max_{\lambda} \left[ -1 - \lambda^T b - \log \sum_{i=1}^{n} e^{(-1 - \lambda^T a_i)} \right] \quad (28) \]
So finally the optimization problem can be written as

$$\max_{\lambda} \left[ -\lambda^T b - \log \sum_{i=1}^{n} e^{(-1-\lambda^T a_i)} \right]$$  \hspace{1cm} (29)

subject to $\lambda \geq 0$

I. KKT CONDITIONS AND DUAL FUNCTIONS

Let $(\lambda^*, v^*)$ be the argument that achieves the maximum value of the dual problem, denoted by $d^*$. Let $x^*$ be the argument that achieves the maximum of the primary problem, denoted by $p^*$. Assuming that strong duality holds ($p^* = d^*$), we may write

$$f_0(x^*) = g(\lambda^*, v^*) = \inf_{x \in dom} \left[ f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i v_i^* h_i(x) \right]$$  \hspace{1cm} (30)

$$\leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i v_i^* h_i(x^*) \overset{(a)}{\leq} f_0(x^*),$$

where $(a)$ follows from (7), so that all the inequalities are equalities (as we started with $f_0(x^*)$ and ended with $f_0(x^*)$). Is so, we have that

$$f_0(x^*) = f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i v_i^* h_i(x^*).$$  \hspace{1cm} (31)

**Theorem 1** KKT Conditions: Let $x^*$ and $(\lambda^*, v^*)$ be any primal and dual optimal points which satisfy $p^* = d^*$. Since $x^*$ minimizes $L(\lambda^*, v^*, x)$ over $x$, we must have

$$0 = \nabla_x L(\lambda, v, x) = \nabla_x f_0(x^*) + \sum_i \lambda_i^* \nabla_x f_i(x^*) + \sum_i v_i^* \nabla_x h_i(x^*)$$  \hspace{1cm} (32)

Thus necessary conditions for $x^*$ and $(\lambda^*, v^*)$ to be primal and dual optimal points are

$$f_i(x^*) \leq 0 \hspace{1cm} \forall i$$  \hspace{1cm} (33)

$$h_i(x^*) = 0 \hspace{1cm} \forall i$$  \hspace{1cm} (34)

$$\lambda_i^* \geq 0 \hspace{1cm} \forall i$$  \hspace{1cm} (35)

$$\lambda_i^* f_i(x^*) = 0 \hspace{1cm} \forall i$$  \hspace{1cm} (36)

$$\nabla_x f_0(x^*) + \sum_i \lambda_i^* \nabla_x f_i(x^*) + \sum_i v_i^* \nabla_x h_i(x^*) = 0$$  \hspace{1cm} (37)

which are called the Karush-Kuhn-Tucker (KKT) conditions.
If the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal.