Duality (Boyd, Ch. 5)

Consider a constrained minimization problem

\[ \min_x f_0(x) \]
\[ \text{s.t. } f_i(x) \leq 0 \quad ; \quad i = 1, \ldots, n \]
\[ h_j(x) = 0 \quad ; \quad j = 1, \ldots, m \]

The Lagrangian is defined as

\[ L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^n \lambda_i f_i(x) + \sum_{j=1}^m \nu_j h_j(x) . \]

\( \lambda_i, \nu_j \) are called the Lagrange multipliers.

Define the Lagrange Dual function as

\[ g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) \]

- if \( L \) is unbounded below on \( x \),
  then \( g(\lambda, \nu) = -\infty \).

For example: \( \min_x c^T x \), with no constraint.

- \( g(\lambda, \nu) \) is concave in \((\lambda, \nu)\):
  \( L \) is affine in \((\lambda, \nu)\), so it is concave (and convex).
  \( \inf_x \) of a concave function is concave.
  This happens even if \( f_0(x) \) & \( f_i(x) \) are not convex.

\[ \inf_L \text{ vs } \min L \]:
Let \( f(x) = \frac{1}{x} \). Then, \( \forall x > 0 \),
\[ \inf_x \frac{1}{x} = 0 \], but \( \min_x \frac{1}{x} \)
is undefined.
So, 
\[ g(\lambda, v) = \begin{cases} -b^T v, & \text{if } c + A^T v - \lambda = 0 \\ -\infty, & \text{else.} \end{cases} \]

* $g(\lambda, v)$ is affine in $(\lambda, v)$.

* The dual of a linear programming problem is a linear problem, Lagrange

---

**Lower bound on primal optimal value:**

**Claim** Let $p^* = f_0(x^*)$, where $x^*$ is the optimal solution of the primal problem.

**Then** 
\[ g(\lambda, v) \leq p^* \]

**for any $\lambda \geq 0$, any $v$.**

**Proof:** Let $\bar{x}$ be a feasible point. Then 
\[ f_i(\bar{x}) \leq 0 \text{ and } h_j(\bar{x}) = 0. \]

So, 
\[ \sum_{i=1}^{n} \lambda_i f_i(\bar{x}) + \sum_{j=1}^{m} \nu_j h_j(\bar{x}) \leq 0 \]

Hence 
\[ L(\bar{x}, \lambda, v) = f_0(\bar{x}) + \sum_{i=1}^{n} \lambda_i f_i(\bar{x}) + \sum_{j=1}^{m} \nu_j h_j(\bar{x}) \leq 0 \]

Since $g(\lambda, v) \defeq \inf_x L(x, \lambda, v)$, and 
\[ \inf_x L(x, \lambda, v) \leq L(\bar{x}, \lambda, v), \]
we have 
\[ g(\lambda, v) \leq L(\bar{x}, \lambda, v) \leq f_0(\bar{x}). \]

This holds for any feasible point $\bar{x}$, so in particular holds for $x^*$. Therefore
\[ g(\lambda, v) \leq f_0(x^*) = p^*. \]
Example: Determine the Lagrange dual function of

\[ \min_x \| x \|_2 \]
\[ \text{s.t. } Ax = b \]

The Lagrange function is

\[ \mathcal{L}(x, \nu) = \| x \|_2^2 + \nu^T (Ax - b) \]

In order to find the Lagrange dual function \( g(\nu) \), we need to find

\[ g(\nu) = \inf_x \mathcal{L}(x, \nu). \]

\[ \frac{\partial \mathcal{L}}{\partial x} = 2x + A^T \nu = 0 \implies x = -\frac{1}{2} A^T \nu \]

Therefore,

\[ g(\nu) = \| x^* \|_2^2 + \nu^T (Ax^* - b) \]
\[ = \| -\frac{1}{2} A^T \nu \|_2^2 + \nu^T (A (\frac{1}{2} A \nu) - b) \]
\[ = \frac{1}{4} \nu A A^T \nu - \frac{1}{2} \nu^T A A^T \nu - b^T \nu \]

\[ g(\nu) = \frac{1}{4} \nu A A^T \nu - b^T \nu. \]

\( g(\nu) \) is concave in \( \nu \).

Example: \( \min_x c^T x \)
\[ \text{s.t. } Ax = b \]
\[ x \geq 0 \]

The Lagrange function is

\[ \mathcal{L}(x, \lambda, \nu) = c^T x - \sum_{i=1}^n \lambda_i x_i + \nu^T (Ax - b) \]
\[ = -\nu^T b + (c + A^T \nu - \lambda)^T x \]

The Lagrange dual function is

\[ g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu) \]
\[ = \inf -\nu^T b + \begin{cases} \infty, & \text{if } c + A^T \nu - \lambda \neq 0 \\ 0, & \text{if } c + A^T \nu - \lambda = 0 \end{cases} \]
Lagrange Dual Problem

Since $g(\lambda, v) \leq p^*$, we ask: What is the best upper bound? This leads to the **Dual Problem**.

$$\max_{(\lambda, v)} g(\lambda, v)$$
$$\text{s.t. } \lambda \geq 0.$$

**Example:**

**(Primal)** $\min_x \|x\|^2$
$$\text{s.t. } Ax = b$$
$$g(v) = \frac{1}{2} v^T A A^T v - b^T v$$
$$x = \frac{1}{2} A^T v$$

$\Rightarrow$ **(Dual)** $\max_v g(v)$
$$\Rightarrow \begin{align*}
\max_v & \quad \frac{1}{2} v^T A A^T v - b^T v \\
\text{s.t. } & \lambda \geq 0
\end{align*}$$

This is an unconstrained optimization.

**Example:**

**(Primal)** $\min_x c^T x$
$$\text{s.t. } Ax = b$$
$$x \geq 0$$

$\Rightarrow$ **(Dual)** $\max_v -b^T v$
$$\Rightarrow \begin{align*}
\min_v & \quad b^T v \\
\text{s.t. } & A^T v + c - \lambda \geq 0 \\
\text{s.t. } & \lambda \geq 0
\end{align*}$$

- The dual problem is almost convex, even if the primal is not convex: $g(\lambda, v)$ is concave. So $\max (\lambda, v)$ is a convex problem. $\lambda \geq 0$ is convex.

- If primal is also convex, then (possibly) the dual will solve the primal. (Need more conditions)
Define duality gap:
\[ d^* = \max_{\lambda, \nu} g(\lambda, \nu) \]
\[ \text{s.t. } \lambda \geq 0 \]

Then since \( g(\lambda, \nu) \leq p^* \) \( \forall \lambda \geq 0 \) and \( \nu \), we also have \[ d^* \leq p^* \]

The gap is called the duality gap: \( p^* - d^* \).

**Strong Duality**

If the primal problem is
\[
\min \ f_0(x) \\
\text{s.t. } f_i(x) \leq 0 \quad i.e. \\
A x = b \quad \text{affine}
\]
then when Slater's constraint qualification is satisfied, i.e.
\[
\begin{cases}
\text{there exists } x \text{ s.t.} \\
 f_i(x) < 0, \ i=1,2,...,n, \\
\text{and } A x = b
\end{cases}
\]

then \( d^* = p^* \) if primal is convex.

**Message:** When primal is convex, then we can solve the dual and get back the primal solution.

When primal is not convex, then the dual solution would be a convex approximation to the primal.
Dual Norm and Conjugate Function

Let $\|x\|$ be the norm of $x$. The dual norm of $\|\cdot\|$ is

$$\|x\|_* = \max_{\|z\| \leq 1} (Z^T x).$$

Properties:
(1) For $\|x\|_2$, the dual norm is $\|x\|_* = \|x\|_2$

(By Cauchy inequality) $(Z^T x)^2 \leq \|z\|^2 \|x\|^2$

$\leq \|x\|^2 + \frac{x}{2}.$

Choose $z = \frac{x}{\|x\|_2}$, then upper becomes tight.

(2) For $\|x\|_1$, the dual norm $\|x\|_* = \|x\|_\infty$

is $\max_x Z^T x$ s.t. $\|z\| \leq 1$

When max is attained, only one of the vertices is activated. This corresponds to $\max(x_1, x_2)$.

(3) $\|x\|_{\infty} = \|x\|_\infty$. So the dual norm of $\|x\|_\infty$ is $\|x\|_\infty$.

Let $f: \mathbb{R}^n \to \mathbb{R}$. The function $f^*: \mathbb{R}^n \to \mathbb{R}$, defined as

$$f^*(y) = \sup_x \{ y^T x - f(x) \},$$

is called the conjugate function.

Example (i) if $f(x) = \|x\|_1$, then (See Boyd Example 3.26)

$$f^*(y) = \begin{cases} 0 & \text{if } \|y\|_\infty \leq 1 \\
\infty & \text{else} \end{cases}$$

(ii) if $f(x) = \log(\sum_{i=1}^n e^{x_i})$, then (See Boyd Example 3.25)

$$f^*(y) = \begin{cases} \sum_{i=1}^n y_i \log y_i & \text{if } y \geq 0, \sum y = 1 \\
\infty & \text{else} \end{cases}$$
Solving Primal Problems Via Dual

Consider \( \min_x f_0(Ax + b) \)

The Lagrange dual is
\[
g(\lambda, v) = \inf_x \left\{ f_0(Ax + b) + \lambda^T 0 + v^T 0 \right\}
\]
\[
= p^*.
\]
So even though we have \( p^* = d^* \) (i.e. strong duality), this dual problem is not informative.

Alternatively, consider
\[
\max_v b^T v - f_0^*(v)
\]
s.t. \( A^T v = 0 \)

The Lagrange dual is
\[
g(v) = \inf_{x,y} L(x, y, v) = \inf_y \left\{ \inf_x L(x, y, v) \right\}
\]

So the dual problem is
\[
\min_y f_0(y) + v^T (Ax + b - y)
\]
\[
\frac{\partial L}{\partial x} = 0 \Rightarrow A^T v = 0
\]

Example
\[
\min_x \log \left( \sum_{i=1}^m \exp \{ a_i^T x + b_i \} \right)
\]
\[
\Rightarrow \min_{x,y} \log \left( \sum_{i=1}^m \exp y_i \right)
\]
s.t. \( Ax + b = y \)

\( \Rightarrow \) dual problem:
\[
\max_v -\sum_{i=1}^m v_i \log v_i + b^T v
\]
s.t. \( v \geq 0, A^T v = 0 \)
Pictorial Illustration of Conjugate Function

Consider the conjugate function:

$$f^*(y) = \sup_x \{ yx - f(x) \}.$$  

Three properties of $f^*(y)$:

(i) $f^*(y)$ is always convex

(ii) $f^{**}(x)$ is always convex

(iii) $f^{**}(x) = f(x)$ if $f(x)$ is convex.

Example: $f$ convex but not differentiable

At $x_1$, there is only one slope that makes $\frac{df(x)}{dx}$, which is $y = f'(x)$.  
At $x_0$, there are multiple slopes that can maximize $yx - f(x)$.  
Let's say this range of slope is $[ya, yb]$. Then there will be a line in $f^*(y)$.
Prove (i)
\[ f^*(\alpha y_1 + (1-\alpha)y_2) \]
\[ = \sup_x \left\{ (\alpha y_1 + (1-\alpha)y_2) - f(x) \right\} \]
\[ = \sup_x \left\{ \alpha y_1 x - f(x) \right\} + (1-\alpha) y_2 x - (1-\alpha) f(x) \]
\[ \leq \sup_x \left\{ \alpha (y_1 x - f(x)) \right\} + \sup_x \left\{ (1-\alpha) (y_2 x - f(x)) \right\} \]
\[ = \alpha f^*(y_1) + (1-\alpha) f^*(y_2). \]

(ii) Follows from (i).

(iii) Define \( f^{**}(x) = \sup_z \left\{ x z - f^*(z) \right\} \).

Let \( \hat{z} \) be the maximizer:
\[ \hat{z} = \arg \max_z \left\{ x z - f^*(z) \right\}. \]

Then \( f^{**}(x) = x \hat{z} - f^*(\hat{z}). \)

Since \( \hat{z} \) is the maximizer, we have
\[ \frac{d}{dz} (x z - f^*(z)) \bigg|_{z=\hat{z}} = 0 \Rightarrow \hat{x} = f^*(\hat{z}). \]

So \( f^{**}(x) = x f^*(\hat{z}) - f^*(\hat{z}) \)
\[ = f^*(\hat{z}) \hat{z} - \sup_u \left\{ u \hat{z} - f(u) \right\}. \]

Let \( \hat{u} = \arg \max_u \left\{ u \hat{z} - f(u) \right\} \). Then
\[ \sup_u \left\{ u \hat{z} - f(u) \right\} = \hat{u} \hat{z} - f(\hat{u}) \]

So \( f^{**}(x) = f^*(\hat{z}) \hat{z} - \hat{u} \hat{z} + f(\hat{u}) \).
Now, if we can show that

\[ (a) \widehat{u} \leq \widehat{z} \]

\[ (b) \widehat{u} = f^*(\widehat{z}) \]

then we are done.

To this end, note that

\[
f^*(z) = \frac{d}{dz} f^*(z) = \frac{d}{dz} \left( \max_u \{ uz - f(u) \} \right)
= \frac{d}{dz} \left( \widehat{uz} - f(\widehat{u}) \right) = \widehat{u}.
\]

And since \( \widehat{u} \leq \widehat{z} \)
Example: \( f \) non-convex

If \( f \) is non-convex, there is a line connecting \( x_1 \) and \( x_2 \). Let the slope be \( y_0 \).

On \( f^*(y) \), at \( y_0 \), there are two slopes \( x_1 \) and \( x_2 \) for any \( x \in \text{conv} \{x_1, x_2\} \).

\( f^{**}(x) \) is the convex envelope of \( f(x) \) if \( f \) is non-convex.