1 Lagrangian

A standard problem in optimization is

$$\begin{aligned} &\textit{min} && f_0(x) && (1)\\ &\text{such that} && f_i(x) < 0,\ i = 1, \ldots, m && (2)\\ & && h_i(x) = 0,\ i = 1, \ldots, p && (3) \end{aligned}$$

The Lagrangian of 1 is,

$$L(\mathbf{X}, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \tag{4}$$

Definition 1.1. *The lagrangian dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is,*

$$g(\lambda, \mu) = \inf_x L(x, \lambda, \mu)$$

Theorem 1.2. *If $\lambda \geq 0$ then $g(\lambda, \mu) \leq p^*$. This is referred to as the lower bound of optimization.*

Remark: g is always concave, and g could be $-\infty$.
Remark: If $\lambda \geq 0$, then $g(\lambda, \mu) \leq p^*$ where p^* is the optimal value of original optimization problem.

Proof. A feasible point x is one such that $f_i(x) \leq 0$ and $h_i(x) = 0$. If $\tilde{X}$ is a feasible point, then,

$$\begin{aligned} f_0(\tilde{x}) &\geq L(\tilde{X}, \lambda, \mu) && \text{for any } \lambda, \mu \text{ if } \lambda \geq 0\\ &\geq \inf_x L(\tilde{X}, \lambda, \mu)\\ &:= g(\lambda, \mu) \end{aligned}$$

■

This provides a nontrivial lower bound.

Definition 1.3. *If $\lambda \geq 0$ and $g(\lambda, \mu) > \infty$, (λ, μ) is called dual feasible.*

Example 1.4. *The least norm problem can be stated as,*

$$\min X^T X \ \textit{such that } Ax = b \tag{5}$$

The dual function of 5 is,

$$\begin{aligned}
L(X,\mu) &= X^T X + \mu^T(AX - b)\\
g(\mu) &= \inf_x L(X,\mu)\\
&= L(-\frac{1}{2}A^T\mu, \mu)\\
&= \frac{1}{4}\mu^T AA^T\mu + \mu^T(-\frac{1}{2}AA^T\mu - b)\\
&= -\frac{1}{4}\mu^T AA^T\mu + (-b\mu^T)
\end{aligned}$$

So, $p^* \geq -\frac{1}{4}\mu^T AA^T\mu + (-b\mu^T)$ *for all* μ.

Example 1.5. *In linear programing we want to find,*

$$\min c^T X \ \textit{such that } AX = b \textit{ and } X \geq 0 \tag{6}$$

The Lagrangian of 6 is,

$$L(X,\lambda,\mu) = c^T X + \mu^T(AX - b) - \lambda^T X$$

Therefore,

$$\begin{aligned}
g(\lambda,\mu) &= \inf_x L(X,\lambda,\mu)\\
&= -\mu^T b + (c + A^T\mu - \lambda)^T X\\
&= \begin{cases} -\mu^T b & \textit{if } c + A^T\mu - \lambda = 0\\ -\infty & \textit{otherwise} \end{cases}
\end{aligned}$$

Example 1.6. *In a partitioning problem we want to find,*

$$X^T W X = \sum w_{ij} X_i X_j \ \textit{such that } X_i^2 = 1 \textit{ for } i = 1, 2, \ldots, n \tag{7}$$

The Lagrangian of 7 is,

$$L(X,\mu) = X^T W X + \sum \mu_i (X_i^2 - 1)$$

Therefore,

$$\begin{aligned}
g(\mu) &= \inf_X X^T W X + \sum \mu_i (X_i^2 - 1)\\
&= \inf_X X^T W X + \sum \mu_i X_i^2 - \sum \mu_i\\
&= \inf_X X^T W X + X^T \textit{diag}(\mu) X - \mu^T \mathbf{1}\\
&= \inf_X X^T (W + \textit{diag}(\mu)) X - \mu^T \mathbf{1}\\
&= \begin{cases} -\mu^T \mathbf{1} & \textit{if } W + \textit{diag}(\mu) \geq 0\\ -\infty & \textit{otherwise} \end{cases}
\end{aligned}$$

If μ *is the smallest eigenvalue then* $W + \textit{diag}(\mu) \geq 0$ *and hence* $p^* \geq n\lambda_0(W)$

Definition 1.7. *For any $f : \mathbb{R}^n \to \mathbb{R}$, the conjugate is defined by,*

$$f^*(Y) = \max_{Y \in dom(f)} (Y^T X - f(X))$$

If f is continuous, f^* is a Legendre transform

Example 1.8. *Let $f(x) = \frac{1}{2} X^T Q X$ for $Q \geq 0$. Then,*

$$Y^T X - \frac{1}{2} X^T Q X$$

is concave and

$$f^*(Y) = \max_x (Y^T X - \frac{1}{2} X^T Q X) = \frac{1}{2} Y^T Q^{-1} Y$$

is the conjugate function to $f(X)$.

Example 1.9. *Let $f(X) = I_S(X)$ where S is some set on $\mathbb{R}^n$.*

Then,

$$f^*(Y) = I_{S^*}(X) = \max_X Y^T X$$

is called the support function of S.

Example 1.10. *Let $f(X) = \|X\|$ where $\| \cdot \|$ is any norm.*

Then,

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

is the indicator function of unit ball of dual norm where $\|Y\| = \sup \left\{ Z^T Y : \|Z\| \leq 1 \right\}$.

Actually,

$$\begin{gathered} \text{if } \| \cdot \| \text{ is } L^2, \| \cdot \|_* \text{ is } L^2 \\ \text{if } \| \cdot \| \text{ is } L^p, \| \cdot \|_* \text{ is } L^q \text{ where } (\frac{1}{p} + \frac{1}{q} = 1) \\ \text{if } \| \cdot \| \text{ is } L^1, \| \cdot \|_* \text{ is } \| \cdot \|_\infty \\ \text{if } \| \cdot \| \text{ is nuclear }, \| \cdot \|_* = \sigma_{max}(\cdot) \end{gathered}$$

2 Conjugate and Dual Problems

Suppose we wish to find,

$$\min_x f(x) + g(x)$$

We will use the trick of rewriting this optimization problem as,

$$\min_{x,z} f(x) + g(z) \text{ such that } z = x \tag{8}$$

The Lagrangian of 8 is,

$$\begin{aligned} \inf x, z f(x) + g(z) + \mu^T (z - x) &= \inf x \left(f(x) + \mu^T x \right) + \inf z \left(g(z) + \mu^T z \right) \\ &= -f^*(\mu) - g^*(-\mu) \end{aligned}$$

and f^*, g^* are all conjugate.

Example 2.1. *Suppose we wish to find,*

$$\min \|x\| \text{ such that } Ax = b \tag{9}$$

The dual function is,

$$\begin{aligned} g(\mu) &= \inf_x \|x\| + \lambda(Ax - b) \\ &= \begin{cases} b^T\mu & \|A^T\mu\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

Definition 2.2. *The* <u>*dual problem*</u> *is* $\max_{\mu,\lambda} g(\mu, \lambda) = d^*$.

Definition 2.3. *An optimization problem has* <u>*weak duality*</u> *if* $d^* \leq p^*$*, that is if the dual problem is smaller than or equal to the solution to the original optimization problem.*

The notion of weak duality is used to determine stopping criterion. E.g. If $|f(X_{n+1}) - d^*| < \epsilon$, then $|f(X_{n+1}) - p^*|$ is also small.

Definition 2.4. *An optimization problem has* <u>*strong duality*</u> *if* $d^* = p^*$*, that is if the dual problem is smaller than or equal to the solution to the original optimization problem.*

<u>**Remark:**</u> Strong duality holds for convex optimization. (usually)

Theorem 2.5. <u>***Slater's Condition***</u> *Consider,*

$$\min f_0(x) \text{ such that } Ax = b, f_i(x) \leq 0 \text{ for all } i = 1, \ldots, m$$

If the problem is strictly feasible, then strong duality is always true. i.e. If there exists $x \in dom(f)$ *such that* $f(x) < 0$ *for all* i, $d^* = p^*$.

Example 2.6. *Recall the linear programming example 1.5.. In that case, the dual problem is,*

$$\min b^T\lambda \text{ such that } A^T\lambda + b = 0, \geq 0$$

If there exists $\tilde{X}$ *such that* $A\tilde{x} < b$*, then* $p^* = d^*$.

Example 2.7. *Consider the quadratic problem,*

$$\min x^T Px \text{ such that } Ax \leq b$$

The dual function is,

$$\begin{aligned} g(\lambda) &= \inf \left(x^T Px + \lambda(Ax + b)\right) \\ &= -\frac{1}{4}\lambda^T AP^{-1}A^T\lambda - b^T\lambda \end{aligned}$$

The dual problem is then,

$$\max_{\lambda \geq 0} -\frac{1}{4}\left(\lambda^T AP^{-1}A^T\lambda\right) - b^T\lambda$$

3 Karush-Kuhn-Tucker (KKT) Conditions

The necessary KKT conditions for an optimal solution are,

1. $0 \in \partial f(x) + \sum \lambda_i \partial f_i(x) + \sum \mu_i \partial h_i(x)$ *(Stationarity)*

2. $\lambda_i \geq 0$ *for all* i *(Dual Feasibility)*

3. $f_i(x) \leq 0, h_i(x) = 0$ *for all* i *(Primal Feasibility)*

4. $\lambda_i f_i(x) = 0$ *for all* i *(Complementary Slackness)*