IE 521: Convex Optimization	**Spring 2017, UIUC**

Lecture 9: Optimality Conditions – February 20

Instructor: Niao He *Scribe: Shuanglong Wang*

Courtesy warning: These notes do not necessarily cover everything discussed in the class. Please email TA (swang157@illinois.edu) if you find any typos or mistakes.

In this lecture, we cover the following topics

- Illustrations of Lagrange Duality
- Saddle Point Formulation
- Optimality Conditions (KKT conditions)

References: Bental & Nemirovski Chapter 3.2

9.1 Recall

- Convex program:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, ..., m \quad (P) \\ & x \in X \end{aligned}$$

  where $f(x), g_1, \ldots, g_m$ are convex and X is convex.

- Lagrange function:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

  where $\lambda = (\lambda_1, ..., \lambda_m)$ is called Lagrange multiplier.

- Lagrange dual function:

$$\underline{L}(\lambda) := \inf_{x \in X} L(x, \lambda) = \inf_{x \in X} \{f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)\}$$

  - Note that $\forall \lambda \geq 0, \underline{L}(\lambda) \leq \text{Opt}(P)$

- Lagrange dual program:

$$\max_{\lambda \geq 0} \quad \underline{L}(\lambda)$$

  We show that $\text{Opt}(D) = \text{Opt}(P)$ when (relaxed) Slater condition holds

9.2 Illustrations of Lagrange Duality

- Linear Program Duality

$$(P)\quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \qquad\qquad (D)\quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y \leq c \end{array}$$

First, rewrite the original problem as:

$$\begin{array}{lll} \min & c^T x & \\ \text{s.t.} & Ax - b \leq 0 & (\lambda_1) \\ & b - Ax \leq 0 & (\lambda_2) \\ & x \in \mathbf{R}^n & \end{array}$$

Introducing the multipliers $\lambda = (\lambda_1, \lambda_2) \geq 0$, the Lagrange function is

$$L(x, \lambda) = c^T x + \lambda_1^T (Ax - b) + \lambda_2^T (b - Ax) = (c + A^T \lambda_1 - A^T \lambda_2)^T x + b^T (\lambda_2 - \lambda_1)$$

The Lagrange dual function is

$$\underline{L}(\lambda) = \inf_{x \in \mathbf{R}^n} (c + A^T \lambda_1 - A^T \lambda_2)^T x + b^T (\lambda_2 - \lambda_1) = \begin{cases} b^T(\lambda_2 - \lambda_1), & c + A^T \lambda_1 - A^T \lambda_2 \geq 0 \\ -\infty, & o.w. \end{cases}$$

The Lagrange dual is $\max_{\lambda \geq 0} \underline{L}(\lambda)$, which is equivalent to

$$\begin{array}{ll} \max\limits_{\lambda \geq 0} & b^T (\lambda_2 - \lambda_1) \\ & c + A^T (\lambda_1 - \lambda_2) \geq 0 \end{array}$$

Substituteing $y = \lambda_2 - \lambda_1$, the formulation above is also equivalent to:

$$\begin{array}{ll} \max\limits_{y} & b^T y \\ \text{s.t.} & A^T y \leq c \end{array}$$

- Quadratic Program Duality:

$$(P)\quad \begin{array}{ll} \min\limits_{x} & \frac{1}{2} x^T Q x + q^T x \\ \text{s.t.} & Ax \leq b \\ & x \geq 0 \end{array} \qquad\qquad (D)\quad \begin{array}{ll} \max\limits_{y, \lambda} & -\frac{1}{2} y^T Q y + b^T \lambda \\ \text{s.t.} & A^T \lambda - Q y = q \\ & \lambda \geq 0 \end{array}$$

where $Q \succ 0$.

The Lagrange function is $L(x,\lambda) = \frac{1}{2}x^TQx + q^Tx + \lambda^T(b - Ax)$.
The Lagrange dual function is $\underline{L}(\lambda) = \inf_x\{\frac{1}{2}x^TQx + (q - A^Tx)x + b^T\lambda\}$.
The infimum is achieved when $Qx = A^T\lambda - q, x = Q^{-1}(A^T\lambda - q)$.

$$\underline{L}(\lambda) = -\frac{1}{2}(A^T\lambda - q)^TQ^{-1}(A^T\lambda - q) + b^T\lambda$$

The Lagrange dual is

$$\max_{\lambda\geq 0}\underline{L}(\lambda) \Longleftrightarrow \max_{\lambda\geq 0}\frac{1}{2}(A^T\lambda - q)^TQ^{-1}(A^T\lambda - q) + b^T\lambda \Longleftrightarrow \begin{array}{rl} \max\limits_{\lambda\geq 0, y} & -\frac{1}{2}y^TQy + b^T\lambda \\ \text{s.t.} & A^T\lambda - Qy = q \\ & \lambda \geq 0 \end{array}$$

In both cases discussed above, strong duality holds true.

9.3 Saddle Point Formulation

Recall

$$\begin{array}{llllll} (P): & \min\limits_{x\in X} & \{f(x) : g_i(x) \geq 0, i = 1, \ldots, m\} & = & \min\limits_{x\in X}\bar{L}(x) & = & \min\limits_{x\in X}\max\limits_{\lambda\geq 0} L(x,\lambda) \\ (D): & \max\limits_{\lambda\geq 0} & \underline{L}(\lambda) & = & \max\limits_{\lambda\geq 0}\min\limits_{x\in X} L(x,\lambda) \end{array}$$

Definition 9.1 *(Saddle point) We call (x^*, λ^*), where $x^* \in X, \lambda^* \geq 0$, a saddle point of $L(x,\lambda)$ if $L(x,\lambda^*) \geq L(x^*,\lambda^*) \geq L(x^*,\lambda), \forall x \in X, \lambda \geq 0$.*

Theorem 9.2 *(x^*, λ^*) is saddle point of $L(x,\lambda)$ if and only if x^* is an optimal solution (P), λ^* is an optimal solution to (D) and $Opt(P) = Opt(D)$.*

Proof:

($\Rightarrow$) Assume (x^*, λ^*) is a saddle point,

$$L(x,\lambda^*) \geq L(x^*,\lambda^*) \geq L(x^*,\lambda), \forall x \in X, \lambda \geq 0$$

$$\text{Opt}(P) = \inf_{x\in X}\bar{L}(x) \leq \bar{L}(x^*) = \sup_{\lambda\geq 0} L(x^*,\lambda) = L(x^*,\lambda^*)$$

$$\text{Opt}(D) = \sup_{\lambda\geq 0}\underline{L}(\lambda) \geq \underline{L}(\lambda^*) = \inf_{x\in X} L(x,\lambda^*) = L(x^*,\lambda^*)$$

Hence, $\text{Opt}(P) \leq L(x^*,\lambda^*) \leq \text{Opt}(D)$
Combined with weak duality, we have $\text{Opt}(P) = \text{Opt}(D)$. Hence, $\text{Opt}(P) = \bar{L}(x^*) = L(x^*,\lambda^*) = \underline{L}(\lambda^*) = \text{Opt}(D)$. Thus, x^* solves (P), λ^* solves (D), and $\text{Opt}(P) = \text{Opt}(D)$

($\Leftarrow$) Assume (x^*, λ^*) are optimal solutions to (P) and (D), and $\text{Opt}(P) = \text{Opt}(D)$. By optimality,

$$\text{Opt}(P) = \bar{L}(x^*) = \sup_{\lambda \geq 0} L(x^*, \lambda) \geq L(x^*, \lambda^*)$$

$$\text{Opt}(D) = \underline{L}(\lambda^*) = \inf_{x \in X} L(x, \lambda^*) \leq L(x^*, \lambda^*)$$

Since $\text{Opt}(D) = \text{Opt}(P)$

$$\sup_{\lambda \geq 0} L(x^*, \lambda) = L(x^*, \lambda^*) = \inf_{x \in X} L(x^*, \lambda)$$

i.e. (x^*, λ^*) is a saddle point of $L(x, \lambda)$. ■

Remark:

- The above theorem holds true for any saddle function $L(x, \lambda)$ and its induced primal and dual problems, not limited to the Lagrange function.
- Saddle point always exists for the Lagrange function of a solvable convex program satisfying the Slater condition. More generally, the existence of saddle points is guaranteed for convex-concave saddle functions over convex compact domains (Minimax Theorem).

9.4 Optimality Conditions

Theorem 9.3 *: Let $x^* \in X$*

(i) (sufficient condition) If there exists $\lambda^* \geq 0$, such that (x^*, λ^*) is a saddle point of $L(x, \lambda)$, then x^* is an optimal solution to (P).

(ii) (necessary condition) Assume (P) is convex and satisfies the slater condition. If x^* is an optimal solution to (P) then $\exists \lambda^* \geq 0$, s.t. (x^*, λ^*) is a saddle point of $L(x, \lambda)$

Proof:

(i) (sufficient part) Follows from previous theorem.

(ii) (necessary part) By strong duality theorem, $\exists$ optimal dual solution $\lambda^* \geq 0$, such that $\text{Opt}(P) = \text{Opt}(D)$. Hence, following from the previous theorem, (x^*, λ^*) is a saddle point of $L(x, \lambda)$.

■

Remark Note that the sufficient condition holds for general constrained program, not necessarily convex ones. However, they are far from being necessary and hardly satisfied.

Definition 9.4 *(Normal Cone) Let $X \subset \mathbf{R}^n$ and $x \in X$. The normal cone of X, denoted as $N_X(x)$, is the set*

$$N_X(x) = \left\{ h \in \mathbf{R}^n : h^T (y - x) \geq 0, \forall y \in X \right\}$$

Note that $N_X(x)$ is a closed convex cone.

Examples:

- $x \in \text{int}(X)$, $N_X(x) = \emptyset$
- $x \in \text{rint}(X)$, $N_X(x) = L^\perp$ where $L =$ Linear subspace parallel to Aff(X)
- $X = \{x : a_i^T x \geq b_i, i = 1, ..., m\}$, $x \notin \text{int}(X)$.

$$N_X(x) = \text{Cone}(\{a_i | a_i^T x = b_i\})$$

Theorem 9.5 *Let (P) be a convex program and let x^* be a feasible solution. Assume $f, g_1, ..., g_m$ are differentiable at x^*.*

(a) (sufficient condition) If there exists $\lambda^* \geq 0$ satisfying

1) $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) \in N_X(x^*)$

2) $\lambda^* g_i(x^*) = 0, \forall i = 1, ..., m$
then x^* is optimal solution to (P) (Karush-Kuhn-Tucker, KKT condition (1951))

(b) (necessary condition)If (P) also satisfies the slater condition. then the above condition is also necessary for x^* to be optimal.

Proof:

(a) (sufficient part) Under the KKT condition

1) implies that $L(x, \lambda^*) \geq L(x^*, \lambda^*), \forall x \in X$ because

$$L(x, \lambda^*) \geq L(x^*, \lambda^*) + \nabla_x L(x^*, \lambda^*)(x - x^*)$$

as the last part is non-negative

2) + feasibility of x^* implies $L(x^*, \lambda^*) \geq L(x^*, \lambda), \forall \lambda \geq 0$ because

$$L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^m \lambda_i g_i(x^*) = f(x^*) \geq f(x^*) + \sum \lambda_i g_i(x^*) = L(x^*, \lambda)$$

Hence (x^*, λ^*) is a saddle point of $L(x, \lambda)$.

(b) (necessary part) From previous theorem, there exists $\lambda \geq 0$ such that (x^*, λ^*) is a saddle point of $L(x, \lambda)$. We have

$$L(x, \lambda^*) \geq L(x^*, \lambda^*), \forall x \in X \Rightarrow (y - x^*)^T \nabla L_x(x^*, \lambda^*) = \lim_{\epsilon \to 0} \frac{L(x^* + \epsilon(y - x^*), \lambda^*) - L(x^*, \lambda^*)}{\epsilon} \geq 0$$

$$\Rightarrow \nabla_x L(x^*, \lambda^*) \in N_X(x^*)$$

$$\begin{aligned} L(x^*, \lambda^*) \geq L(x^*, \lambda), \forall \lambda \geq 0 &\Rightarrow \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \geq \sum_{i=1}^{m} \lambda_i g_i(x^*), \forall \lambda \geq 0 \\ &\Rightarrow \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \geq 0 \quad \text{(note we also have } \lambda_i^* g_i(x^*) \leq 0) \\ &\Rightarrow \lambda_i^* g_i(x^*) = 0, \forall i \end{aligned}$$

This leads to the KKT condition.

■

Remark To summarize, (x^*, λ^*) is an optimal primal-dual pair if it satisfies:

- *Primal feasibility:* $x^* \in X, g_i(x^*) \leq 0$
- *Dual feasibility:* $\lambda^* \geq 0$
- *Lagrange optimality:* $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) \in N_X(x^*)$
- *Complementary slackness:* $\lambda_i^* g_i(x^*) = 0, \forall i = 1, ..., m$

Example: Given $a_i > 0, i = 1, ..., n$, solve the problem

$$\begin{aligned} \min_x \quad & \sum\nolimits_{i=1}^{n} \frac{a_i}{x_i} \\ \text{s.t.} \quad & x > 0 \\ & \sum\nolimits_{i=1}^{n} x_i \leq 1 \end{aligned}$$

The Lagrange function $L(x, \lambda) = \sum_{i=1}^{n} \frac{a_i}{x_i} + \lambda(\sum_{i=1}^{n} x_i - 1)$. The KKT optimality conditions yield

$$\begin{cases} x_i^* > 0, \sum_{i=1}^{n} x_i^* \leq 1 \\ \lambda^* \geq 0 \\ -\frac{a_i}{(x_i^*)^2} + \lambda^* = 0 \\ \lambda^*(\sum_{i=1}^{n} x_i^* - 1) = 0 \end{cases} \Rightarrow x_i = \sqrt{\frac{a_i}{\lambda^*}} \text{ and } \sum_{i=1}^{m} x_i = 1 \Rightarrow \begin{cases} \lambda^* = \sum_{i=1}^{n} \sqrt{a_i} \\ x_i^* = \frac{\sqrt{a_i}}{\sum_{i=1}^{n} \sqrt{a_i}}, i = 1, \ldots, m \end{cases}$$