Lecture 9

Necessary and Sufficient Conditions for Optimality of Primal-Dual Pairs

September 29, 2008

Outline

- Necessary and Sufficient Optimality Condition for Primal-Dual Pairs
- Karush-Kuhn-Tucker (KKT) Conditions
- Examples

General Convex Problem and Its Dual

Primal Problem

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & g_j(x) \le 0, \quad j = 1, \ldots, m \\
& a_i^T x = b_i, \quad i = 1, \ldots, r \\
& x \in X
\end{aligned}
$$

When equalities are kept, we have:

Lagrangian Function

$$\mathcal{L}(x, \mu, \lambda) = f(x) + \mu^T g(x) + \lambda^T (Ax - b), \qquad \mu \in \mathbb{R}^m,\ \mu \succeq 0,\ \lambda \in \mathbb{R}^r$$

where $g = (g_1, \ldots, g_m)^T$ and A is a matrix with rows a_i^T, $i = 1, \ldots, r$

Dual Function

$$q(\mu, \lambda) = \inf_{x \in X} \mathcal{L}(x, \mu, \lambda) = \inf_{x \in X} \left\{ f(x) + \mu^T g(x) + \lambda^T (Ax - b) \right\}$$

The infimum is actually taken over $X \cap \mathrm{dom} f \cap \mathrm{dom} g_1 \cap \ldots \cap \mathrm{dom} g_m$

Dual Problem

$$\max_{\mu \ge 0,\ \lambda \in \mathbb{R}^r} q(\mu, \lambda)$$

Optimality Conditions for Primal-Dual Pairs

Theorem Consider convex primal problem with finite optimal value f^*. Assume there is no duality gap, i.e., $q^* = f^*$. Then:
x^* is a primal optimal and (μ^*, λ^*) is a dual optimal if and only if

- **Primal Feasibility**: x^* primal feasible i.e.,

$$g(x^*) \le 0, \quad Ax^* = b, \quad x^* \in X \cap \mathrm{dom} f$$

- **Dual Feasibility**: (μ^*, λ^*) is dual feasible i.e., $\mu^* \succeq 0$

- **Lagrangian Optimality in x**:
  x^* attains the minimum in $\inf_{x \in X} \mathcal{L}(x, \mu^*, \lambda^*)$

- **Lagrangian Optimality in (μ, λ)**:
  (μ^*, λ^*) attains the maximum in $\sup_{\mu \ge 0,\ \lambda \in \mathbb{R}^r} \mathcal{L}(x^*, \mu, \lambda)$

Proof

($\Rightarrow$) Let x^* be primal optimal and (μ^*, λ^*) be dual optimal. Then, they are feasible for primal and dual problems, respectively. By the optimality of these vectors, we have $f^* = f(x^*)$ and $q^* = q(\mu^*, \lambda^*)$. By the no gap relation $f^* = q^*$, we obtain

$$\begin{aligned} f(x^*) = q(\mu^*, \lambda^*) &= \inf_{x \in X} \left\{ f(x) + (\mu^*)^T g(x) + (\lambda^*)^T (Ax - b) \right\} \\ &\le f(x^*) + (\mu^*)^T g(x^*) + (\lambda^*)^T (Ax^* - b) \\ &\le f(x^*) \end{aligned}$$

Hence, the inequalities must hold as equalities, implying that

$$\begin{aligned} \inf_{x \in X} \mathcal{L}(x, \mu^*, \lambda^*) &= \inf_{x \in X} \left\{ f(x) + (\mu^*)^T g(x) + (\lambda^*)^T (Ax - b) \right\} \\ &= f(x^*) + (\mu^*)^T g(x^*) + (\lambda^*)^T (Ax^* - b) \\ &= \mathcal{L}(x^*, \mu^*, \lambda^*) \end{aligned}$$

Thus, x^* attains the minimum in $\inf_{x \in X} \mathcal{L}(x, \mu^*, \lambda^*)$

Also, it follows that

$$q(\mu^*, \lambda^*) = f(x^*) + (\mu^*)^T g(x^*) + (\lambda^*)^T (Ax^* - b) = \mathcal{L}(x^*, \mu^*, \lambda^*) \quad (1)$$

Furthermore, we have

$$\begin{aligned} f(x^*) = q(\mu^*, \lambda^*) &= \sup_{\mu \geq 0, \lambda \in \mathbb{R}^r} q(\mu, \lambda) \\ &= \sup_{\mu \geq 0, \lambda \in \mathbb{R}^r} \inf_{x \in X} \left\{ f(x) + \mu^T g(x) + \lambda^T (Ax - b) \right\} \\ &\leq \sup_{\mu \geq 0, \lambda \in \mathbb{R}^r} \left\{ f(x^*) + \mu^T g(x^*) + \lambda^T (Ax^* - b) \right\} \\ &\leq f(x^*) \end{aligned}$$

Again, it follows that the inequalities hold as equalities, implying that

$$\begin{aligned} q(\mu^*, \lambda^*) &= \sup_{\mu \geq 0, \lambda \in \mathbb{R}^r} \left\{ f(x^*) + \mu^T g(x^*) + \lambda^T (Ax^* - b) \right\} \\ &= \sup_{\mu \geq 0, \lambda \in \mathbb{R}^r} \mathcal{L}(x^*, \mu, \lambda) \end{aligned}$$

This relation and Eq. (1) show that (μ^*, λ^*) attains the maximum in $\sup_{\mu \geq 0, \lambda \in \mathbb{R}^r} \mathcal{L}(x^*, \mu, \lambda)$

Proof continues

($\Leftarrow$) Suppose now that x^* is primal feasible and attains the minimum in $\inf_{x\in X} \mathcal{L}(x,\mu^*,\lambda^*)$, and that (μ^*,λ^*) is dual feasible and attains the maximum in $\sup_{\mu\ge 0,\lambda\in\mathbb{R}^r} \ \mathcal{L}(x^*,\mu,\lambda)$ Thus, we have:

$$
\begin{aligned}
q(\mu^*,\lambda^*) = \inf_{x\in X} \mathcal{L}(x,\mu^*,\lambda^*) \ &= \ \mathcal{L}(x^*,\mu^*,\lambda^*) \\
&= \sup_{\mu\ge 0,\lambda\in\mathbb{R}^r} \ \mathcal{L}(x^*,\mu,\lambda) \\
&\ge \sup_{\mu\ge 0,\lambda\in\mathbb{R}^r} \ \inf_{x\in X} \mathcal{L}(x,\mu,\lambda) \\
&= \sup_{\mu\ge 0,\lambda\in\mathbb{R}^r} q(\mu,\lambda)
\end{aligned}
$$

This and the dual feasibility of (μ^*,λ^*) implies that (μ^*,λ^*) is dual optimal. From the preceding we also have

$$
q(\mu^*,\lambda^*) = \sup_{\mu\ge 0,\lambda\in\mathbb{R}^r} \ \mathcal{L}(x^*,\mu,\lambda) \ge \mathcal{L}(x^*,0,0) = f(x^*) \qquad (2)
$$

By optimality of (μ^*,λ^*), we have $q^* = q(\mu^*,\lambda^*)$. No duality gap relation implies that $f^* = q(\mu^*,\lambda^*)$. This and Eq. (2), yield $f^* \ge f(x^*)$. Since x^* is primal feasible, it follows that x^* is primal optimal.

Implications of the optimality condition theorem

Suppose there is no duality gap and we have an optimal dual multiplier (μ^*, λ^*). As suggested by the preceding theorem, we may consider minimizing over $x \in X$ the Lagrangian $\mathcal{L}(x, \mu^*, \lambda^*)$:

$$\text{minimize} \quad f(x) + (\mu^*)^T g(x) + (\lambda^*)^T (Ax - b) \quad \text{over } x \in X$$

Possibilities for this problem:

- If a unique minimizer exists and it is feasible, then it is primal optimal (for example, a minimizer is unique when $\mathcal{L}(x, \mu^*, \lambda^*)$ is strictly convex in x)

- If a unique minimizer exists but it is not feasible, then the primal problem has no optimal solution (no primal feasible x achieving f^*)

- If multiple minimizers exist only those that are primal feasible are actually primal optimal

Example of Entropy Maximization

Consider the entropy maximization problem equivalent to:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \ln x_i \\ \text{subject to} & Ax \preceq b, \ \mathbf{1}^T x = 1 \end{array}$$

with domain $x \succeq 0$. Its dual is given by

$$\begin{array}{ll} \text{maximize} & -b^T\mu - \lambda - e^{-\lambda-1} \sum_{i=1}^n e^{-a_i^T \mu} \\ \text{subject to} & \mu \succeq 0 \end{array}$$

Suppose that the Slater condition holds: there is a vector $\bar{x}$ such that

$$A\bar{x} \preceq b, \quad \mathbf{1}^T\bar{x} = 1, \quad \bar{x} \succ 0$$

Thus, there is no gap and the dual optimal solution (μ^*, λ^*) exists.

Example continues

The Lagrangian $\mathcal{L}(x,\mu,\lambda)$ at (μ^*,λ^*) is given by

$$\mathcal{L}(x,\mu^*,\lambda^*) = \sum_{i=1}^{n} x_i \ln x_i + (\mu^*)^T(Ax-b) + (\lambda^*)^T\left(\mathbf{1}^T x - 1\right)$$

which is strictly convex over the domain and has a unique minimizer x^* (over the domain), with components x_i^* given by

$$x_i^* = e^{-\left(a_i^T\mu^*+\lambda^*+1\right)} \qquad \text{for all } i = 1,\ldots,n$$

- If x^* is primal feasible, then x^* is a primal optimal solution
- If it is not primal feasible, then the primal problem has no solutions (no feasible vector attaining the primal optimal value)

Equivalent Version of the Optimality Condition

Theorem Consider convex primal problem with finite optimal value f^*. Assume there is no duality gap, i.e., $q^* = f^*$. Then:

x^* is a primal optimal and (μ^*, λ^*) is a dual optimal if and only if

- **Primal Feasibility**: x^* is primal feasible i.e.,

$$g(x^*) \le 0, \quad Ax^* = b, \quad x^* \in X \cap \mathrm{dom} f$$

- **Dual Feasibility**: (μ^*, λ^*) is dual feasible i.e., $\mu^* \succeq 0$

- **Lagrangian Optimality in x**:

$$x^* \text{ attains the minimum in } \quad \inf_{x \in X} \mathcal{L}(x, \mu^*, \lambda^*)$$

- **Complementary Slackness**:

$$(\mu^*)^T g(x^*) = 0$$

Proof

(⇒) We can use all of the relations listed in the Saddle-Point theorem to establish the relations in this theorem. Note, *we need to establish only the complementary slackness*

From the relation:

$\mathcal{L}(x^*, \mu, \lambda) \le \mathcal{L}(x^*, \mu^*, \lambda^*)$ for all (μ, λ) with $\mu \succeq 0$, $\mu \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^r$

it follows $\mathcal{L}(x^*, 0, 0) \le \mathcal{L}(x^*, \mu^*, \lambda^*)$, implying that

$$f(x^*) \le f(x^*) + (\mu^*)^T g(x^*) + (\lambda^*)^T (Ax^* - b)$$

and therefore: $(\mu^*)^T g(x^*) + (\lambda^*)^T (Ax^* - b) \ge 0$

Since x^* is feasible, we have $Ax^* - b = 0$. Hence $(\mu^*)^T g(x^*) \ge 0$

Again, by feasibility of x^* and μ^*, we also have $(\mu^*)^T g(x^*) \le 0$, thus implying that $(\mu^*)^T g(x^*) = 0$

Proof

($\Leftarrow$) We can use all of the relations listed in this theorem to establish the previous optimality relations. Note that *we only need to establish the max-relation for multipliers*:

$\mathcal{L}(x^*, \mu, \lambda) \leq \mathcal{L}(x^*, \mu^*, \lambda^*)$ for all (μ, λ) with $\mu \succeq 0$, $\mu \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^r$

Consider arbitrary (μ, λ) with $\mu \succeq 0$, $\mu \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^r$. By feasibility of x^* and $\mu \succeq 0$, we have

$$f(x^*) + \mu^T g(x^*) + \lambda^T (Ax^* - b) \leq f(x^*)$$

By the complementary slackness and feasibility of x^*, we also have $(\mu^*)^T g(x^*) = 0$ and $(\lambda^*)^T (Ax^* - b) = 0$. Hence,

$$f(x^*) + \mu^T g(x^*) + \lambda^T (Ax^* - b) \leq f(x^*) + (\mu^*)^T g(x^*) + (\lambda^*)^T (Ax^* - b)$$

Since this holds for arbitrary (μ, λ) with $\mu \succeq 0$, $\mu \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^r$, the multipliers (μ^*, λ^*) attain the maximum in $\mathcal{L}(x^*, \mu, \lambda)$.

KKT Conditions

Primal Problem

Convex minimization problem with $X = \mathbb{R}^n$, and differentiable f and all g_j

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \ \ g(x) \leq 0, Ax = b \end{aligned}$$

Theorem Assume that the **optimal value f^* is finite and $q^* = f^*$**. Then: **x^* is a primal optimal and (μ^*, λ^*) is a dual optimal if and only if** the following KKT conditions are satisfied:

- **Primal/Dual Feasibility**: $g(x^*) \leq 0, \ \ Ax^* = b, \ \ \mu^* \succeq 0$
- **Lagrangian Optimality in x**: $\nabla_x \mathcal{L}(x^*, \mu^*, \lambda^*) = 0$

$$\nabla f(x^*) + \sum_{j=1}^{m} \mu_j^* \nabla g_j(x^*) + \sum_{i=1}^{r} \lambda_i^* a_i = 0$$

- **Complementary Slackness**:
  $(\mu^*)^T g(x^*) = 0 \iff \mu_j^* g_j(x^*) = 0$ for all j

Importance of KKT Conditions

KKT Conditions:

- Provide a certificate of optimality for primal-dual pairs
- Exploited in algorithm design and analysis
  - To verify optimality/suboptimality
  - As design principle (algorithms designed for solving KKT equations)

Power Allocation to Communication Channels

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \ln(\alpha_i + x_i) \\ \text{subject to} & x \succeq 0, \ \mathbf{1}^T x = 1 \end{array}$$

where $\alpha_i > 0$. The problem arises in information theory when allocating power to n channels

- x_i is a decision variable representing the power allocated to the i-th channel
- $\ln(\alpha_i + x_i)$ gives the capacity (communication rate) of the i-th channel
- The problem consists of allocating a total power of one unit to the channels so as to maximize the total communication rate.

The domain of the objective function is $\mathrm{dom} f = \{x \mid x + \alpha \succ 0\}$
Slater condition is satisfied: $\bar{x} = 1/n$;
hence, there is no duality gap. The constraint set is compact and contained in the domain $\mathrm{dom} f$. Hence, the objective function is bounded over the constraint set, and therefore f^* is finite

Power Allocation: Solving KKT Conditions

KKT Conditions for the power allocation problem:

$$x^* \succeq 0,\ \mathbf{1}^T x^* = 1,\ \mu^* \succeq 0, \quad \mu_i^* x_i^* = 0,\ \lambda^* - \frac{1}{\alpha_i + x_i^*} - \mu_i^* = 0 \quad \text{for all } i$$

Eliminate μ_i^* using the last relation:

$$x^* \succeq 0, \quad \mathbf{1}^T x^* = 1, \quad \lambda^* \geq \frac{1}{\alpha_i + x_i^*}, \quad x_i^* \left(\lambda^* - \frac{1}{\alpha_i + x_i^*}\right) = 0 \quad \text{for all } i$$

If $\lambda^* < 1/\alpha_i$, by the third relation: $x_i^* > 0$. By Complementarity Slackness (CS), it follows $\lambda^* = 1/(\alpha_i + x_i^*)$, so that $x_i^* = 1/\lambda^* - \alpha_i$.
If $\lambda^* \geq 1/\alpha_i$, by CS it follows $x_i^* = 0$. Hence,

$$x_i^* = \begin{cases} 1/(\alpha_i + x_i^*) & \text{if } \lambda^* < 1/\alpha_i \\ 0 & \text{otherwise} \end{cases} \quad \Leftrightarrow \quad x_i^* = \max\{0, 1/\lambda^* - \alpha_i\}$$

λ^* is determined from $\mathbf{1}^T x^* = 1$, i.e., $\sum_{i=1}^n \max\{0, 1/\lambda^* - \alpha_i\} = 1$