Strong duality, Karush—Kuhn—Tucker theorem

Peter Hajnal

2024. Fall

Strong Duality: Reminder

Reminder

The optimal value of the primal problem is denoted by p^*, and the optimal value of the dual problem is denoted by d^* ($d^*, p^* \in \mathbb{R} \cup \{-\infty, \infty\}$). The following is true (**Weak Duality Theorem**): $d^* \leq p^*$.

• We talk about strong duality when we can guarantee $d^* = p^*$ under certain conditions.

• There are various options for these *certain* conditions.

• An entire *industry* has developed around the development of such conditions. We only discuss one possibility.

Slater's Theorem

Slater's Theorem

Consider the following optimization problem:

Minimize	$c(x)$,-t	
subject to	$f_i(x) \leq 0$	$i = 1, \ldots, k$
	$g_i(x) = 0$	$i = 1, \ldots \ell$

Suppose that

(1) The problem is convex. Thus, c and f_i are convex functions, and g_i are affine functions. This means that the $g_i(x) = 0$ $(i = 1, \ldots, \ell)$ constraints can be written in the following form: $Ax - b = 0$, where $A \in \mathbb{R}^{\ell \times n}$ and $b \in \mathbb{R}^{\ell}$.

(S) There exists $s \in \mathcal{D}$ such that (i) $f_i(s) < 0$ $(i = 1, \ldots, k)$ and $g_i(s) = 0$ $(i = 1, \ldots, \ell)$. Specifically, $s \in \mathcal{L}$. (ii) Moreover, $s \in \text{int}\, \mathcal{D} = \{x : \exists r > 0 \ \ B(x, r) \subset \mathcal{D}\}$, the set of interior points of $\mathcal{D}$, where $B(x, r)$ is the ball centered at x with radius r.

Then, strong duality holds, i.e., $d^* = p^*$.

About the Conditions

- We call (S) Slater's condition.

- Points satisfying condition (S) are called Slater points.

- (S)(i) and (S)(ii) can be weakened. The statement of the theorem remains true under the following (weakened) conditions:

(S) $(\mathrm{i})_0$ We only require from the Slater point s that $f_i(s) < 0$ if f_i is not affine, and $f_i(s) \leq 0$ if it is affine.

(S) $(\mathrm{ii})_0$
$s \in \operatorname{relint} \mathcal{D} = \mathcal{D}$ relative interior $\mathcal{D}$ in the affine hull of $\mathcal{D}$ ($\operatorname{int} \mathcal{D} \subset \operatorname{relint} \mathcal{D}$).

- Below, we present the proof under an important assumption: A (the matrix of equality and affine constraints) has full row rank.

- Without this assumption, the essence of the proof remains, with only a few technical complications making it lengthier.

Proof: 1st Observation

1st Observation

We can assume that

$$p^* \in \mathbb{R}.$$

- From (S), it follows that $\mathcal{L} \neq \emptyset$, implying $p^* < \infty$.

- Moreover, from weak duality, it follows that strong duality holds if $p^* = -\infty$. Thus, we can assume $p^* > -\infty$. Combining these, we conclude .

Proof: $\mathcal{E}$, $\mathcal{F}$, and Observations

- Let

$$\mathcal{E} = \{(\varphi_1, \varphi_2, \ldots, \varphi_k, \gamma_1, \gamma_2, \ldots, \gamma_\ell, \tau) \in \mathbb{R}^k \times \mathbb{R}^\ell \times \mathbb{R} : \exists x \in \mathcal{D}, \text{ such that } \varphi_i \geq f_i(x)\ i = 1, \ldots, k;\ \gamma_i = g_i(x)\ i = 1, \ldots, \ell;\ \tau \geq c(x)\},$$

$$\mathcal{F} = \{(0, 0, \ldots, 0, \tau) \in \mathbb{R}^k \times \mathbb{R}^\ell \times \mathbb{R} :\ \tau < p^*\}.$$

2nd Observation

$\mathcal{E}$ and $\mathcal{F}$ are convex sets.

3rd Observation

$\mathcal{E}$ is closed under increasing the coordinates φ_i and τ.

Proof: Lemma

Lemma

Lemma

$$\mathcal{E} \cap \mathcal{F} = \emptyset.$$

- We will prove this indirectly.
- Assume $v \in \mathcal{E} \cap \mathcal{F}$, i.e., $v \in \mathcal{E}$ and $v \in \mathcal{F}$.
- $v \in \mathcal{F}$ implies $v = (0, \ldots, 0, \tau)$, where $\tau < p^*$.
- $v \in \mathcal{E}$ implies that there exists $x \in \mathcal{D}$ such that $f_i(x) \leq 0$, $g_i(x) = 0$, and $\tau \geq c(x)$.
- Hence, $x \in \mathcal{L}$, and $c(x) \leq \tau < p^*$, which is a contradiction.

Proof: Application of the Separation Theorem

Separation theorem for convex sets ≈ Farkas' Lemma

K, L convex sets and $K \cap L = \emptyset$ then there exists a hyperplane H, which separates the two sets.

That is, it divides the space into closed half-spaces $H^{\leq}$ and $H^{\geq}$, such that $H^{\leq} \supset K$ and $H^{\geq} \supset L$.

• From the theorem and Lemma 2, it follows that there exists an $n = (\lambda_1, \lambda_2, \ldots, \lambda_k, \mu_1, \ldots, \mu_\ell, \nu)$ ($n \in \mathbb{R}^{k+\ell+1} = R^k \times \mathbb{R}^\ell \times \mathbb{R}$) nonzero vector and a real number α, such that the hyperplane $H_{n,\alpha} = \{x \in \mathbb{R}^{k+\ell+1}, n^\top x = \alpha\}$ divides into two half-spaces:

$$H^{\geq}_{n,\alpha} = \{x \in \mathbb{R}^{k+\ell+1} : n^\top x \geq \alpha\} \supseteq \mathcal{E},$$

$$H^{\leq}_{n,\alpha} = \{x \in \mathbb{R}^{k+\ell+1} : n^\top x \leq \alpha\} \supseteq \mathcal{F}.$$

Proof: Further Observations

• From Observation 3, we know that by increasing the first k and last coordinates, we remain in $\mathcal{E}$ and thus in $H^{\geq}$.

Observation 4

$\lambda \succeq 0$ and $\nu \geq 0$.

• $(0, 0, p^* - \epsilon) \in \mathcal{F}$, which implies $(0, 0, p^* - \epsilon) \in H^{\leq}$, thus $\nu(p^* - \epsilon) \leq \alpha$. Since $\epsilon > 0$ is arbitrary, we get limit transitions, yielding

Observation 5

$\nu p^* \leq \alpha$.

• For $x \in \mathcal{D}$, $(f(x), g(x), c(x)) \in \mathcal{E}$, specifically in $H^{\geq}$.

Observation 6

For every $x \in \mathcal{D}$:

$$\sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{i=1}^{\ell} \mu_i g_i(x) + \nu c(x) \geq \alpha.$$

Proof: Case 1: $\nu \neq 0$ ($\nu > 0$)

- Then for every $x \in \mathcal{D}$,

$$L\left(\frac{\lambda_i}{\nu}, \frac{\mu_i}{\nu}, x\right) = \sum_{i=1}^{k} \frac{\lambda_i}{\nu} f_i(x) + \frac{\mu_i}{\nu} g_i(x) + c(x) \geq \frac{\alpha}{\nu}.$$

- This yields

$$\widetilde{c}\left(\frac{\lambda_i}{\nu}, \frac{\mu_i}{\nu}\right) \geq \frac{\alpha}{\nu} \geq p^*.$$

$\left(\frac{\lambda_i}{\nu}\right)_{i=1}^{k}$ are the feasible solutions of the dual optimization problem.

- From this and the previous inequality, it follows that

$$d^* \geq \widetilde{c}\left(\frac{\lambda_i}{\nu}, \frac{\mu_i}{\nu}\right) \geq p^*.$$

- Comparing with weak duality, we obtain strong duality.

Proof: Case 2: $\nu = 0$

- In this case, $\sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{i=1}^{\ell} \mu_i g_i(x) \geq \alpha \geq \nu p^* = 0$ for all $x \in \mathcal{D}$.
- Write the inequality for $x = s$, where s is a Slater point.

$$\sum_{i=1}^{k} \lambda_i f_i(s) + \sum_{i=1}^{\ell} \mu_i g_i(s) \geq 0, \quad \text{where} \quad \lambda_i \geq 0, f_i(s) < 0 \text{ and } g_i(s) = 0.$$

- Then each λ_i must be zero.
- Our initial inequality simplifies to:

$$\sum_{i=1}^{\ell} \mu_i g_i(x) \geq 0,$$

which rewritten becomes $\mu^\top (Ax - b) \geq 0$ for all $x \in \mathcal{D}$.

Proof: Case 2 (continued)

- Let $x = s + \delta$, where $\delta \in \mathbb{R}^{k+\ell+1}$ and $|\delta| < r_0$, with r_0 so small that $B(s, r_0) \subset \mathcal{D}$.

$$\begin{aligned}\mu^\top(A(s+\delta) - b) =& \mu^\top(As + A\delta - b) = \mu^\top(b + A\delta - b) \\ =& \mu^\top A\delta = \sum_{i=1}^{\ell} (\mu^\top A)_i \delta_i \geq 0.\end{aligned}$$

- This holds for $-\delta$ as well, implying $\mu^\top A = 0$.
- Due to our initial assumption (full row rank of A), $\mu = 0$.
- Thus $n = (\lambda, \mu, \nu) = 0$, which is a contradiction.

Break

Slack Conditions, Weak Duality Reminder

Notation

Let $x \in \mathcal{L}$. We say that the i-th inequality constraint is slack at x if

$$f_i(x) < 0.$$

- Let x^* be a primal optimal point and (λ^*, μ^*) be a dual optimal point. Specifically, $\lambda^* \succeq 0$.
- The weak duality is summarized.
- Let $L(\lambda, \mu, x) = c(x) + \lambda^\top f(x) + \mu^\top g(x)$.
- Then

$$d^* = \widetilde{c}(\lambda^*, \mu^*) = \inf L(\lambda^*, \mu^*, x) = \inf_{x \in \mathcal{D}} (c(x) + (\lambda^*)^\top f(x) + (\mu^*)^\top g(x))$$
$$\leq c(x^*) + (\lambda^*)^\top f(x^*) + (\mu^*)^\top g(x^*) \leq c(x^*) = p^*.$$

- If strong duality holds, then equality holds throughout.

Weak Duality: Analysis of the Second Inequality

$$\begin{aligned} d^* =& \widetilde{c}(\lambda^*, \mu^*) = \inf L(\lambda^*, \mu^*, x) = \inf_{x \in \mathcal{D}} \left(c(x) + (\lambda^*)^\top f(x) + (\mu^*)^\top g(x) \right) \\ \leq& c(x^*) + (\lambda^*)^\top f(x^*) + (\mu^*)^\top g(x^*) \leq c(x^*) = p^*. \end{aligned}$$

Definition

Definition Let x_0 be a primal feasible solution, i.e., $f_i(x_0) \leq 0$, $g_i(x_0) = 0$. Let (λ_0, μ_0) be a dual feasible solution, i.e., $(\lambda_0)_i \geq 0$. This solution pair exhibits *complementary slackness* if

(i) $f_i(x_0) < 0$ implies $(\lambda_0)_i = 0$.

(ii) $(\lambda_0)_i > 0$ implies $f_i(x_0) = 0$.

Observation

In the second inequality, equality holds if and only if x^* and (λ^*, μ^*) exhibit complementary slackness.

Weak Duality: Analysis of the First Inequality

$$d^* = \widetilde{c}(\lambda^*, \mu^*) = \inf L(\lambda^*, \mu^*, x) = \inf_{x \in \mathcal{D}} (c(x) + (\lambda^*)^\top f(x) + (\mu^*)^\top g(x))$$
$$\leq c(x^*) + (\lambda^*)^\top f(x^*) + (\mu^*)^\top g(x^*) \leq c(x^*) = p^*.$$

Observation

If the first inequality is an equality, then $c(x) + (\lambda^*)^\top f(x) + (\mu^*)^\top g(x)$ functions attains a minimum at x^*.

Suppose c and f_i functions are differentiable. Then

$$\nabla c(x^*) + (\lambda^*)^\top \nabla f(x^*) + (\mu^*)^\top \nabla g(x^*) = 0.$$

Suppose c, f_i are convex and g_i are affine. Then $c(x) + (\lambda^*)^\top f(x) + (\mu^*)^\top g(x)$ is also convex ($\lambda^* \succeq 0$). In this case, the above condition is both necessary and sufficient for the equality in the second inequality to hold.

Karush—Kuhn—Tucker Theorem: Background

- The theorem was Karush's master's thesis in the 1930s.

- Later, Kuhn and Tucker discovered the theorem and made it known in the 1950s.

Assume that c, f_i, g_j are differentiable. Moreover, c, f_i are convex, and g_j are affine.

Definition: Karush—Kuhn—Tucker Conditions

For $x^* \in \mathbb{R}^n$, $(\lambda^*, \mu^*) \in \mathbb{R}^k \times \mathbb{R}^\ell$, the conditions are

(KKT1) $f_i(x^*) \leq 0$ and $g_i(x^*) = 0$, i.e., x is primal feasible.

(KKT2) $\lambda_i^* \geq 0$, i.e., (λ^*, μ^*) is dual feasible.

(KKT3) x^* and (λ^*, μ^*) exhibit complementary slackness.

(KKT4) $(\nabla c)(x^*) + (\lambda^*)^\top \nabla f(x^*) + (\mu^*)^\top \nabla g(x^*) = 0$.

Karush—Kuhn—Tucker Theorem: The Theorem

KKT Theorem

Suppose g_i are affine functions, c, f_i are convex and differentiable functions.

If strong duality holds with optimal points, then there exist x_0 and (λ_0, μ_0) that satisfy the (KKT1), (KKT2), (KKT3), (KKT4) conditions.

Conversely, if there exist x_0, (λ_0, μ_0), satisfying the (KKT1), (KKT2), (KKT3), (KKT4) conditions, then strong duality holds and these are primal and dual optimal points.

- We have already seen the first part of the theorem.

Establishing Sufficiency

$$\begin{aligned}
\widetilde{c}(\lambda_0, \mu_0) &= \inf(c(x) + \lambda_0^\top f(x) + \mu_0^\top g(x)) \\
&\underset{\text{(KKT4)}}{=} c(x_0) + \lambda_0^\top f(x_0) + \mu_0^\top g(x) \\
&\underset{\text{(KKT3)}}{=} c(x_0).
\end{aligned}$$

Since KKT4 is necessary and sufficient for x_0 to be an optimum point.

Then

$$d^* \underset{\text{(KKT2)}}{\geq} \widetilde{c}(\lambda_0, \mu_0) = c(x_0) \underset{\text{(KKT1)}}{\geq} p^* \underset{\text{weak duality}}{\geq} d^*.$$

From the chain of inequalities, it is evident that equality holds throughout, i.e., strong duality holds, x_0 is a primal optimal point, and (λ_0, μ_0) is a dual optimal point.

KKT: Example I

Minimize	$\frac{1}{2}x^\top Px + q^\top x + r$-t
subject to	$Ax = b$,

where $P \in \mathcal{S}_+^n$.

• $c(x)$ is convex (since $P \in \mathcal{S}_+^n$) and differentiable, hence KKT theorem can be applied.

• We need to find x_0, μ_0 that satisfy all four Karush—Kuhn-Tucker conditions:
(KKT1): $Ax_0 = b$.
(KKT2): $\emptyset$.
(KKT3): $\emptyset$.
(KKT4): $\nabla c(x_0) + \mu_0^\top \nabla(Ax - b)|_{x=x_0} = 0$, i.e.,

$$Px_0 + q + A^\top \mu_0 = 0.$$

KKT: Example I (continued)

- Summarizing the properties of the sought x_0, μ_0:

$$\begin{pmatrix} P_{n\times n} & A^\top_{n\times k} \\ A_{k\times n} & 0 \end{pmatrix} \begin{pmatrix} x_0 \\ \mu_0 \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}.$$

- The discussion of the solvability of this system of equations, and finding the solution in case of solvability, is a straightforward linear algebraic task.

KKT: Example II

Example

Example

Minimize	$2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2$-t
subject to	$x_1^2 + x_2^2 \leq 5,$
	$3x_1 + x_2 \leq 6.$

- In our case, $\mathcal{D} = \mathbb{R}^2$.

- It can be easily verified that the objective function is convex, and the inequality constraints f_i are also convex functions.

- All occurring functions are differentiable.

KKT: Example II (continued)

- The KKT searches for primal/dual $x_1, x_2, \lambda_1, \lambda_2$ instead, satisfying the primal/dual conditions ((KKT1) and (KKT2)):

$$x_1^2 + x_2^2 \leq 5, \quad 3x_1 + x_2 \leq 6, \quad \lambda_1 \geq 0, \quad \lambda_2 \geq 0.$$

- (KKT4) is crucial to find the optimal place. For this:

$$\nabla(2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2) = \begin{pmatrix} 4x_1 + 2x_2 - 10 \\ 2x_1 + 2x_2 - 10 \end{pmatrix},$$

$$\nabla(x_1^2 + x_2^2 - 5) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}, \nabla(3x_1 + x_2 - 6) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

- Hence, expressing the satisfaction of (KKT4):

$$4x_1 + 2x_2 - 10 + 2\lambda_1 x_1 + 3\lambda_2 = 0, \qquad 2x_1 + 2x_2 - 10 + 2\lambda_1 x_2 + \lambda_2 = 0.$$

KKT: Example II (continued)

• What our number four should also know is the complementary slackness property.

• This can be fulfilled in four different ways:

$$I: \quad x_1^2 + x_2^2 = 5 \text{ and } \lambda_1 \geq 0, \quad 3x_1 + x_2 = 6 \text{ and } \lambda_2 \geq 0.$$

$$II: \quad x_1^2 + x_2^2 < 5 \text{ and } \lambda_1 = 0, \quad 3x_1 + x_2 < 6 \text{ and } \lambda_2 = 0.$$

$$III: \quad x_1^2 + x_2^2 < 5 \text{ and } \lambda_1 = 0, \quad 3x_1 + x_2 = 6 \text{ and } \lambda_2 \geq 0.$$

$$IV: \quad x_1^2 + x_2^2 = 5 \text{ and } \lambda_1 \geq 0, \quad 3x_1 + x_2 < 6 \text{ and } \lambda_2 = 0.$$

KKT: Example II (continued)

• By elementary methods, it can be determined that I, II, and III do not lead to appropriate quadruples.

• The possibility IV, however, leads to the

$$x_1 = 1, x_2 = 2, \lambda_1 = 1, \lambda_2 = 0$$

solution.

• From this, it follows that $(1, 2)$ is a primal optimal solution, $(1, 0)$ is a dual optimal solution. Furthermore, strong duality holds.

This is the End!

Thank you for your attention!