Notes 07: Complementary Slackness, KKT Conditions

1. Max-min characterization of strong duality [BV §5.4.1]

For simplicity, assume no equality constraints.

$$\sup_{\lambda \geqslant 0} L(x, \lambda) = \sup_{\lambda \geqslant 0} \left(f_0(x) + \sum_{1 \leqslant i \leqslant m} \lambda_i f_i(x) \right) = \begin{cases} f_0(x) & f_i(x) \leqslant 0, \quad 1 \leqslant i \leqslant m \\ \infty & \text{otherwise} \end{cases}$$

This is because when x is feasible ($f_i(x) \leqslant 0$ for all $1 \leqslant i \leqslant m$), the optimal choice is $\lambda = 0$. If x is infeasible and $f_i(x) > 0$ for some $1 \leqslant i \leqslant m$, then the supremum can be arbitrarily large by taking λ_i to be arbitrarily large.

Therefore optimal primal value is

$$p^\star = \inf_x \sup_{\lambda \geqslant 0} L(x, \lambda).$$

By definition, dual value is

$$d^\star = \sup_{\lambda \geqslant 0} \inf_x L(x, \lambda).$$

Weak duality means the inequality

$$\sup_{\lambda \geqslant 0} \inf_x L(x, \lambda) \leqslant \inf_x \sup_{\lambda \geqslant 0} L(x, \lambda)$$

Strong duality means the equality

$$\sup_{\lambda \geqslant 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geqslant 0} L(x, \lambda)$$

Strong duality means we can change the order of minimizing over x and maximizing over $\lambda \geqslant 0$, without changing the result.

In fact the above weak duality inequality holds for any function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ (and not just L):

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leqslant \inf_{w \in W} \sup_{z \in Z} f(w, z).$$

This inequality is called max-min inequality.

When equality holds, f (and W and Z) satisfy strong max-min property. Equality only holds for special f and W and Z.

2. Game interpretation of strong duality [BV §5.4.3]

We can interpret the max-min inequality in terms of continuous zero-sum game. The first player chooses $w \in W$, the second player chooses $z \in Z$, and player 1 pays an amount $f(w, z)$ to player 2. Player 1 wants to minimize f and Player 2 wants to maximize f.

If Player 1 makes a choice first, then Player 2 chooses after learning Player 1's choice. Player 2 wants to choose $z \in Z$ to maximize $f(w, z)$, depending on first player's choice $w \in W$. Player 2's payoff is therefore $\sup_{z \in Z} f(w, z)$. Player 1 assumes player 2 will follow this strategy, and will choose $w \in W$ to minimize worst-case payoff. Therefore Player 1 tries to achieve payoff

$$\inf_{w \in W} \sup_{z \in Z} f(w, z).$$

If the order of play is reversed, then Player 2 chooses $z \in Z$ first, and Player 1 chooses $w \in W$ after learning Player 2's choice. Then Player 1 tries to achieve $\inf_{w \in W} f(w, z)$ to minimize payment. Player 2 assumes Player 1 will follow this strategy, and tries to maximize payment $\sup_{z \in Z} \inf_{w \in W} f(w, z)$.

Min-max inequality states the intuitive fact that it's better to go second (to know opponent's choice before choosing). When equality holds, there is no advantage in playing second.

3. Complementary Slackness [BV §5.5.2]

Suppose primal and dual optimal values are attained and equal (strong duality holds). Let $x^\star$ be primal optimum and $(\lambda^\star, \nu^\star)$ be dual optimum. Then

$$\begin{aligned}
f_0(x^\star) &= g(\lambda^\star, \nu^\star) && p^\star = d^\star \text{ by assumption} \\
&= \inf_x \left(f_0(x) + \sum_{1\leqslant i\leqslant m} \lambda_i^\star f_i(x) + \sum_{1\leqslant i\leqslant p} \nu_i^\star h_i(x) \right) && \text{definition of g} \\
&\leqslant f_0(x^\star) + \sum_{1\leqslant i\leqslant m} \lambda_i^\star f_i(x^\star) + \sum_{1\leq i\leqslant p} \nu_i^\star h_i(x^\star) && \text{definition of inf over } x \\
&\leqslant f_0(x^\star) && f_i(x^\star) \leqslant 0, \lambda_i \geqslant 0, h_i(x^\star) = 0
\end{aligned}$$

The above inequalities must all be equalities. Equality on the 3rd line means $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$. Equality on the 4th line means

$$\sum_{1\leqslant i\leqslant m} \lambda_i^\star f_i(x^\star) = 0.$$

Since each term in the sum is nonpositive, we have

$$\lambda_i^\star f_i(x^\star) = 0, \qquad 1 \leqslant i \leqslant m.$$

So $\lambda_i > 0$ implies $f_i(x^\star) = 0$.
And $f_i(x^\star) < 0$ implies $\lambda_i = 0$.
So a dual variable $\lambda_i^\star$ is positive only when the corresponding constraint is tight at $x^\star$.

4. KKT optimality conditions

We consider general optimization problem (not necessarily convex) where $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ are all differentiable.

Suppose primal and dual optimal values are equal. Let $x^\star$ be primal optimum and $(\lambda^\star, \nu^\star)$ be dual optimum.

Since $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$ over x, the gradient must vanish at $x^\star$,

$$\nabla f_0(x^\star) + \sum_{1\leqslant i\leqslant m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{1\leqslant i\leqslant p} \nu_i^\star \nabla h_i(x^\star) = 0.$$

Altogether,

$$\begin{aligned}
f_i(x^\star) \leqslant 0, \quad 1 \leqslant i \leqslant m, \qquad h_i(x^\star) = 0, \quad 1 \leqslant i \leqslant p \qquad & \text{(primal feasibility)} \\
\lambda_i^\star \geqslant 0, \qquad 1 \leqslant i \leqslant m \qquad & \text{(dual feasibility)} \\
\lambda_i^\star f_i(x^\star) = 0, \quad 1 \leqslant i \leqslant m \qquad & \text{(complementary slackness)} \\
\nabla f_0(x^\star) + \sum_{1\leqslant i\leqslant m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{1\leqslant i\leqslant p} \nu_i^\star \nabla h_i(x^\star) = 0 \qquad & \text{(stationarity)}
\end{aligned}$$

These conditions together are known as Karush–Kuhn–Tucker (KKT) conditions. They are necessary conditions for any differentiable optimization problem (not necessarily convex) to attain primal and dual optimum solutions with zero duality gap.

We now assume our optimization problem is convex, in addition to $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ being differentiable. KKT conditions are also sufficient for primal solution $\tilde{x}$ and dual solution $(\tilde{\lambda}, \tilde{\nu})$ to be optimal, with zero duality gap.

Indeed, $\tilde{x}$ is primal feasible and $(\tilde{\lambda}, \tilde{\nu})$ is dual feasible. Further,

$$\begin{aligned}
g(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) && \tilde{x} \text{ is a minimizer by stationarity} \\
&= f_0(\tilde{x}) + \sum_{1\leqslant i\leqslant m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{1\leqslant i\leqslant p} \tilde{\nu}_i h_i(\tilde{x}) && \text{definition of } L \\
&= f_0(\tilde{x}) && \text{complementary slackness \& primal feasibility}
\end{aligned}$$