MATH 565: Lecture 20 (03/26/2026)

Today: * minimax optimality
* strong duality
* KKT optimality conditions

Recall: $\pi^* = \min_{\bar{w}} \{F(\bar{w}) \mid f_i(\bar{w}) \le 0, i=1,\dots,m\}$ (P) primal

$$L(\bar{\alpha}) = \min_{\bar{w}} F(\bar{w}) + \sum_{i=1}^{m} \alpha_i f_i(\bar{w}) \quad \text{(Lagrangian Relaxation)}$$

$$\delta^* = \max_{\bar{\alpha} \geq \bar{0}} L(\bar{\alpha}) \quad \text{(D) dual}$$

$$\Rightarrow \quad \delta^* = \max_{\bar{\alpha} \geq \bar{0}} \min_{\bar{w}} \left\{ F(\bar{w}) + \sum_{i=1}^{m} \alpha_i f_i(\bar{w}) \right\} \quad \text{(D)}$$

you can't swap!

The order of max-min (v/s min-max) is important here. Under certain settings, one may be able to switch the order of max-min to min-max. But the result may be different if one does this in general. Here is a simple example.

Example Consider a simple 2-player game with a 2×2 payoff matrix as shown. We want to minimize across rows and maximize across columns. The order in which we do these two steps matters here.

	y_1	y_2	row min
x_1	5	2	2
x_2	3	4	3
column max	5	4	

Max-min, i.e., taking the maximum of row minimums, gives 3; while min-max, i.e., taking the minimum of column maximums, gives 4.

More generally, we get the **minimax inequality**:

$$\max_{x} \min_{y} f(x,y) \leq \min_{y} \max_{x} f(x,y)$$

the result holds for vectors $\bar{x}, \bar{y}$ as well

The difference between the two expressions is the duality gap.

Proof For a given (x_0, y_0), we have that

$$\min_{y} f(x_0, y) \leq f(x_0, y_0) \leq \max_{x} f(x, y_0)$$

Combining, we get $\min_{y} f(x,y) \leq \max_{x} f(x,y)$

as the result should hold for any (x,y) in place of (x_0,y_0).

This inequality holds for all choices of x on the LHS and all choices of y on the RHS. Hence, we get that

$$\max_{x} \min_{y} f(x,y) \leq \min_{y} \max_{x} f(x,y) \qquad \square$$

Let's look at the Lagrangian relaxation. We define

$$H(\bar{w}, \bar{\alpha}) = F(\bar{w}) + \sum_{i=1}^{m} \alpha_i f_i(\bar{w})$$

where $\bar{w}$: minimization variables and $\bar{\alpha}$: maximization variables.

Then, (D) solves $\max_{\bar{\alpha} \geq \bar{0}} L(\bar{\alpha}) = \max_{\bar{\alpha} \geq \bar{0}} \min_{\bar{w}} H(\bar{w}, \bar{\alpha}) \leq \pi^*$

as $L(\bar{\alpha})$ gives a lower bound on π^*.

But $\min_{\bar{w}} \max_{\bar{\alpha} \geq \bar{0}} H(\bar{w}, \bar{\alpha})$ always gives the primal problem (P).

Lemma 11 The minimax problem $\min_{\bar{\omega}} \max_{\bar{\alpha} \geq \bar{0}} H(\bar{\omega}, \bar{\alpha})$ is equivalent to the original unrelaxed problem (P).

Note: F, f_i need not be convex for this result to hold.

Proof We make the following two observations.

* Since we want to maximize H, if any $f_i(\bar{\omega}) > 0$, i.e., the i^{th} constraint is violated, then $H \to \infty$ by setting $\alpha_i \to \infty$, and hence we cannot honor the minimization objective.

$\Rightarrow \bar{\omega} \in C$ (feasible, $f_i(\bar{\omega}) \leq 0 \ \forall i$) for a valid optimal solution.

of the minimax problem

Equivalently, if $\bar{\omega} \in C$ exists (feasible) then $\max_{\bar{\alpha}} H \not\to \infty$ (by weak duality).

* Since $\bar{\omega}$ is feasible, $f_i(\bar{\omega}) \leq 0 \ \forall i$.

$\Rightarrow \alpha_i f_i(\bar{\omega}) \leq 0$ as $\alpha_i \geq 0$ (for the max).

$\Rightarrow$ to maximize H, set $\alpha_i = 0 \ \forall i$.

$\Rightarrow$ Contribution to $H(\bar{\omega}, \bar{\alpha})$ from $\sum_{i=1}^{m} \alpha_i f_i(\bar{\omega})$ is 0.

Hence, solution to the min-max problem is feasible ($f_i(\bar{\omega}) \leq 0 \ \forall i$) and the penalty term in $L(\bar{\alpha}) = 0$. Hence, it minimizes $F(\bar{\omega})$ over feasible $\bar{\omega}$, i.e., it solves the primal problem (P). $\square$

So, $(P) \equiv \min_{\bar{w}} \max_{\bar{\alpha} \geq \bar{0}} H(\bar{w}, \bar{\alpha}) = \pi^*$

and $(D) \equiv \max_{\bar{\alpha} \geq \bar{0}} \min_{\bar{w}} H(\bar{w}, \bar{\alpha}) = \delta^*$

Strong Duality

Result Let $F(\bar{w})$ and $f_i(\bar{w})$ be convex functions. Then $\delta^* = \pi^*$ provided Slater's condition holds, i.e., $\exists\, \bar{w}$ s.t. $f_i(\bar{w}) < 0 \ \forall i$, i.e., a strictly feasible point exists.

Most ML problems satisfy Slater's condition.

Karush-Kuhn-Tucker (KKT) Optimality Conditions

Recall (P) and (D), the primal and dual problems.

* For $(\bar{w}, \bar{\alpha})$ to be optimal for (P),

  $f_i(\bar{w}) \leq 0 \ \forall i$ ($\bar{w}$ is feasible)

* Furthermore, if $f_i(\bar{w}) < 0$, then setting $\alpha_i = 0$ ensures the maximality of (P) w.r.t. $\bar{\alpha}$.

  $\Rightarrow \underline{\alpha_i f_i(\bar{w}) = 0}$ $\forall i$ for any optimal solution to (P).

These are the complementary slackness conditions (CSCs).

* primal feasibility : $f_i(\bar{w}) \leq 0 \ \forall i$
* dual feasibility : $\alpha_i \geq 0 \ \forall i$

CSCs say that at most one of these two constraints (for each i) can be "slack" or "loose" at optimality (i.e., satisfied as strict inequality).

* We also need $\nabla_{\bar{w}} H(\bar{w},\bar{\alpha}) = \bar{0}$ as the first order optimality condition (for fixed $\bar{\alpha}$).

$$\nabla_{\bar{w}} H(\bar{w},\bar{\alpha}) = \nabla F(\bar{w}) + \sum_{i=1}^{m_1} \alpha_i \nabla f_i(\bar{w}) = \bar{0}$$

These are the **stationarity conditions**.

Theorem 12 (KKT Optimality conditions) Consider $\min\{F(\bar{w}) \mid f_i(\bar{w}) \leq 0, i=1,\dots,m\}$ for convex $F(\cdot)$ and convex $f_i(\cdot)$. Then $(\bar{w},\bar{\alpha})$ is optimal for (P) and (D) **iff** all the following conditions hold.

* feasibility : $f_i(\bar{w}) \leq 0, \ i=1,\dots,m$
  $\bar{\alpha} \geq \bar{0}$ (or $\alpha_i \geq 0 \ \forall i$).
* CSCs : $\alpha_i f_i(\bar{w}) = 0 \ \forall i$
* stationarity : $\nabla F(\bar{w}) + \sum_{i=1}^{m_1} \alpha_i \nabla f_i(\bar{w}) = \bar{0}$.

Note: We do not need to check the second order conditions here as $F(\cdot)$ and $f_i(\cdot)$ are convex.

Hessian $\succeq 0$