Lecture 5: Lagrange Multipliers and KKT

TTIC 31070 / CMSC 35470 / BUSF 36903 / STAT 31015

Convex Optimization

Prof. Zhiyuan Li

Spring 2026

From Perturbation to Optimality Certificates

Lecture 4: dual variables arise from perturbation value functions and give **lower bounds** on the primal value.

Today: when do those lower bounds become **exact certificates**?

Three steps:

1. **Weak duality**: every dual-feasible multiplier gives a lower bound
2. **KKT**: a primal–dual triple is an exact certificate once the bound is tight
3. **Slater's condition**: guarantees this certificate system is *complete*

Convex Constrained Program

Definition 5.1 (Convex constrained program $\mathcal{P}$).

$$\begin{aligned} \min_{x \in C} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \dots, m, \\ & h_j(x) = 0, \quad j = 1, \dots, p, \end{aligned}$$

where $C \subseteq E$ is nonempty convex, $f_0, f_1, \dots, f_m : C \to \mathbb{R}$ are convex, $h_1, \dots, h_p : C \to \mathbb{R}$ are affine.

Optimal value: $\mathrm{value}(\mathcal{P}) := \inf\{f_0(x) : x \text{ feasible for } \mathcal{P}\}$.

Convention: $\mathrm{value}(\mathcal{P}) = +\infty$ if infeasible.

Lagrangian and Dual Program

Definition 5.2 (Lagrangian, dual function, dual program). For $x \in C$, $\lambda \in \mathbb{R}^m_+$, $\nu \in \mathbb{R}^p$:

$$L(x, \lambda, \nu) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x).$$

Dual function: $q(\lambda, \nu) := \inf_{x \in C} L(x, \lambda, \nu)$.

Dual program $\mathcal{D}$: $\max_{\lambda \geq 0, \nu} q(\lambda, \nu)$.

q is concave (infimum of affine functions of (λ, ν)), even if the primal is not convex.

Weak Duality

Theorem 5.1 (Weak duality). If $x \in C$ is primal feasible and $(\lambda, \nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$, then

$$q(\lambda, \nu) \leq f_0(x).$$

Therefore $\mathrm{value}(\mathcal{D}) \leq \mathrm{value}(\mathcal{P})$.

Proof.

$$q(\lambda, \nu) = \inf_{z \in C} L(z, \lambda, \nu) \leq L(x, \lambda, \nu) = f_0(x) + \underbrace{\sum_i \lambda_i f_i(x)}_{\leq 0} + \underbrace{\sum_j \nu_j h_j(x)}_{= 0} \leq f_0(x). \ \square$$

The gap $\mathrm{value}(\mathcal{P}) - \mathrm{value}(\mathcal{D}) \geq 0$ is the **duality gap**. When is it zero?

KKT Conditions

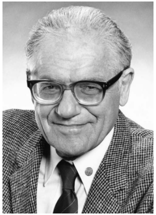
Definition 5.3 (KKT point of $\mathcal{P}$). A triple (x^*, λ^*, ν^*) is a *KKT point* if:

1. **Primal feasibility:** $x^* \in C$, $f_i(x^*) \leq 0$, $h_j(x^*) = 0$.
2. **Dual feasibility:** $\lambda^* \in \mathbb{R}_+^m$.
3. **Stationarity:** $x^* \in \operatorname{argmin}_{x \in C} L(x, \lambda^*, \nu^*)$.
4. **Complementary slackness:** $\lambda_i^* f_i(x^*) = 0$ for all i.

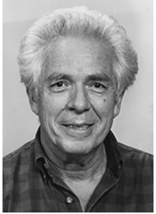
Stationarity $=$ x^* minimizes the Lagrangian over C.

Complementary slackness $=$ each multiplier is zero or its constraint is tight.

Karush–Kuhn–Tucker: A Brief History



William Karush
1917–1997



Harold W. Kuhn
1925–2014

Albert W. Tucker
1905–1995

1939. Karush discovers the conditions in his **Master's thesis at the University of Chicago**. The thesis goes unpublished and unnoticed.

1951. Kuhn & Tucker independently rediscover and publish the conditions at the Second Berkeley Symposium: *"Nonlinear Programming."*

Later. When Karush's 1939 priority was recognized, "Kuhn–Tucker" became **Karush–Kuhn–Tucker (KKT)**.

Example: Projection onto the Simplex

Example 5.1. Given $z \in \mathbb{R}^n$, project onto the probability simplex $\Delta_n := \{x \in \mathbb{R}^n : x_i \geq 0,\ \sum_i x_i = 1\}$:

$$\min_{x \in \mathbb{R}^n} \ \frac{1}{2}\|x - z\|_2^2 \quad \text{s.t.} \quad x_i \geq 0 \ \forall\, i, \quad \sum_i x_i = 1.$$

Setup: $f_0(x) = \frac{1}{2}\|x - z\|_2^2$, $f_i(x) = -x_i$ ($m = n$ inequality constraints), $h(x) = \sum_i x_i - 1$ ($p = 1$ equality constraint).

Before proving KKT $\Rightarrow$ optimality, let us first see what it means to **solve** the KKT system concretely.

Simplex Projection: KKT Derivation

KKT conditions for (x^*, λ^*, ν^*):

Stationarity: $\nabla_{x_i} L = (x_i^* - z_i) - \lambda_i^* + \nu^* = 0$, so $x_i^* = z_i + \lambda_i^* - \nu^*$.

Complementary slackness: $\lambda_i^* x_i^* = 0$ for all i.

Two cases for each coordinate:

- $x_i^* > 0$: then $\lambda_i^* = 0$, so $x_i^* = z_i - \nu^*$
- $x_i^* = 0$: then $\lambda_i^* = \nu^* - z_i \geq 0$, i.e. $z_i \leq \nu^*$

Threshold form:

$$\boxed{x_i^* = \max\{z_i - \nu^*,\ 0\},}$$

where ν^* is chosen so that $\sum_i \max\{z_i - \nu^*, 0\} = 1$.

Simplex Projection: Threshold Picture

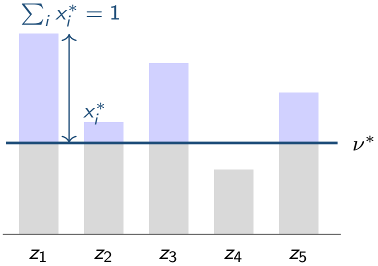
The formula $x_i^* = \max\{z_i - \nu^*, 0\}$ means:

- subtract threshold ν^* from every z_i
- discard coordinates that fall below 0
- choose ν^* so that $\sum_i x_i^* = 1$

The blue caps are the surviving coordinates:

$$\sum_i \max\{z_i - \nu^*, 0\} = 1.$$

This is a soft-thresholding / water-filling operation.



KKT Implies Optimality

Theorem 5.2. If (x^*, λ^*, ν^*) is a KKT point of $\mathcal{P}$, then

- x^* is primal optimal,
- (λ^*, ν^*) is dual optimal,
- $\text{value}(\mathcal{D}) = \text{value}(\mathcal{P})$.

Proof. By stationarity and complementary slackness:

$$q(\lambda^*, \nu^*) = \inf_{x \in C} L(x, \lambda^*, \nu^*) \overset{\text{stat.}}{=} L(x^*, \lambda^*, \nu^*) = f_0(x^*) + \underbrace{\sum_i \lambda_i^* f_i(x^*)}_{\text{CS: } = 0} + \underbrace{\sum_j \nu_j^* h_j(x^*)}_{= 0} = f_0(x^*).$$

Weak duality gives $q(\lambda^*, \nu^*) \leq f_0(x)$ for any feasible x.
So $f_0(x^*) = q(\lambda^*, \nu^*) \leq f_0(x)$: x^* is primal optimal, (λ^*, ν^*) is dual optimal, and the gap is zero. □

Slater's Condition

Definition 5.4 (Slater's condition). $\mathcal{P}$ satisfies *Slater's condition* if there exists $\tilde{x} \in \mathrm{ri}(C)$ such that

$$f_i(\tilde{x}) < 0 \quad \forall\, i = 1, \ldots, m, \qquad h_j(\tilde{x}) = 0 \quad \forall\, j = 1, \ldots, p.$$

Reading:

- The Slater point $\tilde{x}$ satisfies all inequalities **strictly**
- Uses $\mathrm{ri}(C)$ (not $\mathrm{int}(C)$) because the feasible set may live in a lower-dimensional affine subspace
- Slater gives "room to perturb" — exactly the margin that Lecture 4's relative-interior condition needs

Morton Slater and the Cowles Commission

Morton L. Slater
1921–2002
Harvard PhD, 1949

1950. Slater writes *"Lagrange Multipliers Revisited"* as a **Cowles Commission Discussion Paper** at the **University of Chicago**. Never published in a journal.

He shows that strict feasibility (now "Slater's condition") replaces the differentiability assumptions in Kuhn–Tucker, using the minimax theorem of von Neumann and Kakutani.

Like Karush's KKT conditions (1939), Slater's condition (1950) was born at the University of Chicago.

Strong Duality under Slater

Theorem 5.3. If $\mathrm{value}(\mathcal{P}) \in \mathbb{R}$ and $\mathcal{P}$ satisfies Slater's condition, then

$$\mathrm{value}(\mathcal{D}) = \mathrm{value}(\mathcal{P})$$

and $\mathcal{D}$ attains its optimal value.

Proof idea. Package $\mathcal{P}$ into a perturbation $\Phi(x,(u,v)) := f_0(x) + \delta_{\{f_i(x)\le u_i,\, h_j(x)=v_j\}}$.

The Slater point gives a positive margin in every inequality, so $(0_m, 0_p) \in \mathrm{ri}(\mathrm{dom}\, p)$.

Then Theorem 4.4 (4) gives strong duality and dual attainment. $\square$

KKT Existence under Slater

Theorem 5.4. If $\mathrm{value}(\mathcal{P})$ is finite, $\mathcal{P}$ satisfies Slater's condition, and x^* is primal optimal, then there exist $\lambda^* \in \mathbb{R}^m_+$, $\nu^* \in \mathbb{R}^p$ such that (x^*, λ^*, ν^*) is a KKT point of $\mathcal{P}$.

Proof idea. By Theorem 5.3, the dual attains at some (λ^*, ν^*). Tightness of weak duality:

$$f_0(x^*) = \mathrm{value}(\mathcal{P}) = \mathrm{value}(\mathcal{D}) = q(\lambda^*, \nu^*).$$

This forces:

- $L(x^*, \lambda^*, \nu^*) = \inf_{x \in C} L(x, \lambda^*, \nu^*)$ (stationarity)
- $\lambda_i^* f_i(x^*) = 0$ for all i (complementary slackness) $\square$

How the Notions Fit Together

Three logically distinct questions:

1. Does the **primal** attain its optimum?
2. Is the **duality gap** zero?
3. Does the **dual** attain its optimum?

Slater + finite value $\Longrightarrow$ no gap + dual attainment (Theorem 5.3)

If also **primal attainment**: $\Longrightarrow$ KKT point exists (Theorem 5.4)

Slater does **not** guarantee primal attainment — that is a separate topological property (e.g. coercivity, compactness).

The counterexamples ahead show that each implication arrow is strict.

Counterexample: Slater Is Not Necessary

Example 5.2. $\min_{x\in\mathbb{R}} x^2$ s.t. $x^2 \le 0$.

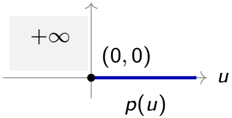
Feasible set: $\{0\}$. $x^* = 0$, $p^* = 0$.

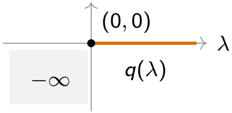
Slater fails: no $\tilde{x}$ with $\tilde{x}^2 < 0$.

But: $q(\lambda) = \inf_x\{(1+\lambda)x^2\} = 0$ for all $\lambda \ge 0$.

$d^* = 0 = p^*$: **no gap**, dual attained, KKT exists.

Slater is sufficient, not necessary.





Counterexample: No Gap $\neq$ Dual Attainment

Example 5.3. $\min_{x\in\mathbb{R}} x \quad \text{s.t. } x^2 \leq 0.$

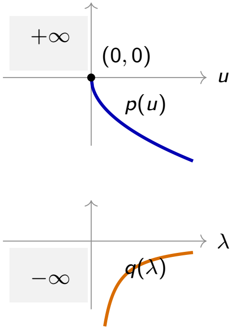
$x^* = 0$, $p^* = 0$. Slater fails.

$q(0) = -\infty$. $q(\lambda) = -\frac{1}{4\lambda}$ for $\lambda > 0$.

$\sup_{\lambda\geq 0} q(\lambda) = 0 = p^*$: **no gap**.

But the sup is **not attained**: $q(\lambda) < 0$ for all $\lambda > 0$.

No gap + primal attainment $\not\Rightarrow$ KKT.



Counterexample: Slater Does Not Guarantee KKT

Example 5.4. $\inf_{x,t\in\mathbb{R}} \; t \quad \text{s.t. } e^{-x} \le t.$

Slater holds (e.g. $(0,2)$).

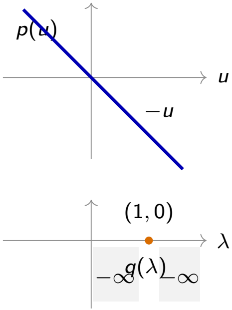
$p^* = 0$ but **not attained**: $e^{-x} > 0$.

$q(\lambda) = 0$ if $\lambda = 1$, else $-\infty$.

$d^* = 0 = p^*$: no gap, dual attained at $\lambda^* = 1$.

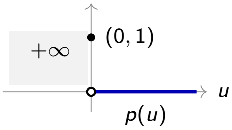
But **no KKT point**: no primal optimizer.

Slater gives gap $= 0$ and dual attainment,
but not primal attainment $\Rightarrow$ no KKT.



Counterexample: Slater Failure Creates a Gap

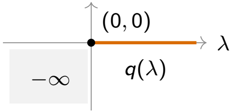
Example 5.5. $C = \{(x, y) : y > 0\}$, $f_0 = e^{-x}$, $f_1 = x^2/y$.
$\min\{e^{-x} : (x, y) \in C,\ x^2/y \le 0\}$.



Feasible: $\{(0, y) : y > 0\}$. $p^* = 1$.

Slater fails: $x^2/y = 0$ only at $x = 0$, not < 0.

$q(\lambda) = \inf_{(x,y)\in C}\{e^{-x} + \lambda x^2/y\} = 0$ for all $\lambda \ge 0$.



$d^* = 0 < 1 = p^*$: **positive duality gap**.

Slater failure can create a gap that no dual certificate can close.

Counterexample: No Gap $\neq$ Primal or Dual Attainment

Example 5.6. Direct product of Ex. 5.4 and 5.3:

$$\inf_{x,t,y\in\mathbb{R}} (t+y) \quad \text{s.t.} \quad e^{-x} \leq t, \quad y^2 \leq 0.$$

$y^2 \leq 0$ forces $y = 0$, reducing to $\inf t$ s.t. $e^{-x} \leq t$.

$p^* = 0$, but primal infimum **not attained**.

$q(\lambda, \mu) = -\frac{1}{4\mu}$ for $\lambda = 1$, $\mu > 0$; else $-\infty$.

$\sup q = 0 = p^*$: **no gap**, but dual supremum **not attained**.

Finite value + no gap $\not\Rightarrow$ attainment on either side.

Theorems and Examples at a Glance

		Slater	finite	prim.	no gap	dual	KKT
Thm 5.2	KKT $\Rightarrow$ all	–	C	C	C	C	A
Thm 5.3	Slater $\Rightarrow$ gap+dual	A	A	–	C	C	–
Thm 5.4	+prim. att. $\Rightarrow$ KKT	A	A	A	–	–	C
Ex 5.2	Slater not needed	$\times$	✓	✓	✓	✓	✓
Ex 5.3	gap $\neq$ dual att.	$\times$	✓	✓	✓	$\times$	$\times$
Ex 5.4	Slater $\neq$ KKT	✓	✓	$\times$	✓	✓	$\times$
Ex 5.5	Slater fail $\Rightarrow$ gap	$\times$	✓	✓	$\times$	✓	$\times$
Ex 5.6	gap $\neq$ any att.	$\times$	✓	$\times$	✓	$\times$	$\times$

A = assumption, C = conclusion, ✓ = holds, $\times$ = fails.

Summary & What's Next

Today:

- Convex constrained program $\mathcal{P}$ and its Lagrangian dual $\mathcal{D}$
- Weak duality: $q(\lambda, \nu) \leq f_0(x)$
- KKT conditions: primal + dual feasibility, stationarity, complementary slackness
- KKT $\Rightarrow$ optimality (Theorem 5.2)
- Slater + finite value $\Rightarrow$ strong duality + dual attainment (Theorem 5.3)
- Slater + primal attainment $\Rightarrow$ KKT existence (Theorem 5.4)
- Simplex projection: KKT gives threshold form $x_i^* = \max\{z_i - \nu^*, 0\}$

Next lecture:

- Cutting-plane methods
- From duality certificates to algorithmic certificates