Lecture 5: Lagrange Multipliers and KKT

Lecture 4 explained where dual variables come from: they arise from perturbation value functions and give lower bounds on the primal value. Lecture 5 asks when those lower bounds become exact certificates of constrained optimality. The story has three steps: weak duality says every dual-feasible multiplier gives a lower bound; KKT says a primal-dual triple is an exact certificate once the lower bound is tight; Slater's condition explains when this certificate system is complete.

5.1 Lagrangian Duality

Definition 5.1 (Convex constrained program)**.** Let E be a finite-dimensional real vector space. A convex constrained program is an optimization problem of the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & x \in C, \\ & f_i(x) \le 0 \quad \forall i \in \{1, \dots, m\}, \\ & h_j(x) = 0 \quad \forall j \in \{1, \dots, p\}, \end{aligned}$$

where $C \subseteq E$ is nonempty and convex, the functions $f_0, f_1, \dots, f_m : C \to \mathbb{R}$ are convex, and the functions $h_1, \dots, h_p : C \to \mathbb{R}$ are affine. We denote this program by P, and its optimal value by $\text{value}(P)$.

Definition 5.2 (Lagrangian, dual function, and dual program)**.** Let P be a convex constrained program as in Definition 5.1. For $x \in C$, $\lambda \in \mathbb{R}^m_+$, and $\nu \in \mathbb{R}^p$, define

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

The dual function is

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

The dual program associated with P, denoted by D, is

$$\begin{aligned} \text{maximize} \quad & q(\lambda, \nu) \\ \text{subject to} \quad & \lambda \in \mathbb{R}^m_+, \\ & \nu \in \mathbb{R}^p. \end{aligned}$$

The value of D is denoted by $\text{value}(D)$.

We write $0_m \in \mathbb{R}^m$ and $0_p \in \mathbb{R}^p$ for the zero vectors in the perturbation coordinates, so the origin of $\mathbb{R}^m \times \mathbb{R}^p$ is denoted by $(0_m, 0_p)$.

Theorem 5.1 (Weak duality)**.** *Let P be a convex constrained program as in Definition 5.1. If $x \in C$ is primal feasible and $(\lambda, \nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$, then*

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

Consequently,

$$\text{value}(D) \leq \text{value}(P).$$

Theorem 5.1 says that every dual-feasible multiplier produces a valid lower bound on the primal objective value. This is the basic certificate statement for constrained convex optimization. In this sense, Theorem 5.1 is the nonlinear continuation of LP weak duality from Lecture 3. The underlying geometry already appeared in Lecture 2: a supporting hyperplane or subgradient gives a valid lower bound, and here the multipliers play the same certifying role for constrained problems.

Proof of Theorem 5.1. Let $x \in C$ be primal feasible and let $(\lambda, \nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$. Because $f_i(x) \leq 0$ for all i and $h_j(x) = 0$ for all j,

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

Since $q(\lambda, \nu) = \inf_{y \in C} L(y, \lambda, \nu) \leq L(x, \lambda, \nu)$, we obtain

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

Taking the supremum over $(\lambda, \nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$ on the left and the infimum over primal-feasible $x \in C$ on the right gives the final inequality. □

5.2 KKT as Optimality Certificates

Definition 5.3 (KKT point)**.** Let P be a convex constrained program as in Definition 5.1. A triple $(x^\star, \lambda^\star, \nu^\star)$ is a KKT point of P if

1. $x^\star \in C$, $f_i(x^\star) \leq 0$ for all i, and $h_j(x^\star) = 0$ for all j;
2. $\lambda^\star \in \mathbb{R}^m_+$;
3. $x^\star \in \arg\min_{x \in C} L(x, \lambda^\star, \nu^\star)$;
4. $\lambda_i^\star f_i(x^\star) = 0$ for all $i \in \{1, \ldots, m\}$.

The third item is the stationarity condition in its clean optimization form: once the multipliers are fixed, the primal point must minimize the Lagrangian over C. Because $\lambda^\star \in \mathbb{R}^m_+$, the map $x \mapsto L(x, \lambda^\star, \nu^\star)$ is convex on C. Thus the argmin condition is exactly the right general replacement for a bare $\nabla_x L = 0$ equation; in the Euclidean unconstrained case $C = E$, it reduces to the familiar first-order stationarity condition via Theorem 1.5.

Before using KKT as a theorem, it is useful to see what it means to solve the KKT system in a concrete problem.

Example 5.1 (Projection onto the simplex)**.** Fix $z \in \mathbb{R}^n$, and consider the Euclidean projection problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|x - z\|_2^2 \qquad \text{subject to} \qquad x_i \geq 0 \ \forall i, \qquad \sum_{i=1}^n x_i = 1.$$

Equivalently, we project z onto the simplex

$$\Delta_n := \left\{ x \in \mathbb{R}^n : x_i \geq 0 \ \forall i, \ \sum_{i=1}^n x_i = 1 \right\}.$$

Write the inequality constraints as $f_i(x) = -x_i \leq 0$, and the equality constraint as $h(x) = \sum_{i=1}^n x_i - 1 = 0$. The Lagrangian is

$$L(x, \lambda, \nu) = \frac{1}{2}\|x - z\|_2^2 - \sum_{i=1}^n \lambda_i x_i + \nu \left(\sum_{i=1}^n x_i - 1 \right).$$

If $(x^\star, \lambda^\star, \nu^\star)$ is a KKT point, then stationarity gives

$$x_i^\star - z_i - \lambda_i^\star + \nu^\star = 0 \qquad \forall i.$$

Together with complementary slackness,

$$\lambda_i^\star x_i^\star = 0, \qquad \lambda_i^\star \geq 0, \qquad x_i^\star \geq 0,$$

this implies

$$x_i^\star > 0 \Longrightarrow \lambda_i^\star = 0 \Longrightarrow x_i^\star = z_i - \nu^\star,$$

while

$$x_i^\star = 0 \Longrightarrow \lambda_i^\star = \nu^\star - z_i \geq 0 \Longrightarrow z_i - \nu^\star \leq 0.$$

Hence any KKT point must satisfy the threshold form

$$x_i^\star = \max\{z_i - \nu^\star, 0\} \qquad \forall i.$$

Finally, primal feasibility imposes

$$\sum_{i=1}^n x_i^\star = 1,$$

so $\nu^\star$ must satisfy

$$\sum_{i=1}^n \max\{z_i - \nu^\star, 0\} = 1.$$

Conversely, if $\nu^\star$ satisfies this equation and we define

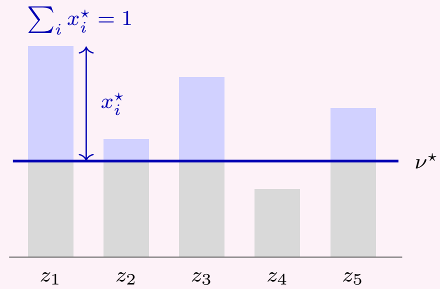
$$x_i^\star := \max\{z_i - \nu^\star, 0\}, \qquad \lambda_i^\star := \max\{\nu^\star - z_i, 0\},$$

then the KKT conditions are satisfied. So solving the KKT system leads directly to the standard threshold formula for simplex projection.

Equivalently, one subtracts a common threshold $\nu^\star$ from every coordinate of z, truncates negative values to 0, and chooses $\nu^\star$ so that the surviving mass sums to 1:

$$\sum_{i=1}^n \max\{z_i - \nu^\star, 0\} = 1.$$

The picture is:



Theorem 5.2 (KKT $\Rightarrow$ strong duality and primal/dual attainment)**.** *Let P be a convex constrained program as in Definition 5.1. If $(x^\star, \lambda^\star, \nu^\star)$ is a KKT point of P, then $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal for D, and*

$$\text{value}(D) = \text{value}(P).$$

Theorem 5.2 says that once a primal-dual triple satisfies the KKT system, the lower bound from Theorem 5.1 becomes exact. In other words, KKT is already an optimality certificate.

Proof of Theorem 5.2. Let $(x^\star, \lambda^\star, \nu^\star)$ be a KKT point of P. Because $x^\star \in \arg\min_{x\in C} L(x, \lambda^\star, \nu^\star)$, one has

$$L(x^\star, \lambda^\star, \nu^\star) = q(\lambda^\star, \nu^\star).$$

Because $x^\star$ is primal feasible and complementary slackness holds,

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

Hence $q(\lambda^\star, \nu^\star) = f_0(x^\star)$. Now fix any primal-feasible $x \in C$. By Theorem 5.1,

$$f_0(x^\star) = q(\lambda^\star, \nu^\star) \le f_0(x).$$

Therefore $x^\star$ is primal optimal. Since $q(\lambda^\star, \nu^\star) = f_0(x^\star) = \text{value}(P)$, the pair $(\lambda^\star, \nu^\star)$ is dual optimal for D, and $\text{value}(D) = \text{value}(P)$. □

5.3 Slater and Exactness

Definition 5.4 (Slater condition)**.** We say that a convex constrained program P defined in Definition 5.1 satisfies Slater's condition if there exists $\widetilde{x} \in \text{ri}(C)$ such that

$$f_i(\widetilde{x}) < 0 \quad \forall i \in \{1, \dots, m\} \qquad \text{and} \qquad h_j(\widetilde{x}) = 0 \quad \forall j \in \{1, \dots, p\}.$$

The relative interior appears here because the feasible geometry may live in a lower-dimensional affine set, so ordinary interior may be empty for ambient reasons even when the problem still has room to move. What Slater gives is not just one feasible point, but a feasible point with strict slack in the inequality directions. This margin is exactly what lets us perturb the right-hand sides slightly while staying feasible. This is where Lecture 5 begins to separate from Lecture 3. In LP, polyhedral structure gave strong duality under much weaker hypotheses; in general convex optimization, Lecture 4 says that exact duality is governed by the perturbation value function, and Slater is the condition that puts the base point of that value function in the relative interior of its domain.

Theorem 5.3 (Slater + finite primal value ⇒ strong duality and dual attainment)**.** *Let P be a convex constrained program as in Definition 5.1. Assume that* $\mathrm{value}(P) \in \mathbb{R}$ *and that P satisfies Slater's condition in the sense of Definition 5.4. Then*

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

Moreover, D attains its optimal value.

Theorem 5.3 answers the next question: when is the lower-bound system complete? Slater's condition is exactly the qualification that lets the perturbation argument from Lecture 4 certify that there is no duality gap and that the dual optimum is attained.

Remark 5.1 (How the perturbation proof recovers the dual function)*.* The proof of Theorem 5.3 packages the constrained problem into the perturbation value function

$$p(u, v) := \inf \{ f_0(x) : x \in C,\ f_i(x) \le u_i\ \forall i,\ h_j(x) = v_j\ \forall j \} .$$

If we write

$$F(x) := (f_1(x), \ldots, f_m(x)), \qquad h(x) := (h_1(x), \ldots, h_p(x)),$$

then the dual function from Definition 5.2 is exactly the sign-flipped conjugate-side object from Lecture 4:

$$\forall \lambda \in \mathbb{R}^m_+,\ \forall \nu \in \mathbb{R}^p, \qquad q(\lambda, \nu) = -p^*(-\lambda, -\nu) = \inf_{u \in \mathbb{R}^m,\ v \in \mathbb{R}^p} (p(u, v) + \lambda^\top u + \nu^\top v).$$

Indeed,

$$p^*(-\lambda, -\nu) = \sup \left\{ -\lambda^\top u - \nu^\top v - f_0(x) : x \in C,\ f_i(x) \le u_i\ \forall i,\ h_j(x) = v_j\ \forall j \right\} .$$

When $\lambda \ge 0$, the supremum in the u-variables is attained at $u_i = f_i(x)$, while the v-variables are forced to equal $h(x)$. Hence

$$p^*(-\lambda, -\nu) = \sup_{x \in C} \left\{ -f_0(x) - \lambda^\top F(x) - \nu^\top h(x) \right\} = -q(\lambda, \nu).$$

If some $\lambda_i < 0$, then the coefficient of u_i in $p^*(-\lambda, -\nu)$ becomes positive; since u_i only has the lower bound $u_i \ge f_i(x)$, the supremum is $+\infty$, so $-p^*(-\lambda, -\nu) = -\infty$. This is why the dual problem naturally restricts to $\lambda \in \mathbb{R}^m_+$. So the Lagrangian dual function is not a separate construction unrelated to marginal duality; it is the same conjugate picture, written in the sign convention natural for inequality multipliers.

Proof idea of Theorem 5.3. Package the constrained problem into its perturbation value function $p(u, v)$, where u records the inequality right-hand sides and v records the equality right-hand sides. Now let $\tilde{x}$ be a Slater point. Because $\tilde{x} \in \mathrm{ri}(C)$ and each f_i is continuous there, the strict inequalities $f_i(\widetilde{x}) < 0$ persist on a relative neighborhood $U_0 \subseteq C$: there is some margin $\eta > 0$ such that $f_i(y) < -\eta$ for every $y \in U_0$ and every i. This means that all perturbations u with $u_i > -\eta$ remain feasible somewhere near $\widetilde{x}$. For the equality block, the map $h : \operatorname{aff}(C) \to \mathbb{R}^p$ is affine, hence open onto its image. Therefore the set $h(U_0)$ contains a relative neighborhood W of 0_p in $h(\operatorname{aff} C)$. So every pair (u, v) with $u \in \prod_i(-\eta, \infty)$ and $v \in W$ still belongs to $\operatorname{dom} p$. Since $\operatorname{dom} p \subseteq \mathbb{R}^m \times h(\operatorname{aff} C)$, this is exactly the statement that $(0_m, 0_p) \in \operatorname{ri}(\operatorname{dom} p)$. Then Theorem 4.4 applies at the origin, and the rest of the proof is the computation of the resulting dual objective.

Theorem 5.4 (Slater + primal attainment ⇒ KKT existence)**.** *Let P be a convex constrained program as in Definition 5.1. Assume that* $\operatorname{value}(P)$ *is finite and that P satisfies Slater's condition in the sense of Definition 5.4. If $x^\star$ is primal optimal, then there exist $\lambda^\star \in \mathbb{R}^m_+$ and $\nu^\star \in \mathbb{R}^p$ such that $(x^\star, \lambda^\star, \nu^\star)$ is a KKT point of P.*

Theorem 5.4 is the converse direction under Slater. Once a dual maximizer exists, equality of the primal and dual values forces equality in the weak-duality estimate, and that equality is precisely what turns lower-bound certificates into the full KKT system.

Proof idea of Theorem 5.4. Take a primal optimizer $x^\star$ and a dual optimizer $(\lambda^\star, \nu^\star)$ from Theorem 5.3. Because the primal and dual values agree, the weak-duality inequality is tight at $(x^\star, \lambda^\star, \nu^\star)$. Tightness forces both complementary slackness and stationarity, so the triple is a KKT point.

Remark 5.2 (How the notions fit together). At this point it is useful to separate three logically distinct questions: primal attainment, no duality gap, and dual attainment. The combined certificate statement is

$$\text{KKT existence} \iff \text{primal attainment} + \text{no gap} + \text{dual attainment}.$$

The forward direction is Theorem 5.2, and the converse follows by equality in the weak-duality chain.
Slater is not one of these layers; it is a regularity condition. In the present lecture, its role is

$$\text{Slater} + \text{finite primal value} \implies \text{no gap} + \text{dual attainment}.$$

So if one also has primal attainment, Theorem 5.4 gives KKT existence.

5.4 Failure Modes

Example 5.2 (Slater is not necessary)**.** Consider

$$\min_{x \in \mathbb{R}} x^2 \qquad \text{subject to} \qquad x^2 \leq 0.$$

The feasible set is $\{0\}$, so the primal optimum is attained at $x^\star = 0$. Slater's condition fails, because the strict inequality $x^2 < 0$ is impossible. Nevertheless,

$$q(\lambda) = \inf_{x \in \mathbb{R}} (1 + \lambda)x^2 = 0 \qquad \forall \lambda \geq 0.$$

Its perturbation value function is

$$p(u) = \begin{cases} 0, & u \geq 0, \\ +\infty, & u < 0. \end{cases}$$

So there is no duality gap, the dual optimum is attained, and KKT points do exist. This example shows that Slater is sufficient, not necessary.

Example 5.3 (No gap does not imply dual attainment)**.** Consider

$$\min_{x \in \mathbb{R}} x \qquad \text{subject to} \qquad x^2 \leq 0.$$

Again the feasible set is $\{0\}$, so the primal optimum is attained at $x^\star = 0$, and again Slater's condition fails. The Lagrangian is

$$L(x, \lambda) = x + \lambda x^2, \qquad \lambda \geq 0.$$

Hence

$$q(0) = -\infty, \qquad q(\lambda) = \inf_{x \in \mathbb{R}} (x + \lambda x^2) = -\frac{1}{4\lambda} \qquad \forall \lambda > 0.$$

Therefore

$$\sup_{\lambda \geq 0} q(\lambda) = 0 = p^\star,$$

and the perturbation value function is

$$p(u) = \begin{cases} -\sqrt{u}, & u \geq 0, \\ +\infty, & u < 0. \end{cases}$$

so there is no duality gap, but the dual optimum is not attained. Consequently there is no KKT point. This example shows that primal attainment together with no gap is still not enough for KKT existence: one also needs dual attainment.

Example 5.4 (Slater does not guarantee KKT existence)**.** Consider

$$\inf_{x, t \in \mathbb{R}} t \qquad \text{subject to} \qquad e^{-x} \leq t.$$

Slater's condition holds; for example, $(x, t) = (0, 2)$ is strictly feasible. The primal value is 0, but it is not attained, because $e^{-x} > 0$ for every finite x. The Lagrangian is

$$L(x, t, \lambda) = t + \lambda(e^{-x} - t) = \lambda e^{-x} + (1 - \lambda)t, \qquad \lambda \geq 0.$$

Hence

$$q(\lambda) = \begin{cases} 0, & \lambda = 1, \\ -\infty, & \lambda \neq 1. \end{cases}$$

The perturbation value function is

$$p(u) = -u \qquad \forall u \in \mathbb{R},$$

but the infimum is not attained for any $u \in \mathbb{R}$. So there is no duality gap and the dual optimum is attained at $\lambda^\star = 1$, but there is no KKT point because there is no primal optimizer. This example shows that Slater guarantees the duality-side conclusions, but not primal attainment.

Example 5.5 (Failure of Slater can create a duality gap)**.** Let

$$C := \left\{(x, y) \in \mathbb{R}^2 : y > 0\right\}, \qquad f_0(x, y) := e^{-x}, \qquad f_1(x, y) := \frac{x^2}{y}.$$

Consider

$$\min \left\{ e^{-x} : (x, y) \in C,\ \frac{x^2}{y} \leq 0 \right\}.$$

The feasible set is $\{(0, y) : y > 0\}$, so the primal optimum is attained and equals

$$p^\star = 1.$$

Slater's condition fails, because $\frac{x^2}{y} < 0$ is impossible on C. For every $\lambda \geq 0$,

$$q(\lambda) = \inf_{(x,y)\in C} \left(e^{-x} + \lambda \frac{x^2}{y} \right).$$

For fixed x, the infimum over $y > 0$ of $\lambda x^2/y$ is 0, so

$$q(\lambda) = \inf_{x \in \mathbb{R}} e^{-x} = 0.$$

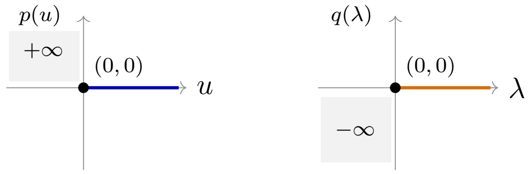
The perturbation value function is

$$p(u) = \begin{cases} +\infty, & u < 0, \\ 1, & u = 0, \\ 0, & u > 0. \end{cases}$$

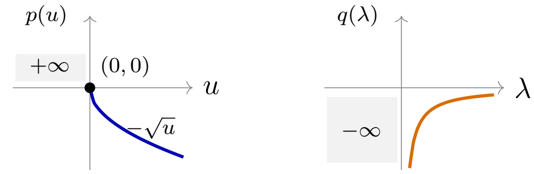
Thus

$$d^\star = 0 < 1 = p^\star.$$

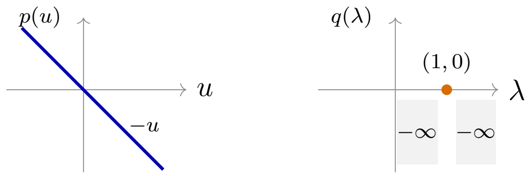
So failure of Slater can indeed produce a positive duality gap.



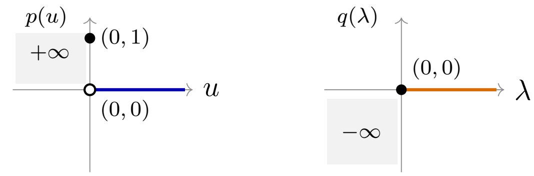
(a) Example 5.2: $p(u)$ and $q(\lambda)$



(b) Example 5.3: $p(u)$ and $q(\lambda)$



(c) Example 5.4: $p(u)$ and $q(\lambda)$



(d) Example 5.5: $p(u)$ and $q(\lambda)$

Figure 1: Value functions and dual functions for Examples 5.2 to 5.5. Filled dots indicate values attained by the function at boundary points; hollow dots indicate limiting branches not taken by the function.

Example 5.6 (No gap does not imply primal or dual attainment)**.** This example is the direct product of Example 5.4 and Example 5.3: the first block contributes failure of primal attainment, and the second contributes failure of dual attainment. Consider

$$\inf_{x,t,y\in\mathbb{R}} (t+y) \qquad \text{subject to} \qquad e^{-x} \le t, \qquad y^2 \le 0.$$

The second constraint forces $y = 0$, so the primal problem reduces to

$$\inf_{x,t\in\mathbb{R}} t \qquad \text{subject to} \qquad e^{-x} \le t.$$

Hence

$$p^\star = 0,$$

but the primal infimum is not attained. The perturbation value function in the two right-hand-side variables $(u, v) \in \mathbb{R}^2$ is

$$p(u,v) = \begin{cases} -u - \sqrt{v}, & v \ge 0, \\ +\infty, & v < 0. \end{cases}$$

Equivalently, if $p_{\exp}$ and $p_{\sqrt{\cdot}}$ are the value functions from Examples 5.3 and 5.4, then

$$p(u,v) = p_{\exp}(u) + p_{\sqrt{\cdot}}(v).$$

Likewise the dual function splits additively:

$$q(\lambda,\mu) = q_{\exp}(\lambda) + q_{\sqrt{\cdot}}(\mu),$$

with $q_{\exp}$ and $q_{\sqrt{\cdot}}$ taken from Examples 5.3 and 5.4. Explicitly,

$$q(\lambda,\mu) = \begin{cases} -\dfrac{1}{4\mu}, & \lambda = 1,\ \mu > 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

Therefore

$$\sup_{\lambda,\mu\ge 0} q(\lambda,\mu) = \sup_{\mu>0} \left(-\frac{1}{4\mu}\right) = 0 = p^\star,$$

but the supremum is not attained. Thus this is a convex problem with finite value and no duality gap, yet neither the primal optimum nor the dual optimum is attained.

Table 1: Theorems and examples at a glance

Item	Note	Slater	finite	prim.	no gap	dual	KKT
Thm. 5.2	KKT => no gap + attainment on both sides.	–	C	C	C	C	A
Thm. 5.3	Slater + finite value => no gap + dual attainment.	A	A	–	C	C	–
Thm. 5.4	Slater + primal attainment => KKT existence.	A	A	A	–	–	C
Thms. 5.4+5.2	Slater + finite value + primal attainment => all three conclusions.	A	A	A	C	C	C
Ex. 5.1	Slater is sufficient, not necessary.	×	✓	✓	✓	✓	✓
Ex. 5.2	No gap + primal attainment still do not imply dual attainment.	×	✓	✓	✓	×	×
Ex. 5.3	Slater still does not guarantee primal attainment.	✓	✓	×	✓	✓	×
Ex. 5.4	Failure of Slater can create a positive duality gap.	×	✓	✓	×	✓	×
Ex. 5.5	No gap can coexist with failure of both attainments.	×	✓	×	✓	×	×

In the theorem rows, A means "assumption" and C means "conclusion." The fourth theorem row records the consequence of combining Theorem 5.4 and Theorem 5.2.

Dependency and proof sketch

1. Theorem 5.1 is the base inequality of Lecture 5: every dual-feasible pair gives a lower bound on the primal objective value.
2. Theorem 5.2 turns KKT into an optimality certificate by combining stationarity and complementary slackness with Theorem 5.1.
3. Theorem 5.3 is a direct application of Theorem 4.4. The proof packages the constrained problem into a convex master function $\Phi(x,(u,v))$, verifies Slater as

$$(0_m, 0_p) \in \operatorname{ri}(\operatorname{dom} p),$$

and then computes the resulting dual objective by evaluating $\Phi^*(0,s,r)$.
4. Theorem 5.4 is the converse direction under Slater: once a dual maximizer exists, equality of the primal and dual optima forces stationarity and complementary slackness.

Proofs

Proof of Theorem 5.3. Step 1: build the perturbation function and its value function. Set $U := \mathbb{R}^m \times \mathbb{R}^p$, let

$$F := (f_1, \dots, f_m) : C \to \mathbb{R}^m, \qquad h := (h_1, \dots, h_p) : C \to \mathbb{R}^p,$$

and define

$$\Phi(x,(u,v)) := \begin{cases} f_0(x) + \delta_{\mathbb{R}^m_+}(u - F(x)) + \delta_{\{0_p\}}(h(x) - v), & x \in C, \\ +\infty, & x \notin C. \end{cases}$$

For $(u,v) \in U$, set

$$p(u,v) := \inf_{x \in E} \Phi(x,(u,v)).$$

Then p is exactly the perturbation value function

$$\forall (u,v) \in \mathbb{R}^m \times \mathbb{R}^p, \qquad p(u,v) = \inf \{ f_0(x) : x \in C,\ f_i(x) \le u_i\ \forall i,\ h_j(x) = v_j\ \forall j \}.$$

By construction, Φ is convex.
Set

$$p^\star := \mathrm{value}(P).$$

By definition,

$$p(0_m, 0_p) = p^\star.$$

Step 2: use Slater's condition to show that the origin is a relative interior point of the perturbation domain.
To apply Theorem 4.4(4), it remains to show

$$(0_m, 0_p) \in \mathrm{ri}(\mathrm{dom}\, p).$$

Let $\widetilde{x}$ be a Slater point. Since each $f_i : C \to \mathbb{R}$ is finite-valued and convex, it is continuous on $\mathrm{ri}(C)$ in finite dimensions. Therefore there exist a relative neighborhood $U_0 \subseteq C$ of $\widetilde{x}$ and a number $\eta > 0$ such that

$$f_i(y) < -\eta \qquad \forall y \in U_0,\ \forall i.$$

Each affine function $h_j : C \to \mathbb{R}$ extends uniquely to an affine function on $\mathrm{aff}(C)$; by abuse of notation we denote this extension again by h_j. Now define the affine map

$$h : \mathrm{aff}(C) \to \mathbb{R}^p.$$

Because h is affine, it is continuous and open onto its image $h(\mathrm{aff}(C))$. Since $h(\widetilde{x}) = 0_p$, there exists a relative neighborhood W of 0_p in $h(\mathrm{aff}(C))$ such that

$$W \subseteq h(U_0).$$

Set

$$V := \prod_{i=1}^{m} (-\eta, \infty).$$

Then V is an open neighborhood of 0_m in $\mathbb{R}^m$. Fix $(u,v) \in V \times W$. Because $v \in W \subseteq h(U_0)$, there exists $y \in U_0$ such that $h(y) = v$. For this same y, one has

$$f_i(y) < -\eta < u_i \qquad \forall i,$$

so $\Phi(y, (u,v)) = f_0(y) < +\infty$. Hence $(u,v) \in \mathrm{dom}\, p$. We have proved

$$V \times W \subseteq \mathrm{dom}\, p.$$

Since $\mathrm{dom}\, p \subseteq \mathbb{R}^m \times h(\mathrm{aff}(C))$, this shows $(0_m, 0_p) \in \mathrm{ri}(\mathrm{dom}\, p)$. Since $p(0_m, 0_p) = p^\star \in \mathbb{R}$, Theorem 4.4(4) gives

$$p^\star = \max_{(s,r) \in \mathbb{R}^m \times \mathbb{R}^p} \{ -\Phi^*(0, s, r) \}.$$

Step 3: compute $\Phi^*(0, s, r)$ and recover the usual dual program.
To compute $\Phi^*(0, s, r)$, we return to the original constraint form of Φ:

$$\Phi^*(0, s, r) = \sup \left\{ s^\top u + r^\top v - f_0(x) : x \in C,\ f_i(x) \le u_i\ \forall i,\ h_j(x) = v_j\ \forall j \right\}.$$

If some $s_i > 0$, then u_i has only a lower bound $u_i \geq f_i(x)$, so for any fixed $x \in C$ we may send $u_i \to +\infty$ and force

$$s^\top u + r^\top v - f_0(x) \geq M$$

for arbitrary $M > 0$. Hence $\Phi^*(0, s, r) = +\infty$, so $-\Phi^*(0, s, r) = -\infty$. Therefore only $s \leq 0$ can contribute to the maximum. For such s, the supremum in the u-variables is attained at $u_i = f_i(x)$, and the v-variables are forced to equal $h_j(x)$. Thus

$$\begin{aligned}
\Phi^*(0, s, r) &= \sup_{x \in C} \left\{ \sum_{i=1}^{m} s_i f_i(x) + \sum_{j=1}^{p} r_j h_j(x) - f_0(x) \right\} \\
&= -\inf_{x \in C} \left\{ f_0(x) - \sum_{i=1}^{m} s_i f_i(x) - \sum_{j=1}^{p} r_j h_j(x) \right\} \\
&= -q(-s, -r).
\end{aligned}$$

After the change of variables

$$\lambda := -s \in \mathbb{R}_+^m, \qquad \nu := -r \in \mathbb{R}^p,$$

the maximizing expression becomes exactly $q(\lambda, \nu)$. Hence

$$p^\star = \max_{\lambda \in \mathbb{R}_+^m,\ \nu \in \mathbb{R}^p} q(\lambda, \nu) = \text{value}(D).$$

Since $p^\star = \text{value}(P)$, we conclude

$$\text{value}(D) = p^\star = \text{value}(P),$$

and the maximizing pair (λ, ν) shows that D attains its optimal value. □

Proof of Theorem 5.4. Let $x^\star$ be a primal optimal point. By Theorem 5.3, there exists a dual optimal pair $(\lambda^\star, \nu^\star)$ with

$$\lambda^\star \in \mathbb{R}_+^m, \qquad q(\lambda^\star, \nu^\star) = \text{value}(D) = \text{value}(P) = f_0(x^\star).$$

Because

$$q(\lambda^\star, \nu^\star) = \inf_{x \in C} L(x, \lambda^\star, \nu^\star) \leq L(x^\star, \lambda^\star, \nu^\star),$$

and because $x^\star$ is primal feasible,

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

Hence

$$f_0(x^\star) = q(\lambda^\star, \nu^\star) \leq L(x^\star, \lambda^\star, \nu^\star) \leq f_0(x^\star).$$

Hence all inequalities are equalities. From

$$q(\lambda^\star, \nu^\star) = L(x^\star, \lambda^\star, \nu^\star) = \inf_{x \in C} L(x, \lambda^\star, \nu^\star),$$

we conclude that $x^\star \in \arg\min_{x\in C} L(x, \lambda^\star, \nu^\star)$. Thus the stationarity/minimization condition in the KKT system holds. Finally, the equality $L(x^\star, \lambda^\star, \nu^\star) = f_0(x^\star)$ and primal feasibility imply

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

Each term in this sum is nonpositive because $\lambda_i^\star \geq 0$ and $f_i(x^\star) \leq 0$. Therefore every term must vanish, so

$$\forall i \in \{1, \ldots, m\}, \qquad \lambda_i^\star f_i(x^\star) = 0.$$

So complementary slackness holds. Together with primal feasibility and $\lambda^\star \in \mathbb{R}_+^m$, this proves that $(x^\star, \lambda^\star, \nu^\star)$ is a KKT point of P. □

Exercises.

1. Consider the trust-region quadratic problem

$$\min_{x\in\mathbb{R}^n} \left(g^\top x + \frac{1}{2} \|x\|_2^2 \right) \quad \text{subject to} \quad \|x\|_2 \leq r,$$

where $g \in \mathbb{R}^n$ and $r > 0$. Rewrite the constraint as

$$\frac{1}{2} \|x\|_2^2 \leq \frac{1}{2} r^2,$$

use the KKT conditions to show that the optimizer has the form

$$x^\star = -\frac{g}{1+\lambda^\star}$$

for some $\lambda^\star \geq 0$, and deduce the explicit formula

$$x^\star = \begin{cases} -g, & \|g\|_2 \leq r, \\ -\dfrac{r}{\|g\|_2} g, & \|g\|_2 > r. \end{cases}$$

2. Derive the KKT conditions for

$$\min_{x\in\mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2 \quad \text{subject to} \quad \|x\|_1 \leq \tau.$$

Show that $x^\star$ is optimal if and only if there exists $\lambda^\star \geq 0$ and $g^\star \in \partial \|x^\star\|_1$ such that

$$A^\top (Ax^\star - b) + \lambda^\star g^\star = 0, \qquad \|x^\star\|_1 \leq \tau, \qquad \lambda^\star(\|x^\star\|_1 - \tau) = 0.$$

3. Consider the water-filling problem

$$\max_{x\in\mathbb{R}^n} \sum_{i=1}^{n} \log(a_i + x_i) \quad \text{subject to} \quad x_i \geq 0 \ \forall i, \qquad \sum_{i=1}^{n} x_i = B,$$

where $a_1, \ldots, a_n > 0$ and $B > 0$. Use the KKT conditions to show that there exists $\nu^\star > 0$ such that

$$x_i^\star = \max\left\{ \frac{1}{\nu^\star} - a_i, 0 \right\} \qquad \forall i,$$

and determine how $\nu^\star$ is fixed by the equality constraint.