From Convex Analysis to Learning, Prediction, and Elicitation*
Lecture 3: Lagrange Duality

Lunjia Hu

Linear programs (LPs) are a basic form of constrained optimization problems. Consider the following LP:

$$\underset{x_1,x_2\in\mathbb{R}}{\text{minimize}} \quad 3x_1 + 7x_2 \tag{P0}$$
$$\text{s.t.} \quad x_1 - 3x_2 \le -2, \tag{1}$$
$$-2x_1 + 5x_2 \le 3. \tag{2}$$

Let OPT denote the optimal value of this LP. If someone wants us to prove $\mathsf{OPT} \le 10$, we just need to show them the solution $x_1 = x_2 = 1$. This solution indeed achieves the desired objective $3x_1 + 7x_2 = 10$ while satisfying both constraints. But if someone wants us to prove $\mathsf{OPT} \ge 10$, what should we do?

Proving $\mathsf{OPT} \ge 10$ feels more challenging because it corresponds to proving "something does NOT exist". It corresponds to proving that there does NOT exist (x_1, x_2) achieving $3x_1 + 7x_2 < 10$ while satisfying the two constraints. This is a "non-existence" statement, whereas a proof, by definition, is an "existence" statement (the proof itself is the object that exists).

How do we turn the "non-existence" statement into an "existence" statement? For this specific LP, we can prove $\mathsf{OPT} \ge 10$ using the *existence* of two coefficients $h_1 = 29, h_2 = 16$ satisfying the following identity:

$$3x_1 + 7x_2 + h_1(x_1 - 3x_2 + 2) + h_2(-2x_1 + 5x_2 - 3) = 10 \quad \text{for all } x_1, x_2 \in \mathbb{R}.$$

This identity can be verified using simple calculations, and it indeed shows that $3x_1 + 7x_2 \ge 10$ whenever the two constraints (1) and (2) are satisfied.

Is it always possible to prove the optimum value of an LP in this way? How about convex optimization problems beyond LP? In this lecture we answer these questions using the theory of Lagrange duality. In the example above, $x_1 = x_2 = 1$ is often termed the optimal *primal* solution to the LP, whereas $h_1 = 29, h_2 = 16$ is the optimal *dual* solution. The coefficients h_1, h_2 are also termed *Lagrange multipliers*. We will make sense of these terms in this lecture.

1 Core Duality Theorem

The following theorem lies at the core of essentially every Lagrange duality theorem:

*`https://lunjiahu.com/convex-analysis/`

Theorem 1. *Let $C \subseteq \mathbb{R}^{m+n}$ be a convex set. Suppose the following condition* cannot *be satisfied by any $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$:*

$$(x, y) \in C \quad \textit{and} \quad y = \mathbf{0}. \tag{3}$$

Then there exists a non-zero vector $h \in \mathbb{R}^n$ such that

$$\langle y, h \rangle \geq 0 \quad \textit{for every } (x, y) \in C. \tag{4}$$

If in addition C is the intersection of a convex open set O and a convex polyhedron P, then there exists $h \in \mathbb{R}^n$ such that

$$\langle y, h \rangle > 0 \quad \textit{for every } (x, y) \in C. \tag{5}$$

Proof. Let $Z \subseteq \mathbb{R}^{m+n}$ be the set of vectors $(z, \mathbf{0}_n)$ where $z \in \mathbb{R}^m$ is arbitrary and $\mathbf{0}_n \in \mathbb{R}^n$ is the zero vector. It is clear that Z is a convex polyhedron. Since condition (3) is not satisfiable, we have $C \cap Z = \emptyset$, so by the separating hyperplane theorem, there exists a non-zero vector $(h_1, h) \in \mathbb{R}^m \times \mathbb{R}^n$ such that

$$\langle x, h_1 \rangle + \langle y, h \rangle \geq \langle z, h_1 \rangle \quad \text{for every } (x, y) \in C \text{ and every } z \in \mathbb{R}^m.$$

Since $z \in \mathbb{R}^m$ can be arbitrary, the above inequality can only hold when $h_1 = \mathbf{0}_m$ is the zero vector in $\mathbb{R}^m$. This proves (4).

When C is the intersection of a convex open set and a convex polyhedron, we have strict hyperplane separation: there exists $(h_1, h) \in \mathbb{R}^m \times \mathbb{R}^n$ such that

$$\langle x, h_1 \rangle + \langle y, h \rangle > \langle z, h_1 \rangle \quad \text{for every } (x, y) \in C \text{ and every } z \in \mathbb{R}^m.$$

Again, we must have $h_1 = \mathbf{0}_m$. Thus (5) holds. □

2 Lagrange Duality

We consider a very general convex optimization problem as follows. Let m, n, n' be nonnegative integers. Let $E \subseteq \mathbb{R}^m, S \subseteq \mathbb{R}^{n'}$ be two convex sets and let $f_0, f_1, \ldots, f_n : E \to \mathbb{R}$ be convex functions defined on E. Let $A \in \mathbb{R}^{n' \times m}$ be a matrix. These objects define a general convex optimization problem:

$$\underset{x \in E}{\text{minimize}} \quad f_0(x) \tag{P1}$$

$$\text{s.t.} \quad f_i(x) < 0 \quad \text{for every } i = 1, \ldots, n, \tag{6}$$

$$Ax \in S. \tag{7}$$

The main goal of this section is to prove the following duality theorem:

Theorem 2 (Lagrange duality)**.** *Assume the following regularity conditions hold in the optimization problem* (P1)*:*

1. *E is a convex open set;*

2. *S is the intersection of a convex open set and a convex polyhedron.*

Assume (P1) *is feasible, that is, there exists* $x \in E$ *satisfying the constraints* (6) *and* (7)*. For some threshold* $t \in \mathbb{R}$*, assume that the objective value of* (P1) *cannot go below* t*, that is, no* $x \in E$ *achieves* $f_0(x) < t$ *while satisfying the constraints* (6) *and* (7)*. Then there exist* Lagrange multipliers $h_1, \ldots, h_n \geq 0$ *and* $h' \in \mathbb{R}^{n'}$ *such that*

$$f_0(x) + \sum_{i=1}^{n} h_i f_i(x) + \langle Ax - s, h' \rangle \geq t \quad \textit{for every } x \in E \textit{ and } s \in S. \tag{8}$$

Remark 1. *The converse of Theorem 2 holds trivially: if there exist* $h_1, \ldots, h_n \geq 0$ *and* $h' \in \mathbb{R}^{n'}$ *such that* (8) *holds, then the objective value of* (P1) *cannot go below* t*. Theorem 2 thus finds a* proof *for the fact that the optimal objective value of* (P1) *is at least* t*.*

Remark 2. *Note how Theorem 2 turns a non-existence statement into an existence statement. It follows the same spirit as the hyperplane separation theorem and the minimax theorem:*

If something doesn't exist, then something else must exist.

Proof. Let C be the set of vectors $(x, y_0, y, y') \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n'}$ satisfying the following conditions (where $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$):

$$\begin{aligned} x &\in E, \\ f_0(x) - y_0 &< t, && (9) \\ f_i(x) - y_i &< 0 \quad \text{for every } i = 1, \ldots, n, && (10) \\ Ax - y' &\in S. && (11) \end{aligned}$$

We first show that C can be written as the intersection of an open convex set and a convex polyhedron. Observe that $C = C_0 \cap C_1 \cap \cdots \cap C_n \cap C_{n+1}$, where

$$\begin{aligned} C_0 &= \{(x, y_0, y, y') \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n'} : x \in E \text{ and } f_0(x) - y_0 < t\}, \\ C_i &= \{(x, y_0, y, y') \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n'} : x \in E \text{ and } f_i(x) - y_i < 0\} \quad \text{for every } i = 1, \ldots, n, \\ C_{n+1} &= \{(x, y_0, y, y') \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n'} : Ax - y' \in S\}. \end{aligned}$$

It is a standard result that every convex function $f : E \to \mathbb{R}$ defined on an open convex set $E \subseteq \mathbb{R}^m$ must be continuous. This implies that $C_0, C_1, \ldots, C_n$ are all open. Also, $C_0, C_1, \ldots, C_n$ are all clearly convex. Since $Ax - y'$ is a linear function of (x, y'), our assumption that S is the intersection of an open convex set and a convex polyhedron ensures that C_{n+1} is also the intersection of an open convex set and a convex polyhedron. Note that the intersection of two open convex sets is still an open convex set, and that the intersection of two convex polyhedra is also a convex polyhedron. We can thus combine the properties of $C_0, C_1, \ldots, C_n, C_{n+1}$ and conclude that C is the intersection of an open convex set and a convex polyhedron.

By assumption, the objective value of (P1) cannot go below t, so there is no $(x, y_0, y, y') \in C$ satisfying $y_0 = 0, y = \mathbf{0}_n, y' = \mathbf{0}_{n'}$. By Theorem 1, there exist $h_0 \in \mathbb{R}, h \in \mathbb{R}^n, h' \in \mathbb{R}^{n'}$ such that

$$y_0 h_0 + \langle y, h \rangle + \langle y', h' \rangle > 0 \quad \text{for every } (x, y_0, y, y') \in C. \tag{12}$$

For every $(x, y_0, y, y') \in C$, we can arbitrarily increase y_0 and every coordinate of y without violating the inequalities (9) and (10). Thus, after increasing y_0 and y, we still have $(x, y_0, y, y') \in C$ and (12) must still be satisfied. This is possible only when $h_0 \geq 0$ and $h \geq \mathbf{0}$.

Now we prove $h_0 > 0$. By our assumption that (P1) is feasible, there exists $(x, y_0, y, y') \in C$ such that $y = \mathbf{0}$ and $y' = \mathbf{0}$. Plugging it into (12), we get $y_0 h_0 > 0$, which implies that $h_0 \neq 0$. Since we have shown $h_0 \geq 0$, we now have $h_0 > 0$.

By scaling h_0, h, h' using the same positive factor, we can assume without loss of generality that $h_0 = 1$.

For arbitrary $x \in E, s \in S$ and $\varepsilon > 0$, let us choose

$$y_0 := f_0(x) - t + \varepsilon, \quad \text{and} \tag{13}$$

$$y_i := f_i(x) + \varepsilon \quad \text{for every } i = 1, \ldots, n, \tag{14}$$

$$y' := Ax - s. \tag{15}$$

These choices ensure that (9), (10), and (11) are satisfied, so $(x, y_0, y, y') \in C$. Plugging (13), (14), and (15) into (12), we get

$$f_0(x) - t + \varepsilon + \sum_{i=1}^{n} h_i(f_i(x) + \varepsilon) + \langle Ax - s, h' \rangle > 0.$$

Sending $\varepsilon \to 0$ proves (8). $\square$

3 Slater's Condition

The convex optimization problems we get often do not exactly have the form of (P1). Nevertheless, Theorem 2 still applies as long as they satisfy *Slater's condition*, which we discuss in this section.

For example, many optimization problems do not have the strict inequalities in (6). Instead, they have non-strict inequalities. The domain of x is often not an open convex set E, but rather is the *closure* $\mathsf{cl}\, E$ of some open convex set E. Similarly, the constraint (7) is often not given by a set S that is the intersection of an open convex set and a convex polyhedron. Instead, it is given by the closure $\mathsf{cl}\, S$ of such S. In summary, we have the following optimization problem:

$$\underset{x \in \mathsf{cl}\, E}{\text{minimize}} \quad f_0(x) \tag{P2}$$

$$\text{s.t.} \quad f_i(x) \leq 0 \quad \text{for every } i = 1, \ldots, n, \tag{16}$$

$$Ax \in \mathsf{cl}\, S. \tag{17}$$

Definition 1 (Slater's condition)**.** *In* (P2)*, assume*

1. *$E \subseteq \mathbb{R}^m$ is a convex open set;*
2. *$S \subseteq \mathbb{R}^{n'}$ is the intersection of a convex open set and a convex polyhedron;*
3. *$f_0, f_1, \ldots, f_n : \mathsf{cl}\, E \to \mathbb{R}$ are convex functions defined on $\mathsf{cl}\, E$;*
4. *$A \in \mathbb{R}^{n' \times m}$.*

We say (P2) *satisfies* Slater's condition *if the corresponding problem* (P1) *is feasible: there exists $x \in E$ that satisfies the two constraints* (6) *and* (7) *of* (P1).

Note that Slater's condition is stronger than the feasibility of (P2), which would correspond to the existence of $x \in \mathsf{cl}\, E$ satisfying the two constraints (16) and (17) of (P2).

We have the following Lagrange duality theorem for (P2) under Slater's condition:

Theorem 3 (Lagrange duality under Slater's condition)**.** *Assume* (P2) *satisfies Slater's condition (as well as the four basic assumptions in Definition 1). For some threshold $t \in \mathbb{R}$, assume that the objective value of* (P2) *cannot go below t, that is, no $x \in \mathsf{cl}\, E$ achieves $f_0(x) < t$ while satisfying the constraints* (16) *and* (17)*. Then there exist $h_1, \ldots, h_n \geq 0$ and $h' \in \mathbb{R}^{n'}$ such that*

$$f_0(x) + \sum_{i=1}^{n} h_i f_i(x) + \langle Ax - s, h' \rangle \geq t \quad \text{for every } x \in \mathsf{cl}\, E \text{ and } s \in \mathsf{cl}\, S. \tag{18}$$

Proof. Our assumption that the objective value of (P2) cannot go below t implies that the objective value of (P1) also cannot go below t. By Theorem 2, there exist $h_1, \ldots, h_n \geq 0$ and $h' \in \mathbb{R}^{n'}$ such that (8) holds. That is,

$$f_0(x) + \sum_{i=1}^{n} h_i f_i(x) + \left(\langle Ax, h' \rangle - \sup_{s \in S} \langle s, h' \rangle \right) \geq t \quad \text{for every } x \in E. \tag{19}$$

Since $\mathsf{cl}\, S$ is the closure of S, we have $\sup_{s \in S} \langle s, h' \rangle = \sup_{s \in \mathsf{cl}\, S} \langle s, h' \rangle$. Thus, to prove (18), it remains to prove that (19) holds not just for every $x \in E$, but also for every $x \in \mathsf{cl}\, E$. Note that the left-hand side of (19) is a convex function of $x \in \mathsf{cl}\, E$, so by Lemma 4 below, (19) must hold for every $x \in \mathsf{cl}\, E$ as well. □

Lemma 4. *Let E be a non-empty convex open set and let $f : \mathsf{cl}\, E \to \mathbb{R}$ be a convex function defined on the closure $\mathsf{cl}\, E$. For some $t \in \mathbb{R}$, if $f(x) \geq t$ holds for every $x \in E$, then it holds for every $x \in \mathsf{cl}\, E$ as well.*

Proof. Consider an arbitrary $x_0 \in E$. For every $x \in \mathsf{cl}\, E$ and every $\alpha \in (0, 1)$, we have $\alpha x + (1 - \alpha) x_0 \in E$, so

$$t \leq f(\alpha x + (1 - \alpha) x_0) \leq \alpha f(x) + (1 - \alpha) f(x_0).$$

Taking the limit $\alpha \to 1$ proves $f(x) \geq t$. □

4 Linear Programming Duality

As an application of the Lagrange duality theorems, we prove the (strong) duality of linear programming (Theorem 5). For two vectors $u_1, u_2 \in \mathbb{R}^n$, we say $u_1 \leq u_2$ (resp. $u_1 \geq u_2$) if every coordinate of u_1 is at most (resp. at least) the corresponding coordinate of u_2.

Theorem 5. *For vectors $v \in \mathbb{R}^m, b \in \mathbb{R}^n$ and matrix $A \in \mathbb{R}^{n \times m}$, consider the following linear program:*

$$\underset{x \in \mathbb{R}^m}{\text{minimize}} \quad \langle x, v \rangle \tag{P3}$$

$$\text{s.t.} \quad Ax + b \leq \mathbf{0}. \tag{20}$$

Assume (P3) *is feasible. For some threshold $t \in \mathbb{R}$, assume that the objective value of* (P3) *cannot go below t. Then the following LP has a feasible solution $h \in \mathbb{R}^n$ achieving objective value at least t:*

$$\begin{aligned} \underset{h\in\mathbb{R}^n}{\text{maximize}} \quad & \langle b, h\rangle & \text{(D3)} \\ \text{s.t.} \quad & A^\mathsf{T} h + v = \mathbf{0}, & (21) \\ & h \geq \mathbf{0}. & (22) \end{aligned}$$

Proof. We start by writing (P3) in the form of (P1). Specifically, let us define $E = \mathbb{R}^m$, $S = \{s \in \mathbb{R}^n : s + b \leq \mathbf{0}\}$. Now (P3) is equivalent to

$$\begin{aligned} \underset{x\in E}{\text{minimize}} \quad & \langle x, v\rangle & \text{(P3')} \\ \text{s.t.} \quad & Ax \in S. & (23) \end{aligned}$$

Note that E is an open convex set and S is a convex polyhedron, so the regularity assumptions of Theorem 2 are satisfied. Therefore, there exists $h \in \mathbb{R}^n$ such that

$$\langle x, v\rangle + \left(\langle Ax, h\rangle - \sup_{s\in S}\langle s, h\rangle\right) \geq t \quad \text{for every } x \in \mathbb{R}^m. \tag{24}$$

The above implies that $\sup_{s\in S}\langle s, h\rangle < +\infty$. By the definition of S, for every $s \in S$, we can reduce each of its coordinates arbitrarily and the result still belongs to S. Therefore, $\sup_{s\in S}\langle s, h\rangle < +\infty$ implies $h \geq \mathbf{0}$. Now we have $\sup_{s\in S}\langle s, h\rangle = -\langle b, h\rangle$. Thus (24) becomes

$$\langle x, v\rangle + \langle Ax, h\rangle + \langle b, h\rangle \geq t \quad \text{for every } x \in \mathbb{R}^m,$$

or equivalently,

$$\langle x, A^\mathsf{T} h + v\rangle + \langle b, h\rangle \geq t \quad \text{for every } x \in \mathbb{R}^m.$$

Since $x \in \mathbb{R}^m$ can be arbitrary, the above inequality holds only when $A^\mathsf{T} h + v = \mathbf{0}$. This proves that h is a feasible solution to (D3) achieving objective value at least t. □

Remark 3. *We say* (D3) *is the* dual *of* (P3)*, and conversely* (P3) *is the* primal *of* (D3)*. For example, the dual of* (P0) *is*

$$\begin{aligned} \underset{h_1,h_2\in\mathbb{R}}{\text{maximize}} \quad & 2h_1 - 3h_2 & \text{(D0)} \\ \text{s.t.} \quad & h_1 - 2h_2 = -3, & (25) \\ & -3h_1 + 5h_2 = -7, & (26) \\ & h_1, h_2 \geq 0. & (27) \end{aligned}$$

The optimal solution to (D0) *is $h_1 = 29, h_2 = 16$. Note that this is in fact the unique feasible solution to* (D0)*, but in general, not every dual LP has a unique feasible solution.*

Remark 4. *In Theorem 5, we assume that* (P3) *is feasible. This assumption is necessary. Consider the following infeasible LP:*

$$\begin{aligned} \underset{x_1,x_2\in\mathbb{R}}{\text{minimize}} \quad & x_2 \\ \text{s.t.} \quad & x_1 \leq 0, \\ & -x_1 \leq -1, \\ & x_2 \leq 0. \end{aligned}$$

Since it is infeasible, for an arbitrary $t \in \mathbb{R}$, its objective value cannot go below t. However, the corresponding dual problem is also infeasible, so in particular, it cannot achieve objective value at least t:

$$\begin{aligned} \underset{h_1,h_2,h_3\in\mathbb{R}}{\text{maximize}} \quad & h_2 \\ \text{s.t.} \quad & h_1 - h_2 = 0, \\ & h_3 = -1, \\ & h_1, h_2, h_3 \geq 0. \end{aligned}$$

5 Semidefinite Programming Duality

As another application of Lagrange duality, we prove the (strong) duality of semidefinite programming under Slater's condition (Theorem 6).

Theorem 6. *Given matrices $V, A_1, \ldots, A_n \in \mathbb{R}^{m\times m}$ and a vector $b = (b_1, \ldots, b_n) \in \mathbb{R}^n$, consider the following semidefinite program*

$$\begin{aligned} \underset{X\in\mathbb{R}^{m\times m}}{\text{minimize}} \quad & \langle X, V\rangle && \text{(P4)} \\ \text{s.t.} \quad & \langle X, A_i\rangle + b_i \leq 0 \quad \text{for every } i = 1, \ldots, n, && (28) \\ & X \succeq 0. \end{aligned}$$

Assume the program satisfies Slater's condition: there exists $X \succ 0$ (not just $X \succeq 0$) such that constraint (28) *is satisfied. For some threshold $t \in \mathbb{R}$, assume that the objective value of* (P4) *cannot go below t. Then the following semi-definite program has a feasible solution $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ achieving objective value at least t:*

$$\begin{aligned} \underset{h=(h_1,\ldots,h_n)\in\mathbb{R}^n}{\text{maximize}} \quad & \langle b, h\rangle && \text{(D4)} \\ \text{s.t.} \quad & \sum_{i=1}^n h_i A_i + V \preceq 0, \\ & h_1, \ldots, h_n \geq 0. \end{aligned}$$

Proof. We start by writing (P4) in the form of (P2). Specifically, we define $E := \{X \in \mathbb{R}^{m\times m} : X \succ 0\}$ and $S := \{s \in \mathbb{R}^n : s + b \leq \mathbf{0}\}$. We have $\mathsf{cl}\, E = \{X \in \mathbb{R}^{m\times m} : X \succeq 0\}$ and $\mathsf{cl}\, S = S$. Now (P4) can be equivalently written as

$$\begin{aligned} \underset{X\in\mathsf{cl}\,E}{\text{minimize}} \quad & \langle X, V\rangle && \text{(P4')} \\ \text{s.t.} \quad & (\langle X, A_1\rangle, \ldots, \langle X, A_n\rangle) \in \mathsf{cl}\, S. \end{aligned}$$

It is easy to verify that E is an open convex set and S is a convex polyhedron. Therefore, the assumptions of Theorem 3 are satisfied. By Theorem 3, there exists $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ such that

$$\langle X, V\rangle + \left(\sum_{i=1}^n h_i \langle X, A_i\rangle - \sup_{s\in S}\langle s, h\rangle\right) \geq t \quad \text{for every } X \succeq 0. \tag{29}$$

The above implies that $\sup_{s\in S}\langle s,h\rangle < +\infty$. By our definition of $S = \{s \in \mathbb{R}^b : s + b \le \mathbf{0}\}$, we have $h \ge \mathbf{0}$ and $\sup_{s\in S}\langle s,h\rangle = -\langle b,h\rangle$. Thus (29) becomes

$$\langle X, V\rangle + \sum_{i=1}^{n} h_i \langle X, A_i\rangle + \langle b, h\rangle \ge t \quad \text{for every } X \succeq 0,$$

or equivalently,

$$\left\langle X, \sum_{i=1}^{n} h_i A_i + V \right\rangle + \langle b, h\rangle \ge t \quad \text{for every } X \succeq 0.$$

Since $X \succeq 0$ can be arbitrary, the above can hold only when $\sum_{i=1}^{n} h_i A_i + V \preceq 0$. Plugging $X = 0$ into the inequality above yields $\langle b,h\rangle \ge t$. Thus h is a feasible solution to (D4) achieving objective value at least t. □

Remark 5. *The Slater's condition in Theorem 3 cannot be replaced with the weaker assumption that* (P4) *is feasible. See a counter example in Anupam Gupta's notes:* `https://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15859-f11/www/notes/lecture12.pdf`.