Lecture 5: Duality Theory

Rajat Mittal

IIT Kanpur

The objective of this lecture note will be to learn duality theory of linear programming. We are planning to answer following questions.

– What are hyperplane separation theorems?
– What is Farkas lemma?
– How do we define the dual of a linear program?

For this lecture, it will be helpful to know the definitions of open, closed and compact sets. We will only require the geometric intuition behind these definitions.

To begin, we will state another useful property of convex sets: there exists a closest point in a convex set for any point outside the convex set. The concept of orthogonal projection is known to us in the context of linear subspaces. Given a point x outside a linear subspace L, the projection of x on L is the closest point inside the subspace L to x. This concept can be extended to have the closest point to x with respect to a convex set C.

Lemma 1. *Suppose x is a point outside a non-empty closed convex set C. Then there exist a unique point $C(x)$, on the boundary of C, which is closer to x than any other point in C. Say $d(x, y)$ denote the Euclidean distance between x and y.*

$$d(x, C(x)) < d(x, y) \quad \forall\ y \neq x \in C$$

Note 1. We will omit the proof of this lemma, but the uniqueness part can be shown easily.

For the relevance of this theorem, note that this is not true for open concave sets. For example, take all points in $\mathbb{R}^n$ not in a circle. The uniqueness is lost if x is the center of the circle.

Exercise 1. Come up with a closed concave set which does not satisfy the theorem above.

Exercise 2. Come up with an example to prove that C is required to be closed in the lemma.

1 Separating hyperplanes

Given a point x and a set of points C, how can we possibly prove that x is not part of C?

Suppose there exist a vector a and a scalar b, s.t., $a^T x > b$. If it can be shown that $a^T y \leq b$ for every $y \in C$, this gives a mathematical proof that x is not part of C. We can say that a, b (which define a hyperplane) separate x and C. But can we find such a, b all the time?

Exercise 3. Show that there exist C and $x \notin C$, s.t., no a, b separate them.

It turns out that this is always possible if C is a closed convex set. Even for two closed convex sets C_1 and C_2, non-intersecting, similar separation can be found.

Exercise 4. Give an example of two sets which cannot be separated.

Lets define this notion of separation more formally. A hyperplane $a^T x - b = 0$ *separates* sets C and D, if

$$a^T x - b \geq 0,\ \forall x \in C \text{ and } a^T x - b \leq 0,\ \forall x \in D.$$

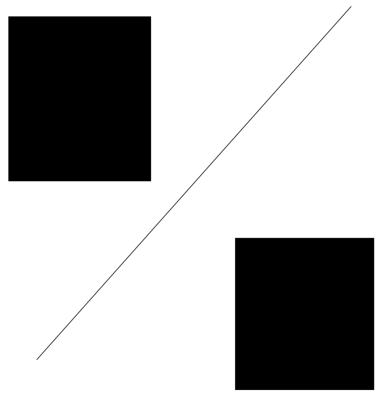

Fig. 1. Convex sets and hyperplanes separating them

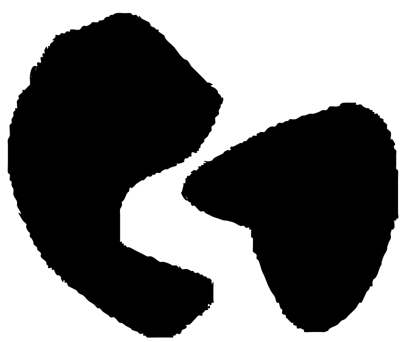

Fig. 2. Non-convex sets

We will be interested in the case when both the convex sets are not contained in the hyperplane (*proper separation*). A stronger separation can be defined by taking strict inequalities. The separation is called a *strict separation* if strict inequalities hold on both sides.

$$a^T x - b > 0, \ \forall x \in C \text{ and } a^T x - b < 0, \ \forall x \in D.$$

Similar to separation there is a concept of supporting hyperplane. Given a convex set C and a point x on the boundary, a hyperplane with normal a is supporting at x, iff,

- The hyperplane passes through x.
- The set lies on one side of the hyperplane, $\forall \ y \in C : \quad a^T x \geq a^T y$.

There is a general theorem that two disjoint convex sets can be *separated* by a hyperplane. Depending on the convex sets the kind of separation can be different.

Similarly, given a point on the boundary of a closed convex set, we can always draw a supporting hyperplane through it.

Geometrically, suppose we are given a convex set and a point on the boundary of the convex set. It is always possible to draw a tangential hyperplane through the point so that the convex set lies only on one side of the hyperplane. Below we look at various cases.

1.1 Two convex sets

Theorem 1. *Given two convex sets C and D, which are mutually disjoint, there always exists a hyperplane $a^T x - b = 0$ separating them. Hence,*

$$a^T x - b \geq 0, \ \forall x \in C \text{ and } a^T x - b \leq 0, \ \forall x \in D.$$

Note 2. The separation here is not strict.

Exercise 5. Construct two convex sets where separation cannot be strict.

We will show the proof for the case when at least one of the sets is compact and the other one is at least closed. We will mostly be dealing with such situations in future.

Proof. In this case actually strict separation can be achieved. Since one set is compact and other is closed, there exist two points $c \in C$ and $d \in D$ whose distance is minimum among any pair of points, one from C (say this is compact) and one from D. This distance can't be zero since C and D are mutually disjoint.

This statement seems quite intuitive but requires a proof. For people with a background of analysis (others can assume this statement), this follows from the fact that $d(x, D)$ (the minimum distance of point x from set D, which exists by Lemma 1) is a continuous function on a compact set C and hence its minimum should exist. For others, this statement can be assumed.

Let us look at the perpendicular bisector of the line between c and d. This will be the separating hyperplane.

We will show that if $\exists \ x \in C$, s.t., $(c-d)^T x < (c-d)^T c$ then c and d are not the closest points.

Exercise 6. Why is it enough to show that for all points x in C, $(c-d)^T x \geq (c-d)^T c$. Hint: $c - d$ is normal to the perpendicular bisector of the line between c and d.

So, suppose there exists a point $x \in C$, s.t.,

$$(c-d)^T x < (c-d)^T c \tag{1}$$

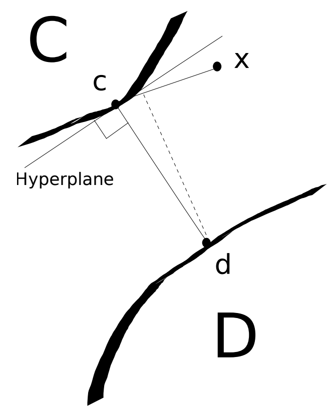

Fig. 3. Closest points

Since C is convex, for every $\theta \in [0,1]$, point $\theta x + (1-\theta)c$ lies in C. What is the distance between this point and d? Consider,

$$
\begin{aligned}
& \|d - (\theta x + (1-\theta)c)\|^2 \\
= & \|(d-c) - \theta(x-c)\|^2 \\
= & \|d-c\|^2 + \theta^2\|x-c\|^2 - 2\theta(d-c)^T(x-c) \\
= & \|d-c\|^2 + \theta^2\|x-c\|^2 - 2\theta(c-d)^T(c-x)
\end{aligned}
$$

The last term $2\theta(c-d)^T(c-x)$ is positive because of our assumption, Eq. 1. Hence, by choosing θ small enough, we can make sure that the distance between point $\theta x + (1-\theta)c$ and d is less than the distance between c and d. This gives us the contradiction.

Exercise 7. What is the exact value of θ at which this will happen?

□

1.2 A point and a convex set

Our next example will be a point and a convex set. In this case we get a strict separation by the hyperplane, s.t., point lies on one side of the hyperplane and the set on the other side. Here strict means both the point and the set are disjoint with the hyperplane.

Theorem 2. *Given a closed convex set C and a point p. There always exist a hyperplane $H = a^T x - b$ which strictly separates them. So $a^T p - b > 0$ and $a^T x - b < 0,\ \forall x \in C$.*

The proof follows by constructing a small enough ball around point p which is compact and disjoint with C. The result follows from Thm. 1 then.

Corollary 1. *Any closed convex set is the intersection of all the half spaces which contain it.*

Proof is left as an exercise. The set of half spaces can be restricted to all the supporting hyperplanes.

1.3 Farkas lemma: A point and a cone

A special case of the previous section is the separation between a point and a finitely generated cone. Remember that a finitely generated cone is convex. The proof follows from the previous theorems, but it will be instructive to see another proof of this.

Lemma 2 (Farkas). *Given a set of vectors $x_1, \cdots, x_k \in \mathbb{R}^n$ (equivalently $C \in \mathbb{R}^{n \times k}$) and a point $b \in \mathbb{R}^n$. Exactly one of the following two conditions are satisfied.*

1. *$\exists\ \alpha \in \mathbb{R}^k_+$ (α in positive orthant), such that, $b = C\alpha$.*
2. *$\exists\ a \in \mathbb{R}^n$, such that, $a^T b > 0$ and $a^T C \leq 0$ (entrywise).*

Note 3. The inequality is strict (strict separation) and the hyperplane is of special kind.

Proof. It is clear that both conditions cannot be satisfied simultaneously. Because then,

$$\begin{gathered} 0 \geq a^T C \\ \Rightarrow 0 \geq a^T C\alpha \\ \Rightarrow 0 > a^T b. \end{gathered}$$

which is a contradiction.

Suppose b can't be written as the conic combination of columns of C, i.e., first condition does not hold. Then look at cone generated by columns of C. It is finitely generated cone and hence by Weyl's theorem, it can be expressed as a bunch of inequalities,

$$Cone(x_1, \cdots, x_k) = \{x : Ax \leq 0\}.$$

Because b is not a member of this cone, there exists a row of A whose inner product with b is strictly positive (call it a_i). Then $a_i^T b > 0$ and $\forall i,\ a_i^T x_i \leq 0 \Rightarrow a_i^T C \leq 0$. ☐

Exercise 8. Interpret Farkas lemma as a hyperplane separation theorem. What do you know about this hyperplane?

This theorem converts the membership question in a cone to finding a separating hyperplane question. The question of membership in the cone is of real importance in convex optimization. It can be shown that optimization over a cone can be done using polynomially many calls to membership/separation algorithm for the cone.

2 Dual of a linear program

One of the strongest aspect of linear optimization is the presence of duality theory. The dual of a program is another linear optimization program which provides tight upper/lower bounds on the original linear program.

We will motivate the duality theory by showing how to take the dual of a linear program. Suppose there is a linear program,

$$\begin{aligned} \max \quad & 2x_1 + 3x_2 + x_3 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 5 \\ & x_1 + 2x_2 = 10 \\ & x_1, x_2, x_3 \geq 0. \end{aligned}$$

What is the value of this program? On a close inspection, it will be clear that the objective function is a linear combination of constraints. Hence whatever be the feasible solution, the objective value will be 15.

Notice that it is required to have a feasible solution. In case there is no feasible solution then the objective value of the program will be $-\infty$. Remember that if a max optimization program does not have a feasible solution then its optimal value is $-\infty$ and if the min optimization program is infeasible then its optimal value is $+\infty$.

Let's take another example,

$$\begin{aligned} \max \quad & 2x_1 + 3x_2 + x_3 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 5 \\ & x_1 + 2x_2 + x_3 = 10 \\ & x_1, x_2, x_3 \geq 0. \end{aligned}$$

In this case it is not clear if 15 is the optimal value. But since $x_3 \geq 0$, the objective value is $\leq 15 - x_3$ and hence lesser than or equal to 15. So 15 is an upper bound on the objective value. Lets make this argument more precise. Suppose,

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0, \end{aligned} \tag{2}$$

is a linear program. If there exist a linear combination of constraints, say y, s.t., $A^T y = c$. Then $b^T y$ will be the optimal value of the program (given that there exists a feasible solution). But getting $A^T y = c$ every time is not easy. Instead, if $A^T y \geq c$, then also $b^T y$ is an upper bound on the value of the program.

Since we want to consider the best upper bound.

$$\begin{aligned} \min \quad & b^T y \\ \text{s.t.} \quad & A^T y \geq c \end{aligned} \tag{3}$$

From the way this program is constructed, the optimal value of Eqn. 3 is always higher than optimal value of Eqn. 2.

Eqn. 2 is known as the *primal linear program.* Eqn. 3 is known as the *dual linear program.*

Exercise 9. What is the dual if the constraint was $Ax \leq b$ instead of $Ax = b$ in the primal linear program?

The argument above showed that the dual objective value is higher than the primal objective value. This is known as the *weak duality.* In case of linear programs it is known that actually dual value is always same as the primal value if primal has a feasible solution. This is known as *strong duality.*

The terms primal and dual are interchangeable. The dual of the dual is the primal program. So if Eqn. 3 is considered as the primal program then the dual will be Eqn. 2.

One important thing to notice is that for every primal constraint there is a dual variable and for every primal variable there is a dual constraint. This relationship is much more stronger than it seems now.

Exercise 10. Find the dual of the following linear program.

$$\begin{aligned} & \max c^T x \\ \text{s.t.} \quad & Ax \geq b \\ & x \leq 0, \end{aligned}$$

Exercise 11. What is the relationship between dual and separating hyperplanes?

For easier access to taking dual, this table might be helpful.

	Primal variable to dual constraint	Primal constraint to dual variable
Max	$\geq \to \geq$, $\leq \to \leq$, unrestricted $\to =$	$\geq \to \leq$, $\leq \to \geq$, $= \to$ unrestricted
Min	$\geq \to \leq$, $\leq \to \geq$, unrestricted $\to =$	$\geq \to \geq$, $\leq \to \leq$, $= \to$ unrestricted

Exercise 12. Verify that the table is correct.

3 Complementary slackness

In the last section we saw how to formulate a dual program from a linear program. From the formulation, it turned out that the dual program gave an upper/lower bound depending on whether the problem was maximization/minimization respectively.

The strength of duality theory lies in the fact that these bounds are in fact tight. This implies that the dual value always agrees with the primal value. This fact has been useful in numerous applications in computer science. Let us study this relationship in detail.

Note 4. The fact that these bounds are tight only holds true for linear programming. For other convex programming instances, we need to be careful.

Let's look at the primal-dual pair for a linear program.

Primal

$$\begin{aligned} &\max c^T x \\ \text{s.t.} \quad &Ax \leq b \\ &x \geq 0 \end{aligned}$$

Dual

$$\begin{aligned} &\min y^T b \\ \text{s.t.} \quad &A^T y - c \geq 0 \\ &y \geq 0 \end{aligned}$$

First, it is instructive to see the direct proof of weak duality. Given any feasible solution X for primal and y for dual,

$$c^T x \leq y^T A x \leq \sum_i y_i b_i = y^T b. \tag{4}$$

Note 5. Above equation implies, all feasible solutions of dual give an upper bound on the optimal value of primal. Similarly any feasible solution for primal gives a lower bound on the optimal value of dual.

Suppose the optimal value of primal is p^*, attained at x^*. Similarly the optimal value of dual is d^* and obtained for y^*. Weak duality implies that $p^* \leq d^*$. Assume that this two values are equal, i.e., $p^* = d^*$ (strong duality, Thm. 3). Then, from Eqn. 4,

$$(A^T y^* - c)^T x^* = 0 \text{ and } (Ax^* - b)^T y^* = 0. \tag{5}$$

The conditions above are called the *complementary slackness* condition. So, for optimal x^*, y^* with strong duality, the complementary slackness condition holds. Conversely, if x, y are feasible solutions of primal and dual respectively and satisfy the complementary slackness condition then strong duality holds and $p^* = d^*$.

What does the complementary slackness conditions tell us?

- Given an optimal solution x^* for the primal, if $x_i^* \neq 0$ then $a_i^T y - c_i = 0$, where a_i is the i^{th} column of A. In other words, if the primal optimal variable is non-zero, then the corresponding constraint in dual is tight.
- Given an optimal solution y^* for the dual, if $y_j^* \neq 0$ then $a_j^T x - b_j = 0$, where a_j is the j^{th} column of A. In other words, if the dual optimal variable is non-zero, then the corresponding constraint in primal is tight.

If we write $a_i^T y - c_i = w_i$ (a_i's are column) and $a_j^T x - b_j = z_j$ (a_j's are rows) as slack variables. Then for primal and dual both we have $m + n$ variables. Complementary slackness tells us that out of these $m + n$ variables, if a primal variable is non-zero then the corresponding dual variable is 0.

4 Extra reading: strong duality

Strong duality states that the optimal value of primal is same as optimal value of dual. The next theorem states that strong duality holds for all linear programs. We look at the following format of primal and dual,

$$\begin{aligned} &\max c^T x \\ \text{s.t.}\quad &Ax - b = 0 \\ &x \geq 0 \end{aligned} \qquad\qquad \begin{aligned} &\min b^T y \\ \text{s.t.}\quad &A^T y - c \geq 0 \end{aligned}$$

Theorem 3. *Given a linear program in above format with parameters c, A, b (say a_i is the i^{th} row of A), suppose the feasible set of primal is $\mathcal{P}$ and feasible set of dual is $\mathcal{D}$. Then one of the following is true,*

- *If $\mathcal{P}$ is empty (primal infeasible), then dual is unbounded or infeasible.*
- *If $\mathcal{D}$ is empty (dual infeasible), then primal is unbounded or infeasible.*
- *If both $\mathcal{D}, \mathcal{P}$ are non-empty, then strong duality holds. So for optimal x^*, y^*, $p^* = c^T x^* = b^T y^* = d^*$.*

In other words, if any one of primal or dual is feasible and bounded then strong duality holds. The proof is based on Farkas Lemma.

Proof. Suppose $\mathcal{P}$ is empty. By Farkas lemma, there exist y, s.t., $b^T y < 0, A^T y \geq 0$. If dual is infeasible, then there is nothing to prove, else there exist $y' : \ A^T y' \geq c$.

Exercise 13. Show that using y, y', $z_i's$ can be constructed, s.t., $A^T z_i \geq c, b^T z_i \to -\infty$ as $i \to \infty$.

From the previous exercise, dual is unbounded. Similarly the case of $\mathcal{D}$ being empty can be dealt.

Exercise 14. Show that if primal is unbounded, then dual is infeasible.

So assume now that both $\mathcal{P}, \mathcal{D}$ are non-empty and the optimal value is bounded. Say p^* (d^*) are the optimal values achieved at x^* (y^* respectively).

Since p^* is optimal, there does not exist x, s.t., $(Ax, c^T x) = (b, p^* + \epsilon)$ (for any $\epsilon > 0$). Considering $A' = (A, c)$ and $b' = (b, p^* + \epsilon)$, the Farkas lemma implies,

$$\exists y : \quad A'^T y \leq 0 \text{ and } b'^T y > 0$$

We can scale y to ensure that the last co-ordinate of y (y_{m+1}) is either +1,−1 or 0. Let y' denotes y except the last co-ordinate (which can be $1, -1$ or 0).

Suppose $y_{m+1} = -1$. Then,

$$\begin{aligned} b^T y' > p^* + \epsilon \text{ and } A^T y' \leq c \\ \Rightarrow b^T y' > p^* \text{ and } (x^*)^T A^T y' \leq c^T x^* = p^* \\ \Rightarrow b^T y' > p^* \text{ and } (b^T) y' = p^* \end{aligned}$$

But this is a contradiction since $Ax^* = b$.

Exercise 15. Show that y_{m+1} is not equal to zero.

So we can assume $y_{m+1} = 1$. Putting back the definition of A', b',

$$\exists y' : \quad A^T y' + c \leq 0 \text{ and } b^T y' + p^* + \epsilon > 0$$

Replacing y' by $-y'$,

$$\exists y' : \quad A^T y' \geq c \text{ and } b^T y' < p^* + \epsilon$$

So, we can have a feasible dual solution with objective value less than $p^* + \epsilon$. That means, the dual optimal value is less than $p^* + \epsilon$ for every $\epsilon > 0$. Hence, the dual optimal value is p^* (remember that it can't be less than p^*).

□

Note 6. There can be cases when both primal and dual are infeasible.

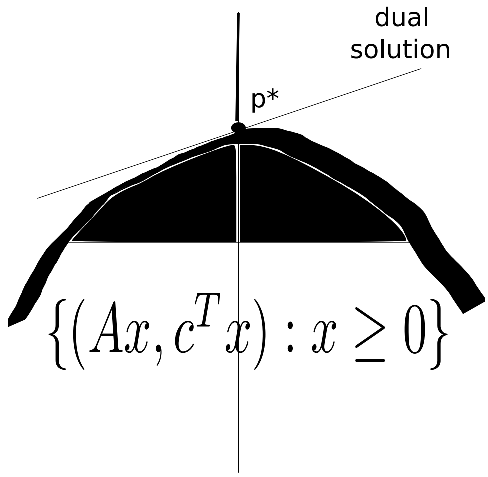

Fig. 4. Strong duality through picture

5 Assignment

Exercise 16. Show that the closest point to a convex set is unique (from Lemma 1).

Exercise 17. Modify the proof of Thm. 1 to show the Thm. 2 directly.

Exercise 18. Prove Cor. 1.

Exercise 19. Show that the dual of the dual is the primal program (for the standard linear program).

Hint: First convert the dual into primal form and then take the dual.