Optimization Methods	Date:	*9th February, 2018*
Instructor: *Sujit Prakash Gujar*	Scribes:	Rohan Karnawat Rohith Veggalam Prudhvi Raju

Lecture 9 : Linear Programming - Duality

1 Recap

1.1 Simplex Algorithm Example

Solving the problem using Largest Co-efficient Algorithm.

$$\begin{gathered} \max\ 9x_1 + 3x_2 + x_3 \\ \text{s.t.} \quad x_1 \leq 1 \\ 6x_1 + x_2 \leq 9 \\ 18x_1 + 6x_2 + x_3 \leq 81 \\ x_1, x_2, x_3 \geq 0 \end{gathered}$$

Let x_4, x_5, x_6 be slack variables , above LP in standard form can be converted to equational form as below

$$\begin{gathered} x_1 + x_4 = 1 \\ 6x_1 + x_2 + x_5 = 9 \\ 18x_1 + 6x_2 + x_3 + x_6 = 81 \end{gathered}$$

Let's assume basic feasible solution be X =(0,0,0,1,9,81) thus $X_B = (x_4, x_5, x_6)$ and $X_N = (x_1, x_2, x_3)$
The simplex tableau looks like this:

$$\begin{array}{l} x_4 = 1 - x_1 \\ x_5 = 9 - 6x_1 - x_2 \\ x_6 = 81 - 18x_1 - 6x_2 - x_3 \\ \hline z = 9x_1 + 3x_2 + x_3 \end{array}$$

Applying largest coefficient algorithm we see that x_1 becomes entering variable and attains a maximum value of 1 making x_4 leaving variable. Thus new basic feasible solution is X=(1,0,0,0,3,63) , z=9 , $X_B = (x_1, x_5, x_6)$ and $X_N = (x_2, x_3, x_4)$
The simplex tableau looks like this:

$$\begin{array}{l} x_1 = 1 - x_4 \\ x_5 = 3 + 6x_4 - x_2 \\ x_6 = 63 + 18x_4 - 6x_2 - x_3 \\ \hline z = 9 - 9x_4 + 3x_2 + x_3 \end{array}$$

Applying largest coefficient algorithm we see that x_2 becomes entering variable and attains a maximum value of 3 making x_5 leaving variable. Thus new basic feasible solution is X=(1,3,0,0,0,45) , z=18 , $X_B = (x_1, x_2, x_6)$ and $X_N = (x_3, x_4, x_5)$
The simplex tableau looks like this:

$$\begin{array}{l} x_1 = 1 - x_4 \\ x_2 = 3 + 6x_4 - x_5 \\ x_6 = 45 - 18x_4 - 6x_5 - x_3 \\ \hline z = 18 + 9x_4 - 3x_5 + x_3 \end{array}$$

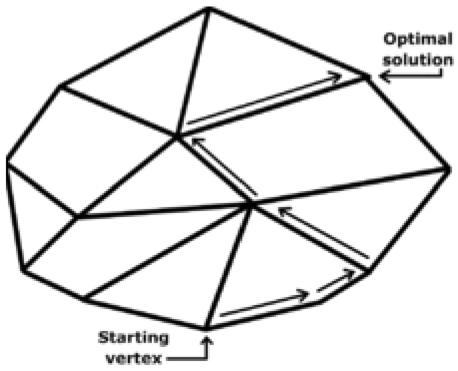

Figure 1: Klee-Minty Polytope: Dantzig's simplex algorithm was not a polynomial-time algorithm when applied to this cube.
Source: Wikipedia

If we continue processing the above problem using largest coefficient algorithm, we will be travel all the vertices of the feasible region before reaching to the step where no more coefficients of X_N in z are positive.

(0,9,0,1,0,27) –> (0,9,27,1,0,0) –> (1,3,45,0,0,0) –> (1,0,63,0,3,0) –> *(0,0,81,1,9,0)*

The simplex algorithm traverses the edges between vertices until it reaches an optimal vertex. However, the simplex algorithm visits every vertex in the worst case (${}^{m}\mathrm{C}_{n}$) of a problem whose feasible region is the Klee Minty cube as above, so the number of steps rises exponentially with the dimension of the problem. However, this does not mean that the pivot rule (largest coefficient) that we used above is the worst rule. Using a different pivot rule may be better for this optimization example but it does not mean that it would always be ideal. *Then why is simplex method widely used?* Because in practice the feasible region is simpler, and the worst case complexity is not always reached. The expected complexity is m^2, where m is the number of constraints.

1.2 Duality

Linear programming problems are optimization problems in which the objective function and the constraints are all linear. In the primal problem, the objective function is a linear combination of n variables. There are m constraints, each of which places an upper bound on a linear combination of the n variables. The goal is to maximize the value of the objective function subject to the constraints. A solution is a vector of n values that achieves the maximum value for the objective function.

In the dual problem, the objective function is a linear combination of the m values that are the limits in the m constraints from the primal problem. There are n dual constraints, each of which places a lower bound on a linear combination of m dual variables.

Advantages of Duality:

1. Reduces the number of variables and hence, the computation complexity. For example complexity is nm^2 in primal, where n is the number of decision variables and m is the number of constraints. Now if $n << m$, and the complexity of the dual is mn^2 which is computationally simpler than the primal.

2. Provides a good bound for the decision variables. Also tells if the feasible region is unbounded.

3. Sometimes finding an initial feasible solution to the dual is much easier than finding one for the primal.

2 Primals and Duals

2.1 Definitions

1. Weak duality theorem
   If x is a feasible solution to Primal(P) and y is a feasible solution to Dual(D), then:
   (a) $c^T x \leq b^T y$
   (b) if equality holds in the above inequality, then x is an optimal solution to (P) and y is an optimal solution to (D).

2. Strong duality theorem or Duality theorem of linear programming
   For a linear program P,
   $$\text{maximize } c^T x \text{ subject to } Ax \leq b,\ x \geq 0$$
   and its Dual D,
   $$\text{minimize } b^T y \text{ subject to } A^T y \geq c,\ y \geq 0$$
   One of the following holds true
   (a) Neither P nor D has a feasible solution
   (b) P has no feasible solution and D is unbounded
   (c) P is unbounded and D has no feasible solution
   (d) If both P and D has an feasible solution and there exists an optimal solution x^* for Primal(P), then there exists an optimal solution y^* for Dual(D) and the value of $c^T x^*$ in (P) equals the value of $b^T y^*$ in (D).
   $$c^T x^* = b^T y^*$$

2.2 Properties and Behavior

Consider the following optimization problem:

$$\begin{aligned} &\max \text{ x}_1 \\ \text{s.t.} \quad & x_1 - x_2 \leq 1 \\ & -x_1 + x_2 \leq 2 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Converting it to dual:

$$\begin{aligned} y_1(x_1 - x_2) &\leq y_1 \\ y_2(-x_1 + x_2) &\leq 2y_2 \\ \implies (y_1 - y_2)x_1 + (y_2 - y_1)x_2 &\leq y_1 + 2y_2 \end{aligned}$$

$\therefore$ The dual optimization problem:

$$\begin{aligned} &\min\ y_1 + 2y_2 \\ \text{s.t.} \quad & y_1 - y_2 \geq 1 \end{aligned}$$

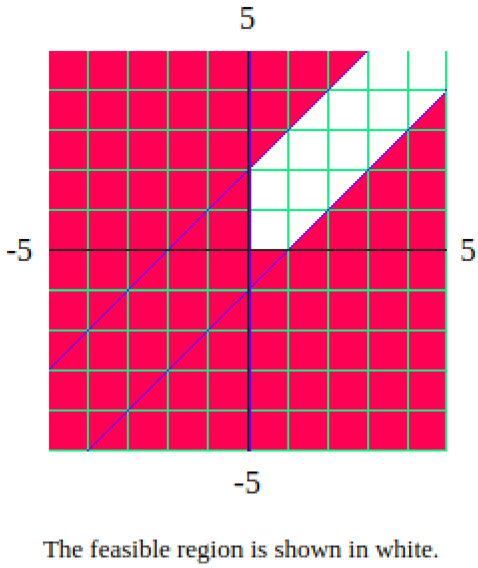

Figure 2: Unbounded feasible region

$$y_2 - y_1 \geq 0$$

$$\implies y_2 \geq y_1 \text{ and } y_2 + 1 \leq y_1, \text{ which is infeasible}$$

Since dual helps find the upper bound of a maximization problem, an infeasible solution to the dual formulation means that the primal problem has no upper bound (unbounded feasible region). This is shown in Figure 2.

Consider the following cases:

1. If the first constraint of the primal was $x_1 + x_2 = 1$, then there is no need to enforce $y_1 \geq 0$ in the dual because equality is assertive and unaffected by sign.
2. If the first constraint of the primal was $x_1 + x_2 \geq 1$, then in the dual space, $y_1 \leq 0$ will be enforced. This will be needed to ensure uniformity in representation in standard form.
3. If a variable in primal is constrained in the form $x_i \leq k$ and $x_i \geq k$, then instead of having two variables in dual for the two primal constraints, one can have an equality involving just one variable y_i.

Both primal and dual cannot be unbounded.
According to the strong duality theorem if one of them(Primal/Dual) is unbounded then other(Dual/Primal) will not have a feasible solution.

$$\begin{array}{rc} \text{Ex:} & \max x_1 \\ \text{s.t.} & x_1 - x_2 \leq 1 \\ & -x_1 + x_2 \leq 2 \\ & x_1, x_2 \geq 0 \end{array}$$

Both primal and dual can be infeasible
According to the Duality Theorem of linear programming both(Primal/Dual) can be infeasible.

$$\begin{array}{rc} \text{Ex:} & \max x_2 \\ \text{s.t.} & -x_1 + x_2 \leq -1 \\ & x_1 - x_2 \leq 0 \\ & x_1, x_2 \geq 0 \end{array}$$

2.3 Dualization Recipe

Let's assume primal LP be

$$\text{P} = \max(c^T x \mid Ax \leq b,\ x \geq 0, x \in \mathbb{R}^n)$$

then corresponding dual LP will be,

$$\text{D} = \min(b^T y \mid A^T y \geq c,\ y \geq 0,\ y \in \mathbb{R}^m)$$

where A is an $m \times n$ matrix.

When there are equality constraints or variables that may be negative, the primal LP

$$\begin{aligned} \text{P} = \max(c^T x) \\ \text{s.t.} \quad a_i x \leq b_i \quad \forall i \in I_1 \\ a_i x = b_i \quad \forall i \in I_2 \\ x_j \geq 0 \quad \forall j \in J_1 \\ x_j \in R \quad \forall j \in J_2 \end{aligned}$$

corresponds to the dual LP

$$\begin{aligned} \text{D} = \min(b^T y) \\ \text{s.t.} \quad y_i \geq 0 \quad \forall i \in I_1 \\ y_i \in R \quad \forall i \in I_2 \\ A_j y \geq c_j \quad \forall j \in J_1 \\ A_j y = c_j \quad \forall j \in J_2 \end{aligned}$$

Table 1: Dualization Recipe

	Primal	*Dual*
Variable	$x_1, x_2, ..x_n$	$y_1, y_2, ...y_m$
Constraints Matrix	A	A^T
Objective Function	max $c^T X$	min $b^T Y$
ith Constraint	$\leq$ $\geq$ $=$	$y_i \geq 0$ $y_i \leq 0$ $y \in \mathbb{R}$
jth Variable	$x_j \geq 0$ $x_j \leq 0$ $x_j \in \mathbb{R}$	$\geq$ $\leq$ $=$

2.4 Proof of strong duality from the Simplex method

Proof. Consider a linear program in standard form

$$\text{P: max } c^T x \text{ s.t. } Ax \leq b,\ x \geq 0$$

By adding m slack variables, we obtain its equational form

$$\text{P': max } \bar{c}^T \bar{x} \text{ s.t. } \bar{A}\bar{x} = b,\ \bar{x} \geq 0$$

where,

$$\bar{x} = (x_1, ..., x_n, x_{n+1}, ..., x_{n+m})$$

$$\bar{A} = (A|I_m)$$

$$\bar{c} = (c_1, ..., c_n, 0, ..., 0)$$

and its dual is

$$\text{D: min } b^T y \text{ s.t. } A^T y \geq c,\ y \geq 0$$

Let x^* be the optimal solution for P using simplex method with a basis B. The first n components of the vector $\bar{x}$ contains the optimal solution(x^*) of P and remaining are slack variables. From the simplex method, in the final simplex tableau the z vector is given by $z = z_0 + r^T x_N$, where $r \leq 0$. From the simplex method $\bar{x}^* = (\bar{x}^*_B|\bar{x}^*_N)$, where $\bar{x}^*_B = \bar{A}_B^{-1} b$ and $\bar{x}^*_N = \bar{0}$

$$c^T x^* = \bar{c}^T \bar{x}^* = \bar{c}_B^T \bar{x}^*_B + \bar{c}_N^T \bar{x}^*_N = \bar{c}_B^T \bar{x}^*_B = \bar{c}_B^T \bar{A}_B^{-1} b$$
$$\text{let } y^{*T} = \bar{c}_B^T \bar{A}_B^{-1}$$
$$\implies c^T x^* = y^{*T} b = b^T y^*$$

Now, If we can prove that y^* is a feasible solution for the Dual(D), from the above equation and the weak duality theorem, we can say that y^* is the optimal solution for the dual. That is $A^T y^* \geq c$ and $y^* \geq \bar{0}$. $y^* \geq \bar{0}$ can also be written as $I_m y^* \geq \bar{0}$ and hence both of the feasibility conditions can be written as $\bar{A}^T y^* \geq \bar{c}$ because $\bar{c}_N$ is $\bar{0}$ and $I_m y^* \geq \bar{0}$. Here $\bar{c}_N$ and $\bar{c}_B$ are non basic and basic components of $\bar{c}$ respectively.
Proof for feasibility:

$$\bar{A}^T y^* = \bar{A}^T (\bar{c}_B^T \bar{A}_B^{-1})^T$$
$$\implies \bar{A}^T y^* = (\bar{c}_B^T \bar{A}_B^{-1} \bar{A})^T = [\bar{c}_B^T \bar{A}_B^{-1} \bar{A}_B | \bar{c}_B^T \bar{A}_B^{-1} \bar{A}_N]^T = [\bar{c}_B^T | \bar{c}_B^T \bar{A}_B^{-1} \bar{A}_N]^T$$

A has two components A_B and A_N. The vector $(\bar{c}_B^T \bar{A}_B^{-1} \bar{A})^T$ can be written using two components. First component $\bar{c}_B^T \bar{A}_B^{-1} \bar{A}_B$ and the second component is $\bar{c}_B^T \bar{A}_B^{-1} \bar{A}_N$
While solving the linear program using simplex method, in final simplex tableau, from the $z = z_0 + r^T x_N$ row, we know that

$$r \leq \bar{0}$$

$$\implies \bar{c}_N - (\bar{c}_B^T \bar{A}_B^{-1} \bar{A}_N)^T \leq \bar{0}$$
$$\implies (\bar{c}_B{}^T \bar{A}_B^{-1} \bar{A}_N)^T \geq \bar{c}_N$$
$$\implies \left[\frac{\bar{c}_B^T}{(\bar{c}_B^T \bar{A}_B^{-1} \bar{A}_N)^T}\right]^T \geq [\bar{c}_B | \bar{c}_N]$$
$$\implies \bar{A}^T y^* \geq \bar{c}$$

Hence we have proved that y^* is feasible solution to the Dual. The weak duality always holds, that is $c^T x^* \leq b^T y^*$. Since the objective of the dual(D) is to minimize $b^T y^*$ and $b^T y^*$ is lower bounded by $\bar{c}^T \bar{x}^*$, y^* is an optimal solution to the Dual.

□

This is a more algorithmic proof. Geometrical or structural proofs are also available; for example Farkas' Lemma.

2.5 An optimization to solve a linear program

Consider a general linear program in equational form with n decision variables and m slack variables

$$\text{P: max } \bar{c}^T \bar{x} \text{ s.t. } Ax \leq b,\ x \geq 0$$

and its dual is

$$\text{D: min } b^T y \text{ s.t. } A^T y \geq c,\ y \geq 0$$

Let us create a new linear program, say N, from the above linear programs with the following objective function and constraints. This linear problem is only polynomially more bigger than the primal and dual problems above.

$$\begin{aligned} \text{N:} \qquad & \max c^T x \\ \text{subject to: } & Ax \leq b \\ & A^T y \geq c \\ & c^T x \geq b^T y \\ & x, y \geq \bar{0} \end{aligned}$$

Now let $k = (x_0|y_0)$ is a feasible solution to the above linear program. So the constraints $c^T x_0 \geq b^T y_0$ are satisfied. With the help of weak duality theorem for any linear program ($c^T x \leq b^T y$) we can say that $b^T y_0 = c^T x_0$. By strong duality theorem, we can say that x_0 is an optimal solution to the Primal and y_0 is an optimal solution for the Dual.
The advantage of creating such a linear program N, using a linear program P and its dual D is that if we can find a feasible solution to N, we can find the optimal solution to P. The worst case complexity to find an optimal solution for a linear program using simplex method is exponential. But a feasible solution can be found in polynomial time. If we find a feasible solution to N in polynomial time means that we found an optimal solution for P in polynomial time. Note that the linear program P and its Dual D must have a feasible solution and should be unbounded for the solution x_0 to be optimal according to strong duality.

3 An alternative approach - Ellipsoid Method

The ellipsoid method, cannot compete with the simplex method in practice, but it had immense theoretical significance. It is the first linear programming algorithm for which it was proved that it always runs in polynomial time (which is not known about the simplex method up to the present, and for many pivot rules it is not even true). The ellipsoid method generates a sequence of ellipsoids whose volume uniformly decreases at every step, thus enclosing a minimizer of a convex function.
For understanding the working of this algorithm some prerequisite understanding is required:

1. Hypersphere:

   (a) Equation for circle (2D Hypersphere) centered at (x_0, y_0):

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \tag{1}$$

(b) Similarly, the equation of a sphere will have a term $(z - z_0)^2$. Equation of a hypersphere of n-dimension is given by:

$$\sum_{i=1}^{n}(x_i - x_{i0})^2 = r^2 \tag{2}$$

(c) Matrix representation- X is a Column Vector:

$$(X - X_0)^T(X - X_0) \tag{3}$$

(d) Region inside hypersphere: $H = \{x \in R^n | x^T x \leq r^2\}$

2. Ellipse:

(a) Equation for 2D ellipse centered at origin:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{4}$$

(b) Equation for an n-dimensional hyper-ellipse centered at origin:

$$\sum_{i=1}^{n}\frac{x_i^2}{a_i^2} = 1 \tag{5}$$

(c) Matrix Representation of hyper-ellipse:

$$X^T M^{-1} X = 1 \tag{6}$$

here M is a diagonal matrix:

$$M = \begin{bmatrix} a_1^2 & & 0 \\ & \ddots & \\ 0 & & a_n^2 \end{bmatrix}$$

(d) Region inside hyper-ellipse centered at origin: $E = \{x \in R^n | X^T M^{-1} X \leq 1\}$

References

[1] Jiří Matoušek and Bernd Gärtner, *Understanding and Using Linear Programming*, Ch 5, Ch. 6 and Ch. 7

[2] Edwin K. P. Chong and Stanislaw H. Zak, *An Introduction to Optimization*, Second Edition, Ch. 16 and Ch. 17