Linear Programming and Combinatorial Optimization

Siddhant Chaudhary

Abstract

These are my complementary notes for a course on *Linear Programming and Combinatorial Optimization*. The reference book used for the course was *Understanding and Using Linear Programming* by Jiri Matousek and Bernd Gartner.

Contents

January 2022

1. Linear Programming

1.1 Definition of an LP

1.1.1 The General Form. A *linear programming* problem (LP) is an optimization problem of the following form: we are given a set of variables $x_1, ..., x_n$, and we want to maximise a linear function in these variables subject to certain constraints. Formally, the problem is written as follows.

$$\begin{aligned} \text{Maximise:} \quad & c_1x_1 + \cdots + c_nx_n \\ \text{Subject to:} \quad & a_{11}x_1 + \cdots + a_{1n}x_n \leq b_1 \\ & a_{21}x_1 + \cdots + a_{2n}x_n \leq b_2 \\ & \vdots \\ & a_{m1}x_1 + \cdots + a_{mn}x_n \leq b_m \end{aligned}$$

This can also be written succinctly as follows.

$$\begin{aligned} \text{Maximise:} \quad & c^T x \\ \text{Subject to:} \quad & Ax \leq b \end{aligned}$$

Here $c = (c_1, ..., c_n)$ and $A = (a_{ij})$ is an $m \times n$ matrix. Note that in such a problem, the solution vector x is allowed to take values in $\mathbf{R}^n$. There is another variation of linear programs, called *integer linear programs*. Everything here is the same, except we have an additional condition, namely $x \in \mathbf{Z}^n$.

1.1.2 The Equational Form. Suppose we are given an LP in general form as mentioned in the above section. It can be shown that such an LP can be converted to an LP of the following form.

$$\begin{aligned} \text{Maximise:} \quad & c^T x \\ \text{Subject to:} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

The idea is to introduce new *slack variables* as follows. Consider some inequality constraint $a_{k1}x_1 + \cdots + a_{kn}x_n \leq b_k$, where $1 \leq k \leq m$. We introduce a new variable s_k along with the following two constraints.

$$\begin{aligned} a_{k_1}x_1 + \cdots + a_{kn}x_n + s_k &= b_k \\ s_k &\geq 0 \end{aligned}$$

This is just a fancy way of enforcing the originaly inequality constraint. We do this for all $1 \leq k \leq m$. However, the job is still not done, as the variables $x_1, ..., x_n$ are still allowed to take negative real values. To solve this problem, we do the following: for each $1 \leq i \leq n$, we introduce variables x_i^+ and x_i^- along with the following conditions.

$$x_i^+ \geq 0$$
$$x_i^- \geq 0$$

Finally, we substitute each x_i with the difference $x_i^+ - x_i^-$. As an exercise, try to prove that the new LP is indeed equivalent to the original LP.

1.2 Basic Feasible Solutions

In our analysis of LPs, we will assume that an LP is always in *equational form* (which was defined in the previous section). Moreover, any point x which satisfies $x \geq 0$ and $Ax = b$ is called a *feasible point.*

1.2.1 A basic assumption. Also, we will make the following assumptions on the matrix A.

(1) The equation $Ax = b$ has a solution. This can be checked efficiently using Gaussian elimination.

(2) We will assume that the rows of the matrix A are linearly independent. So, if A is an $m \times n$ matrix, then $\operatorname{rank}(A) = m$. Also, it must be true that $n \geq m$, because the column rank is also m.

1.2.2 Basic Feasible Solutions. Consider an LP of maximising $c^T x$ subject to $Ax = b$ and $x \geq 0$. A feasible solution x of this LP is said to be a *basic feasible solution* if there exists an m-element set $B \subset \{1, 2, ..., n\}$ such that the following are true.

(1) The columns of A indexed by B are linearly independent.

(2) $x_j = 0$ for all $j \notin B$.

1.2.3 Equivalence of two definitions of BFSs. In the previous section, we saw one definition of a *basic feasible solution.* Here is another definition: a feasible solution $v \in \mathbf{R}^n$ is said to be a *basic feasible solution* if the columns indexed by $K = \{i \mid v_i > 0\}$ are linearly independent.

Lemma 1.1. *The above two definitions of basic feasible solutions are equivalent.*

Proof. Suppose v is a feasible solution. If it satisfies the first definition, then it clearly satisfies the second definition. So we only need to prove the converse.

So suppose v satisfies the second definition. So, the columns indexed by the set K are linearly independent. First, suppose $|K| = m$. Then we can simply take $B = K$, and the first definition's conditions will follow. Also, note that $|K|$ cannot be $> m$ since the rank of A is m. So, the only case which is left to consider the the case when $|K| < m$. Now, because the column space of A has dimension m, there are m columns, say $A^{j_1}, ..., A^{j_m}$ of A which are linearly independent and span the column space. Let $K = \{i_1, ..., i_k\}$. We know that $A^{i_1}, ..., A^{i_k}$ are linearly independent. Now, we keep on adding a column among $A^{j_1}, ..., A^{j_m}$ to the set $\left\{A^{i_1}, ..., A^{i_k}\right\}$ until this new set remains linearly independent; clearly, when this process stops, we will have found a basis for the column space, and that can only happen at a point when we have added exactly $m - k$ elements to the set. In that case, we again will have found the set B. This completes the proof. ■

Remark 1.1.1. It follows that one can use either definition of a BFS. The first solution is helpful in computations, while the second definition if more helpful in proofs.

1.2.4 Optimum occurs at a BFS. In this section, we will mention one of the main results on the nature of solutions to an LP in equational form.

Theorem 1.2. *Suppose there exists $D \in \mathbf{R}$ such that $c^T v \leq D$ for every feasible solution v (in other words, suppose that the cost function is bounded). Then,*

(1) The LP has an optimum.

(2) The optimum occurs at a BFS.

Proof. This is **Theorem 4.2.3** of the reference book. The proof covered was the one in the book. ■

1.3 Simplex Algorithm

Throughout this section, we will assume that we are working with an LP in equational form, i.e we have an LP in which we have to maximise $c^T x$ subject to $Ax = b$ and $x \geq 0$. We will assume that x varies in n dimensional Euclidean space, i.e $x \in \mathbf{R}^n$. It's components will be denoted by $x_1, ..., x_n$. Also, A is an $m \times n$ matrix, and hence $b \in \mathbf{R}^m$.

Definition 1.1. A subset $B \subset \{1, ..., n\}$ is said to be *feasible basis* if $|B| = m$ and if it admits a basic feasible solution.

1.3.1 Tableau of a BFS. Let B be any feasible basis. A *simplex tableau* $\mathcal{T}(B)$ is a system of $m + 1$ linear equations in the variables $x_1, ..., x_n, z$ that has the same set of solutions as the system $Ax = b$, $z = c^T x$, and in matrix notation this system looks like the following.

$$\begin{array}{c} x_B = p + Qx_N \\ \hline z = z_0 + r^T x_N \end{array}$$

Above, x_B is the vector of the basic variables, $N = \{1, ..., n\} \setminus B$, x_N is the vector of non-basic variables, $p \in \mathbf{R}^m$ is any vector, $r \in \mathbf{R}^{n-m}$, $z_0 \in \mathbf{R}$ and Q is some $m \times (n - m)$ matrix.

Lemma 1.3. *For each feasible basis B, there exists exactly one tableau $\mathcal{T}(B)$, and it is given by the following.*

$$\begin{aligned} p &= A_B^{-1} b \\ Q &= -A_B^{-1} A_N \\ z_0 &= c_B^T A_B^{-1} b \\ r^T &= c_N^T - (c_B^T A_B^{-1} A_N) \end{aligned}$$

Proof. We know that B is a feasible basis. Note that the equation $Ax = b$ can be written as

$$A_B x_B + A_N x_N = b$$

Here, A_B is the $m \times m$ matrix in which only those columns indexed by B are taken, and similarly A_N is the $m \times (n - m)$ matrix consisting of the rest of the columns. This means that

$$A_B x_B = b - A_N x_N$$

Also, we know that A_B has rank m (since B is a feasible basis), and hence A_B is invertible. So, we get that

$$x_B = A_B^{-1} b - A_B^{-1} A_N x_N$$

So, for the tableau $\mathcal{T}(B)$, we take $p = A_B^{-1}b$ and $Q = -A_B^{-1}A_N$.

Now, also note that

$$\begin{aligned} c^T x &= c_B^T x_B + c_N^T x_N \\ &= c_B^T (A_B^{-1} b - A_B^{-1} A_N x_N) + c_N^T x_N \\ &= c_B^T A_B^{-1} b + (c_N^T - c_B^T A_B^{-1} A_N) x_N \end{aligned}$$

So, we take $z_0 = c_B^T A_B^{-1} b$ and $r^T = c_N^T - c_B^T A_B^{-1} A_N$. This proves the existence of a tableau $\mathcal{T}(B)$ for this feasible basis.

Uniqueness is easy to prove and I won't do it here (look at **Lemma 5.5.1** of the book for the proof). ∎

Lemma 1.4. *If $\mathcal{T}(b)$ is a simplex tableau such that $r \leq 0$, then the optimum cost is z_0 attained at the BFS corresponding to B.*

Proof. Note that for any feasible solution, the cost is given by

$$z = z_0 + r^T x_N$$

Since $x_N \geq 0$ and $r \leq 0$, the cost is bounded above by z_0. Since the BFS corresponding to B attains z_0, this is the optimum. ∎

1.3.2 The Pivoting Step. It can be shown that the pivoting step yields another feasible basis, but I didn't have time to write down a proof of this here. Checkout **Lemma 5.6.1** of the book for this.

1.3.3 Bland's Rule. In this section, we will look at *Bland's Rule*, which is a pivoting rule to avoid cycling between degenerate cases. The rule is quite simple: first, we arbitrarily index the variables using indices in $\{1, ..., n\}$ (if there are n variables). At any pivoting step, if there is a choice of variables to bring into the basis, we pick the one with the least index. This is known as *Bland's rule*. It can be shown that this rule prevents any degenerate cycles (but I won't do this here).

1.4 Duality of LPs

1.4.1 The Primal and Dual of an LP. Suppose we have an LP in which we want to maximise $c^T x$ subject to the conditions $Ax \leq b$ and $x \geq 0$, where A is some $m \times n$ matrix, and b is an $m \times 1$ column vector. This LP is called a *primal* LP. Writing everything in coordinate form, the goal of the LP is to maximise

$$c_1 x_1 + \cdots + c_n x_n$$

subject to the conditions

$$\begin{aligned} a_{11} x_1 + \cdots a_{1n} x_n &\leq b_1 \\ a_{21} x_1 + \cdots a_{2n} x_n &\leq b_2 \\ &\vdots \\ a_{m1} x_1 + \cdots a_{mn} x_n &\leq b_m \\ x_1, ..., x_n &\geq 0 \end{aligned}$$

The *dual* LP of the above *primal* LP is defined as follows: the *dual* LP has m variables $y_1, ..., y_m$, and the goal is to minimize

$$b_1y_1 + \cdots + b_my_m$$

subject to the following constraints.

$$\begin{aligned}
a_{11}y_1 + \cdots + a_{m1}y_m &\geq c_1 \\
a_{12}y_1 + \cdots + a_{m2}y_m &\geq c_2 \\
&\vdots \\
a_{1n}y_1 + \cdots + a_{mn}y_m &\geq c_n
\end{aligned}$$

In matrix form, the LP is the following.

$$\begin{array}{rccc}
\text{Minimize:} & b^Ty & & \\
\text{Subject to:} & A^Ty & \geq & c \\
& y & \geq & 0
\end{array}$$

1.4.2 The Weak Duality Theorem. Consider a primal LP

$$\begin{array}{rccc}
\text{Maximize:} & c^Tx & & \\
\text{Subject to:} & A^Tx & \leq & b \\
& x & \geq & 0
\end{array}$$

and it's corresponding dual LP

$$\begin{array}{rccc}
\text{Minimize:} & b^Ty & & \\
\text{Subject to:} & A^Ty & \geq & c \\
& y & \geq & 0
\end{array}$$

Theorem 1.5 (**Weak Duality Theorem**)**.** *For every feasible solution x' of the primal LP and for every feasible solution y' of the dual LP, the following holds.*

$$c^Tx' \leq b^Ty'$$

In other words, the cost of the primal LP at x' is always bounded above by the cost of the dual LP at y'.

Proof. The proof of this is quite straightforward. First, we have the following by the definition of the dual LP.

$$\begin{aligned}
c^Tx' &= c_1x'_1 + \cdots + c_nx'_n \\
&\leq \sum_{i=1}^{n}(a_{1i}y'_1 + \cdots + a_{mi}y'_m)x'_i \\
&= \sum_{j=1}^{m}(a_{j1}x'_1 + \cdots + a_{jn}x'_n)y'_j \\
&\leq \sum_{j=1}^{m} b_jy'_j \\
&= b^Ty'
\end{aligned}$$

This proves the claim. ∎

Corollary 1.5.1. *The previous theorem implies the following.*

(1) If the primal LP is unbounded above, the dual LP is infeasible.
(2) If the dual LP is unbounded below, the primal is infeasible.

1.4.3 A Dualization Reciple. Infact, for any primal LP (not necessarily in the form we looked at in the previous section), a dual LP can be constructed. This dualization recipe is given in section 6.2 in a great way; check that out. The **Weak Duality Theorem** 1.5 goes through for any dual LP, and the proof is very similar.

1.4.4 The Strong Duality Theorem. In this section, we will look at a stronger version of the **Weak Duality Theorem** 1.5 that we saw before.

Theorem 1.6 (**Strong Duality Theorem**). *Consider any primal LP (which we'll denote by P) and it's dual LP (which we'll denote by D); we assume that P is a maximization problem, while D is a minimization problem. Then, only the following can occur.*
(1) Either both P and D are infeasible.
(2) P is unbounded above and D is infeasible.
(3) P is infeasible and D is unbounded below.
(4) Both P and D have optima, and their optimal costs are equal.

The proof of this theorem is a bit involved; just look at section 6.4 of the book.

1.4.5 Farkas' Lemma and the Convex Cone. *Farkas' Lemma* is another useful tool, which leads to a neat proof of the **Strong Duality Theorem** 1.6.

Proposition 1.7 (**Farkas' Lemma**). *Let A be a real matrix with m rows and n columns. Let $b \in \mathbf{R}^m$ be a vector. Then, exactly one of the following two possibilities occur.*
1. *There exists a vector $x \in \mathbf{R}^n$ satisfying $Ax = b$ and $x \geq 0$.*
2. *There exists a vector $y \in \mathbf{R}^m$ satisfying $y^T A \geq 0$ and $y^T b < 0$.*

Definition 1.2. Let $a_1, ..., a_n \in \mathbf{R}^m$ be a set of points. The *convex cone* generated by this set of points is defined to be the following set.

$$\{t_1 a_1 + \cdots + t_n a_n \mid t_1, ..., t_n \geq 0\}$$

So, the *convex cone* is the set of all non-negative linear combinations of the points.

Convex cones give a very nice geometric interpretation of **Farkas's Lemma**. To be more precise, the lemma says that either the point b lies in the convex cone generated by the columns of A, or there is a hyperplane passing through the origin that separates the columns of A and the point b.

We easily see that the following proposition is equivalent to **Proposition** 1.7.

Proposition 1.8 (**Variants of Farkas' Lemma**). *Let A be an $m \times n$ real matrix, and let $b \in \mathbf{R}^m$. The following statements are equivalent to each other, and are also equivalent to Farkas's Lemma.*
(1) $Ax = b$ has a non-negative solution iff. for every $y \in \mathbf{R}^m$ such that $y^T A \geq 0$, we have $y^T b \geq 0$.
(2) $Ax \leq b$ has a non-negative solution iff. for every non-negative $y \in \mathbf{R}^m$ such that $y^T A \geq 0$, we have $y^T b \geq 0$.
(3) $Ax \leq b$ has a solution iff. for every non-negative $y \in \mathbf{R}^m$ such that $y^T A = 0$, we have $y^T b \geq 0$.

Proof. First, we show that statement (1) is equivalent to **Proposition** 1.7, and then we will prove that all of the above statements are equivalent to one another.

First, suppose **Proposition** 1.7 is true. Given this, if $Ax = b$ has a non-negative solution, then there cannot be any vector $y \in \mathbf{R}^m$ satisfying $y^T A \geq 0$ and $y^T b < 0$; hence, for all

y with $y^T A \geq 0$, it must be true that $y^T b \geq 0$, and this proves one direction of (1). The converse of this is similarly shown. So, it follows that **Proposition** 1.7 implies (1). It is not hard to show that (1) implies **Proposition** 1.7 as well.

For the rest of the proofs, refer to **Proposition 6.4.3** of the main book. ∎

1.4.6 Complementary Slackness. Consider the LP given by

$$\begin{array}{rccc} \text{Maximize:} & c^T x & & \\ \text{Subject to:} & Ax & \leq & b \end{array}$$

and it's dual, which is given by

$$\begin{array}{rccc} \text{Minimize:} & b^T y & & \\ \text{Subject to:} & A^T y & = & c \\ & y & \geq & 0 \end{array}$$

We will now prove a theorem. Throughout our discussion, we will assume that the first LP is the primal (P) and that the second LP is it's dual (D).

Theorem 1.9. *Let x_0, y_0 be feasible solutions of the primal and the dual respectively. Then,*

$$x_0 \text{ and } y_0 \text{ are optima} \iff c^T x_0 = b^T y_0$$

Proof. First, suppose x_0 and y_0 are optima of the corresponding LPs. From the strong duality theorem, the optimum costs of the primal and the dual are equal. And hence,

$$c^T x_0 = b^T y_0$$

and the forward direction is proven.

Now, suppose x_0, y_0 are feasible solutions of the primal and the dual such that $c^T x_0 = b^T y_0$. Now, note that

(1) Both primal and the dual must have optima, since the very existence of x_0, y_0 shows that the primal and the dual are feasible. Infact, the above equality shows that both are bounded (by the **Weak Duality Theorem** 1.5).
(2) By the same **Weak Duality Theorem** 1.5), note that for all feasible solutions x of the primal, we must have $c^T x \leq b^T y_0$. Hence, x_0 is really the optimum point of the primal.
(3) Similarly, it can be concluded that y_0 is the optimum point of the dual.

∎

Theorem 1.10 (Complementary Slackness Condition). *Let the LPs be as above, and let x_0, y_0 be feasible solutions of the primal and the dual respectively. Then, $c^T x_0 = b^T y_0$ if and only if*

$$(y_0)_i > 0 \implies A_i x_0 = b \qquad \forall i \in \{1, 2, ..., m\}$$

Here, A_i is the ith row of A.

Remark 1.10.1. Another way of saying this theorem in words is the following.

(1) If the dual variable is *slack*, then the primal inequality is *tight*.
(2) If the primal inequality is *slack*, then the dual variable is *tight*.

Proof. Assume that the matrix A is $m \times n$.

First, we prove the forward direction. So, assume that $c^T x_0 = b^T y_0$. We know that $A^T y_0 = c$, since y_0 is a feasible point of the dual. This clearly means that $A^i y_0 = c_i$, where A^i is the ith column of A. So, we have the following.

$$\begin{aligned} c^T x_0 &= \sum_{i=1}^{n} c_i (x_0)_i \\ &= \sum_{i=1}^{n} (A^i y_0)(x_0)_i \end{aligned}$$

Now, the last sum can be rearranged as follows.

$$\sum_{i=1}^{n} (A^i y_0)(x_0)_i = \sum_{i=1}^{m} (A_i x_0)(y_0)_i$$

where A_i is the ith row of A. Now, we're given that $c^T x_0 = b^T y_0$. So, we get that

$$\sum_{i=1}^{m} (A_i x_0)(y_0)_i = \sum_{i=1}^{m} (y_0)_i b_i$$

and rearranging, we get

$$\sum_{i=1}^{m} (y_0)_i (b_i - A_i x_0) = 0$$

Now, because $y_0 \geq 0$ and because $b_i - A_i x_0 \geq 0$ for each i, this is only possible if

$$(y_0)_i (b_i - A_i x_0) = 0$$

for all i. This proves the forward direction.

Now, let us prove the reverse direction. We know from weak duality that

$$c^T x_0 \leq b^T y_0$$

And, because of the fact that $A^T y_0 = c$, we have the following

$$c^T x_0 - b^T y_0 = \sum_{i=1}^{n} (A^i y_0)(x_0)_i - \sum_{i=1}^{m} b_i (y_0)_i$$

As before, rearranging the terms to get the $(y_0)_i$s together, we get the following.

$$\begin{aligned} c^T x_0 - b^T y_0 &= \sum_{i=1}^{m} (A_i x_0)(y_0)_i - \sum_{i=1}^{m} b_i (y_0)_i \\ &= \sum_{i=1}^{m} (y_0)_i (A_i x_0 - b_i) \end{aligned}$$

From our hypothesis, we know that $(y_0)_i > 0 \implies A_i x_0 = b_i$. So, this means that

$$c^T x_0 = b^T y_0$$

and this completes the proof. ■

1.4.7 Complementary slackness for other primal-dual pairs. For this section, we consider the following primal-dual pair. The primal LP is given by Consider the LP given by

$$\begin{array}{rccc} \text{Maximize:} & c^T x & & \\ \text{Subject to:} & Ax & \leq & b \\ & x & \geq & 0 \end{array}$$

and it's dual is given by

$$\begin{array}{rccc} \text{Minimize:} & b^T y & & \\ \text{Subject to:} & A^T y & \geq & c \\ & y & \geq & 0 \end{array}$$

A complementary slackness theorem can be stated for this pair too.

Theorem 1.11. *Let x_0, y_0 be feasible solutions of the primal and the dual LPs respectively. Then, $c^T x_0 = b^T y_0$ iff. the following hold.*

- $(y_0)_i > 0 \implies A_i x_0 = b_i$ *for all* $i \in \{1, ..., m\}$.
- $(x_0)_j > 0 \implies A_j^T y_0 = c_j$ *for all* $j \in \{1, ..., n\}$.

Here, A_i and A_j^T are the ith and the jth rows of A and A^T respectively.

Proof. Let us prove the reverse direction first. So, suppose both the points hold. We want to show that $c^T x_0 = b^T y_0$. First, note that

$$c^T x_0 = \sum_{j=1}^{n} c_j (x_0)_j$$

From the second implication, we know that if $(x_0)_j > 0$, then $c_j = A_j^T y_0$. So, we see that

$$c^T x_0 = \sum_{j=1}^{n} (A_j^T y_0)(x_0)_j$$

The left hand side can be rearranged in terms of $(y_0)_1, ..., (y_0)_m$, and we get

$$c^T x_0 = \sum_{i=1}^{m} (A_i x_0)(y_0)_i$$

Then, from the first implication, this in turn can be written as

$$\begin{aligned} \sum_{i=1}^{m} (A_i x_0)(y_0)_i &= b_1 (y_0)_1 + \cdots + b_m (y_0)_m \\ &= b^T y_0 \end{aligned}$$

and this proves that $c^T x_0 = b^T y_0$.

Now, let us prove the forward direction. So, suppose $c^T x_0 = b^T y_0$. Since y_0 is a feasible point of the dual, we know that $c \leq A^T y$, which means that $c_j \leq A_j^T y$ for all $j \in \{1, ..., n\}$. Clearly, this means that

$$\begin{aligned} c^T x_0 &\leq (A_1^T y_0)(x_0)_1 + \cdots + (A_n^T y_0)(x_0)_n \\ &= (A_1 x_0)(y_0)_1 + \cdots + (A_m x_0)(y_0)_m \\ &\leq b_1 (y_0)_1 + \cdots + b_m (y_0)_m \\ &= b^T y_0 \end{aligned}$$

where the last inequality is true because x_0 is a feasible point of the primal LP. But, because we know that $c^T x_0 = b^T y_0$, it follows that

$$\begin{aligned} c^T x_0 &= \sum_{i=1}^{m} (A_i x_0)(y_0)_i \\ &= b^T y_0 \end{aligned}$$

From here, the same proof trick as in the proof of **Theorem** 1.10 can be repeated to prove the claim. ■

Finally, here's a general complementary slackness theorem that we will state without proof (although the proof uses ideas similar to those above).

Theorem 1.12. *Let x_0, y_0 be feasible solutions of primal and dual LPs respectively. Then, $c^T x_0 = b^T y_0$ iff.*
1. $(y_0)_i(A_i x_0 - b_i) = 0$ *for all* $i \in \{1, ..., m\}$.
2. $(x_0)_j(A_j^T y_0 - c_j) = 0$ *for all* $j \in \{1, ..., n\}$.

2. Zero Sum Games

2.1 Introduction

2.1.1 Introducing the model. Here we will introduce the model under which we will work. There are two players: the *maximizer* and the *minimizer*. The maximizer has m strategies, which we will label with $\{1, ..., m\}$. Similarly, the minimizer has n strategies, labelled by $\{1, ..., n\}$. We are also given an $m \times n$ *payoff matrix*; the entry m_{ij} for $1 \leq i \leq m$, $1 \leq j \leq n$ tells us that when the maximizer plays strategy i and the minimizer plays strategy j, the *payoff* is m_{ij}, i.e the maximizer receives m_{ij} from the minimizer (if $m_{ij} < 0$, then the maximizer has to give this quantity to the minimizer). The goal of the maximizer is to *maximize* it's payoff, and the goal of the minimizer is to *minimize* it's payoff (and hence the names). These kind of games are called *zero-sum games* because a player's loss is the other player's gain.

2.1.2 Maxmin and Minmax. Given a game represented by payoff matrix M, we define the following two quantities.

$$\begin{aligned} \text{max-min-pure}(M) &= \max_{i \in \{1,...,m\}} \min_{j \in \{1,2,...,n\}} m_{ij} \\ \text{min-max-pure}(M) &= \min_{j \in \{1,...,n\}} \max_{i \in \{1,2,...,m\}} m_{ij} \end{aligned}$$

The choices $1, .., m$ for the maximizer are called her *pure* or *deterministic* strategies. The same definition holds for the choices $1, ..., n$ for the minimizer.

Lemma 2.1. *For any game with payoff matrix M, we have*

$$\text{max-min-pure}(M) \leq \text{min-max-pure}(M)$$

Proof. The proof is straightforward. If we fix a row i, then note that

$$\min_{j \in \{1,...,n\}} m_{ij} \leq \min_{j \in \{1,...,n\}} \max_{i \in \{1,...,m\}} m_{ij}$$

Taking the max over the left hand side, we can conclude the proof. ■

2.1.3 Saddle Points. An entry m_{kl} of the payoff matrix M is said to be a *saddle point* for M if

$$m_{kl} = \min_{j \in \{1,..,n\}} m_{kj} = \max_{i \in \{1,...,m\}} m_{il}$$

In simple words, a *saddle point* is an entry of the payoff matrix which is the minimum entry of it's row and the maximum entry of it's column.

Proposition 2.2. *For any game, max-min-pure=min-max-pure iff there is a saddle point.*

Proof. First, suppose for a game with payoff matrix M, we have that

$$v^+ = \min_{j \in \{1,...,n\}} \max_{i \in \{1,...,m\}} m_{ij} = \max_{i \in \{1,...,m\}} \min_{j \in \{1,...,n\}} m_{ij} = v^-$$

Now, let j^* be the specific column such that $v^+ = \max_{i \in \{1,...,m\}} m_{ij^*}$ and let i^* be the specific row such that $v^- = \min_{j \in \{1,...,n\}} m_{i^*j}$. We will show that $m_{i^*j^*}$ is a saddlepoint.

Note that for any $i \in \{1, ..., n\}$ and $j \in \{1, ..., m\}$ we have the following.

$$m_{i^*j} \geq \min_{k \in \{1,...,n\}} m_{i^*k} = v^- = v^+ = \max_{k \in \{1,...,m\}} m_{kj^*} \geq a_{ij^*}$$

In the above inequality, take $j = j^*$ on the left hand side and take $i = i^*$ on the right hand side to obtain

$$m_{i^*j^*} \geq v^- = v^+ \geq m_{i^*j^*}$$

which implies that $m_{i^*j^*} = v^+ = v^-$. This also shows that for any $i \in \{1, ..., m\}$ and $j \in \{1, ..., n\}$ we have

$$m_{i^*j} \geq m_{i^*j^*} \geq m_{ij^*}$$

and this shows that $m_{i^*j^*}$ is a saddle point.

Conversely, suppose there is a saddle point in M. Suppose $m_{i^*j^*}$ is a saddle point. Then we have the following.

$$v^+ = \min_{j \in \{1,...,n\}} \max_{i \in \{1,...,m\}} m_{ij} \leq \max_{i \in \{1,...,m\}} m_{ij^*} = m_{i^*j^*} = \min_{j \in \{1,...,n\}} m_{i^*j} \leq \max_{i \in \{1,...,m\}} \min_{j \in \{1,...,n\}} m_{ij} = v^-$$

which shows that $v^- \geq v^+$. But, we already know that $v^- \leq v^+$ always holds. So, it must be true that $v^+ = v^-$. This completes the proof. ■

Remark 2.2.1. Such a situation is said to be a *Nash equilibrium*.

2.1.4 Mixed Strategies. A *mixed strategy* is a probability distribution over pure strategies. For instance, if the maximizer's pure strategies are $\{1, ..., m\}$, then a *mixed strategy* for the maximizer will be a probability distribution $(x_1, ..., x_m)$ over these strategies. A similar case holds for the minimizer's strategies.

Given mixed strategies $\sigma := (x_1, ..., x_m)$ and $\tau := (y_1, ..., y_n)$ for the maximizer and the minimizer, the *expected payoff* is defined as follows.

$$\text{Payoff}(\sigma, \tau) := \sum_{i \in \{1,...,m\}} \sum_{j \in \{1,...,n\}} x_i y_j m_{ij} = x^T M y$$

Next, suppose the strategy x of the maximizer is fixed. We are interested in finding

$$\min_y x^T M y$$

i.e the best strategy for the minimizer. Since we have fixed x, this becomes a linear program as follows.

$$\begin{array}{lccccccccc} \text{Minimize:} & x^T M y & & & & & & & & \\ \text{Subject to:} & y_1 & + & y_2 & + & \cdots & + & y_n & = & 1 \\ & & & y_1 & y_2, & \ldots, & & y_n & \geq & 0 \end{array}$$

The dual of the above primal LP is the following LP, which can be easily checked.

$$\begin{array}{lcl} \text{Maximize:} & x_0 & \\ \text{Subject to:} & \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} x_0 & \leq \quad M^T x \end{array}$$

Now, observe that the primal LP is bounded, because the feasible region (which is the simplex) is compact, and hence the primal LP has an optimum. By the **Strong Duality Theorem** 1.6, we see that the optimum of the primal LP is equal to the optimum of the dual LP.

To put everything together, for a fixed x, to find the value

$$\min_y x^T M y$$

it is enough to solve the dual LP (which has no mention of the strategy y of the minimizer). So, in the dual LP, we treat x as a variable itself, i.e we add the conditions of x to the dual LP to get the following LP.

$$\begin{array}{lcccccccccc} \text{Maximize:} & x_0 & & & & & & & & & \\ \text{Subject to:} & \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} x_0 & \leq & M^T x & & & & & & & \\ & x_1 & + & x_2 & + & \cdots & + & x_m & = & 1 \\ & & & x_1, & x_2, & \ldots, & & x_m & \geq & 0 \end{array}$$

The optimum of the above LP gives us the following quantity.

$$\max_x \min_y x^T M y = \text{max-min}(M)$$

Now, let's try to repeat the same process for the case when the strategy y of the minimizer is fixed. In that case, we are interested in finding

$$\max_x x^T M y$$

i.e the best strategy for the maximizer. Again, because y is fixed, this can be written as the following linear program.

$$\begin{array}{lccccccccc} \text{Maximize:} & x^T M y & & & & & & & & \\ \text{Subject to:} & x_1 & + & x_2 & + & \cdots & + & x_m & = & 1 \\ & & & x_1 & x_2, & \ldots, & & x_m & \geq & 0 \end{array}$$

Let's now write down the dual of the above primal LP, which is easily seen to be the following LP.

$$\begin{array}{lcl} \text{Minimize:} & y_0 & \\ \text{Subject to:} & \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} y_0 & \geq \quad M y \end{array}$$

Note that above, there is no mention of the vector x. So, again, we treat y as a variable in the above LP to get the following LP.

$$\begin{array}{rl} \text{Minimize:} & y_0 \\ \text{Subject to:} & \begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} y_0 \geq My \\ & y_1 + y_2 + \cdots + y_n = 1 \\ & y_1, y_2, \ldots, y_n \geq 0 \end{array}$$

The optimum of the above LP gives us the following quantity.

$$\min_y \max_x x^T My = \text{min-max}(M)$$

Finally, it can be checked that the two duals we have derived are actually form a primal-dual pair. Since both have optimums, their optimum is actually the same, and this gives us the following equation for mixed strategies.

$$\text{max-min}(M) = \text{min-max}(M)$$

3. Combinatorial Optimization

3.1 Integer Linear Programs

3.1.1 Overview. An *integer linear program* (ILP) is just a linear program with an additional integrality constraint on the variables. This is really all that is there to these problems.

3.1.2 Complexity of solving ILPs. It can be shown that the decision version of the problem of solving ILPs is NP-complete (we won't show this here). So, our goal will be to approximate the solution of an ILP using LP techniques. This will be our focus for the next section.

3.1.3 Totally Unimodular Matrices. Let A be an $m \times n$ matrix. A is said to be *totally unimodular* if all *square submatrices* of A have determinant in the set $\{-1, 0, 1\}$. A *square submatrix* is a square matrix obtained from A by deleting some rows or columns.

Proposition 3.1. *If A is a totally unimodular matrix, then all entries of A are in the set $\{-1, 0, 1\}$.*

Proof. This is clear from the definition; if we remove all rows and all columns barring a single cell of the matrix, we can conclude the claim. ■

Example 3.1. Let $G = (V, E)$ be a bipartite graph. Consider the problem of finding the size of a maximum matching in G. This problem can be formulated as an ILP as follow: for each edge $e \in E$, we will have a variable x_e. The constraint on x_e will be that $x_e \in \mathbf{Z}$ and that $0 \leq x \leq 1$. Then, for a vertex v, we will have the constraint

$$\sum_{e \in E \text{ s.t } e \text{ is incident on } v} x_e \leq 1$$

We can think of each variable x_e as an indicator variable, denoting whether the edge e is chosen or not. Finally, the objective that we want to maximise will be

$$\sum_{e \in E} x_e$$

It can be shown that the matrix corresponding to this LP is totally unimodular (Show this! Induction might be used here; induction on the size of the square submatrix).

Theorem 3.2. *Let A be an $m \times n$ totally unimodular matrix. Consider the following ILP.*

$$\begin{array}{rccc} \text{Maximize:} & c^T x & & \\ \text{Subject to:} & Ax & \leq & b \\ & x & \geq & 0 \\ & x & \in & \mathbf{Z}^n \end{array}$$

If $b \in \mathbf{Z}^m$, then every BFS of this ILP is integral.

Proof. First, we claim that if we add a standard basis vector to A as a column, then the resultant matrix will be totally unimodular (here by a standard basis vector we mean the usual basis $\{e_1, ..., e_n\}$). This is actually very easy to see, and we will not explicitly write the proof.

Now, having the above fact in our toolbox, we do the following: first, we convert the given ILP to equational form by introducing m slack variables; the resultant matrix will be the matrix

$$\begin{bmatrix} A & I_{m \times m} \end{bmatrix} \tag{1}$$

and this matrix will be totally unimodular by the above claim. Now, let B be the set of basic coordinates, and let N be the set of non-basic. So, in any BFS, every variable in N is zero, and we have

$$x_B = A_B^{-1} b$$

where A_B is the matrix obtained by the columns indexed by B. Then, for each i in B, we will have

$$(x_B)_i = \frac{\det(i(A_B, b))}{\det(A_B)}$$

by Cramer's rule. Here, $i(A_B, b)$ is the matrix A_B with the ith column replaced by the vector b. Since b is also integral, and since $\det(A_B)$ is in $\{-1, 1\}$, we see that x_B is also integral. This completes the proof. ∎

Corollary 3.2.1. *If A is unimodular, then every ILP with matrix A and it's corresponding LP (without the integrality constraint) have the same optimum, provided the the LP is unbounded.*

3.1.4 Matchings, Covers and Kőnig's Theorem. In this section, we will prove a useful theorem using LP techniques.

Theorem 3.3 (Kőnig). *Let $G = (X \cup Y, E)$ be a bipartite graph. The size of a maximum matching equals the size of a minimum vertex cover.*

Proof. Recall from **Example** 3.1 the ILP for maximum matchings. We write it again here for the sake of brevity: we want to maximise $\sum_{e \in E} x_e$ such that for any vertex v,

$$\sum_{e \text{ incident on } v} x_e \leq 1$$

and $x_e \in \{0, 1\}$ for all $e \in E$.

Now, let us write down the ILP for computing the size of a minimum cover. For each vertex v, we maintain a variable x_v; we then want to minimize

$$\sum_{v \in V} x_v$$

such that $x_u + x_v \geq 1$ for all edges $(u, v) \in E$, and $x_v \in \{0, 1\}$ for all $v \in V$.

Now we have two ILPs. The idea will be to convert these to LPs, and show that they are duals of each other. Since the matrices in both the LPs will be totally unimodular, we can conclude the statement of the theorem.

For the first ILP, we remove the integrality constraint $x_e \in \{0, 1\}$, and instead we add the constraint $0 \leq x_e$ (note that $x_e \leq 1$ is already forced by the summation constraint). For the second ILP, we remove the integrality constraint $x_v \in \{0, 1\}$, and instead we add the constraint $x_v \geq 0$. That the resultant LPs are duals of each other is not hard to see; the statement of the theorem follows. ■

Next, consider the case of non-bipartite graphs. It turns out that **Konig's Theorem** 3.3 doesn't hold anymore; for a counterexample, consider a triangle. If one investigates the incidence matrix of a triangle, then one can see that the incidence matrix is not totally unimodular.

3.1.5 Minimum vertex covers for general graphs. Now, again consider the case of a general graph. In this section, we will see how the solution of the ILP for the minimum vertex cover problem and the solution of it's relaxation to an LP are related. Recall that the ILP is to minimize

$$\sum_{v \in V} x_v$$

such that $x_u + x_v \geq 1$ for all edges $(u, v) \in E$, and $x_u \in \{0, 1\}$. The relaxed LP for this ILP is the same problem, with the integrality constraint removed, and the constraint $x_v \geq 0$ added for all $v \in V$.

Now, suppose x^* is the optimum of the LP. We claim that all coordinates of x^* are atmost 1. Suppose not, and let $v \in V$ be such that $x_v^* > 1$. In that case, reducing x_v^* to 1 gives us a feasible solution with a *lesser* cost, which is a contradiction. So, it certainly is true that all coordinates of x^* are atmost 1.

Now, consider the set S_{LP} defined as follows.

$$S_{LP} := \left\{ v \in V \mid x_v^* \geq \frac{1}{2} \right\}$$

Here, the coordinates x_v^* are the coordinates of the optimal solution x^* of the LP. We claim that S_{LP} is a vertex cover, which is pretty easy to see: note that for each edge (u, v), we must have $x_u^* + x_v^* \geq 1$. Hence, atleast one of x_u^* or x_v^* is $\geq \frac{1}{2}$, and hence one of u or v is in S_{LP}.

Next, we claim that

$$|S_{LP}| \leq 2|S_{ILP}|$$

where $|S_{ILP}|$ is the size of a minimum vertex cover of the graph. So suppose y^* is the ILP optimum. Clearly, we see that

$$\sum_{v\in V} x_v^* \leq \sum_{v\in V} y_v^*$$

Now, observe that

$$|S_{LP}| \leq 2\sum_{v\in V} x_v^*$$

and this just follows from the definition of S_{LP}. So, it follows that

$$|S_{LP}| \leq 2\sum_{v\in V} x_v^* \leq 2\sum_{v\in V} y_v^* = 2|S_{ILP}|$$

and this completes the proof.

3.2 Primal Dual Algorithms

In combinatorial optimization, a *primal dual algorithm* is an algorithm which solves a combinatorial problem by making use of a pair of primal-dual LPs. In this section, we will see examples of this.

3.2.1 A general template. Suppose that the primal LP is of the form

$$\begin{array}{rccc} \text{Minimize:} & c^T x & & \\ \text{Subject to:} & Ax & \geq & b \\ & x & \geq & 0 \end{array}$$

and it's dual will be given by

$$\begin{array}{rccc} \text{Maximize:} & b^T y & & \\ \text{Subject to:} & A^T y & \leq & c \\ & y & \geq & 0 \end{array}$$

The general strategy is the following.

(1) Frame the optimization problem as an integer linear program (ILP).
(2) Generate an LP from the given ILP.
(3) Write down the dual of the LP.
(4) Find a feasible solution y_0 of the dual. Typically for the problems under consideration, $c \geq 0$ and hence $y = 0$ will be feasible. This is called the *initialization step* (or the zeroth step).
(5) After the ith step, say we have a feasible solution y_i. Let $(A^T)'$ denote the rows of the dual that are tight at y_i, i.e

$$(A^T)' y_i = c'$$

where c' is c restricted to the rows in $(A^T)'$.
(6) If possible, find a $\overline{y}$ such that $(A^T)'\overline{y} \leq 0$ and $b^T\overline{y} > 0$.
(7) Then, find the greatest $\epsilon > 0$ such that

$$A^T(y_i + \epsilon\overline{y}) \leq c$$

(8) Set $y_{i+1} = y_i + \epsilon \overline{y}$.
(9) When the iterative step cannot be performed anymore, we *terminate*.
(10) From the final y that is obtained at the end of the algorithm, generate a primal solution x and use either complementary slackness or strong duality to show that x and y are optima of the primal and the dual respectively.

3.2.2 Shortest Paths. We are given a directed graph G, and for every edge a non-negative weight. There are source and target vertices s and t. Our goal is to compute the weight of the shortest path from s to t. We will also assume that s only has outgoing edges, and t only has incoming edges.

We use an incidence matrix to represent the graph. Here, the rows of the matrix will be indexed by the vertices, and the columns will be indexed by the edges. For a pair (u, e) of a vertex and an edge, the entry in row u and column e of the matrix will contain
(1) A $+1$, if u is the source vertex of the edge.
(2) A -1, if u is the target vertex of the edge.
(3) 0 if u is not a part of the edge.

Let us now write the shortest path problem as an ILP. For each edge $e \in E$, we will have a variable x_e which takes value in the set $\{0, 1\}$. Now, note that a subset P of edges is a path from s to t iff.
- there is exactly one outgoing edge from s in P.
- for every $v \notin \{s, t\}$ in the path, the number of incoming edges to v in P equals the number of outgoing edges from v in P.
- There is exactly one incoming edge to t in P.

Also, we assume that the graph has m vertices and n edges. So, the incidence matrix has dimension $m \times n$. Then, the above constraint can be represented as the equation

$$A \begin{bmatrix} x_{e_1} \\ x_{e_2} \\ x_{e_3} \\ x_{e_4} \\ \vdots \\ x_{e_n} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ -1 \end{bmatrix}$$

Our objective will be to minimize the cost

$$\sum_{e \in E} c_e x_e$$

Now with our ILP set up, we will form an LP out of it. To do this, we will relax the inequality constraint to

$$0 \leq x_e \leq 1$$

Infact, we will remove the $x_e \leq 1$ constraint, and we will just put $0 \leq x_e$. It will turn out that for the final LP that we will obtain, there is an optimum that assigns either 0 or 1 to each x_e, and hence the constraint $x_e \leq 1$ can be safely removed.

So to put everything together, we have the following LP.

$$\begin{array}{lccc} \text{Minimize:} & \sum_e c_e x_e & & \\ \text{Subject to:} & Ax & = & \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ -1 \end{bmatrix} \\ & x & \geq & 0 \end{array}$$

Moreover, we know that

$$\text{optimum}_{LP} \leq \text{optimum}_{ILP}$$

We can see that the dual of the above LP is the following.

$$\begin{array}{lccc} \text{Maximize:} & y_s - y_t & & \\ \text{Subject to:} & y_u - y_v & \leq & c_{uv}, \text{ for all edges } (u,v) \\ & y & \geq & 0 \end{array}$$

We can now repeat the steps of the general template and apply it to this primal-dual pair (This needs to be completed!).

3.2.3 Minimum Vertex Cover. We are given an undirected graph $G = (V, E)$ with weights on vertices $w : V \to \mathbf{N}$. Our goal is to find a vertex cover of minimum weight. In this section, we will see how to come up with a primal-dual algorithm for this problem.

It turns out that the decision version of this problem is **NP**-hard. So, we will see that the primal-dual algorithm that we come up with gives us a 2-approximate solution to the problem.

Now, let us construct the ILP for our problem. For each vertex u, we will maintain a variable x_u that will take a value in $\{0, 1\}$. Our constraints will be

$$x_u + x_v \geq 1 \qquad \forall (u,v) \in E$$

and the objective that we want to minimize will be

$$\sum_{u \in V} w_u x_u$$

Here, our matrix is the transpose of the incidence matrix, in which the rows are indexed by the edges, and the columns indexed by the vertices. The entry (e, u) of the matrix will be 1 iff. edge e is incident on u.

Next, as per our primal-dual template, we try to relax this ILP. To do this, we will remove the integrality constraint, and add the constraint $0 \leq x_u \leq 1$ for each vertex u. Infact, we will even remove the constraint $x_u \leq 1$, and only have $x_u \geq 0$ for each vertex u. This will be our primal LP. Again, note that

$$\text{optimum}_{LP} \leq \text{optimum}_{ILP}$$

Let us now compute the dual of the above primal LP. For each edge $e \in E$, we will have a variable y_e, and for each vertex u we will have the constraint

$$\sum_{e \text{ incident on } u} (y_i)_e \leq w_u$$

and we will have $y_e \geq 0$ for each $e \in E$. Our objective will be to maximise the quantity

$$\sum_{e \in E} y_e$$

This LP is our dual LP.

Now, we pick $y_0 = 0$ as our initial feasible point of the dual. Suppose we have some y_i, and we want to compute y_{i+1}. Let $T \subseteq V$ be the set of all vertices for which the dual constraint is tight, i.e for all $u \in T$,

$$\sum_{e \text{ incident on } u} y_e = w_u$$

Next as per our template, we want to find some $\overline{y}$ such that $(A)'\overline{y} \leq 0$ and $\sum_e \overline{y}_e > 0$. For this, we will do the following.

(1) Suppose there is some edge $e_i = (u_i, v_i)$ such that the dual constraints of u_i, v_i on y_i are not tight. Let S be the set of all such edges. In that case, we let $\overline{y}$ to be such that $\overline{y}_{(u,v)} = 1$ if $(u, v) \in S$, and 0 otherwise. Note that no edge in S can be incident to any vertex in T. It is clear that $\overline{y}$ satisfies $\sum_e \overline{y}_e > 0$ and that $(A)'\overline{y} \leq 0$ (infact, $(A)'\overline{y} = 0$).

(2) If there is no such edge, we terminate the algorithm.

Once we've computed S as in point (1) above, we will increase the value of some edge $e \in S$ until the constraint of some vertex becomes tight (i.e, we will choose an appropriate ϵ).

From here, it is not that hard to argue that this algorithm gives a 2-approximation to the optimum solution. (Because of time constraints, I couldn't finish this section)