5-8-1991

The Theorms of the Alternative

Kathryn L. Balsman '91
Illinois Wesleyan University

The Theorems of the Alternative

Kathryn L. Balsman

May 8, 1991

In layman's terms, a Theorem of the Alternative is a theorem which states that given two conditions, one of the two conditions is true. It further states that if one of those conditions fails to be true, then the other condition must be true. Mathematically, this does not give a very clear picture of this group of theorems, however. Katta Murty defines a typical Theorem of the Alternative as one which shows that corresponding to any given system of linear constraints, system I, there is another associated system of linear constraints, system II, based on the same data, satisfying the property that one of the systems among I and II is feasible if and only if the other is infeasible. These theorems are not extremely well-known although they do have direct applications in the derivation of optimality conditions. The task at hand was to explore the Theorems of the Alternative as they are found in linear programming, projection theory, and linear algebra.

Although not overly familiar as theorems, many mathematicians have found the Theorems of the Alternative to be fascinating. These mathematicians have explored the possibilities of deriving a group of these theorems from one or more of these theorems. Two mathematicians who have engaged in just such an exploration are W.C. Pye and P.G. Webster, both from the University of Southern Mississippi - Hattiesburg. They present a geometrically motivated result from which they claim, in conferences with Dr. Melvyn Jeter, that all of the Theorems of the Alternative can be easily derived from their main theorem simply by selecting the proper subspace. Their main theorem is as follows:

Theorem (Geometric form of Gordan's theorem) Let S be a subspace of $\mathbf{R}_{n\times 1}$. Then one and only one of the following is true:

I. S contains a positive vector.

II. $S^{\perp}$ contains a semipositive vector.

Pye has not, however, produced any proof of this claim. Several months were spent by the author of this paper in an effort to prove the validity of this claim. Virtually all of the different Theorems of the Alternative as stated in [3] were subjected to manipulations in this quest. The general consensus would appear to be that not all of the common Theorems of the Alternative, as found in references such as [3] or [4], can be derived as corollaries from Pye's Theorem of the Alternative. Marlow and Murty do suggest, however, that there are a few theorems which are fundamental and that the other theorems can be proven from them. This hypothesis seems to be much more valid than the one made by Pye and Webster.

As a result of the failure to derive all Theorems of the Alternative from Pye and Webster's main theorem, this paper focuses on derivations of various Theorems of the Alternative from different mathematical perspectives. The first chapter focuses on Farkas' Lemma as it is found in linear programming. The proof of Farkas' Lemma proves to be rather straightforward after first proving the Fundamental Duality Theorem which states that if either the primal or the dual problems has an optimal solution, then the other problem does as well. (The appendix provides further information on finding an optimal solution for a linear programming problem.) In fact, both optimal solutions are equal. Next, the focus shifts to Pye and Webster's approach to the Theorems of the Alternative which uses a projection theorem and is more geometrically motivated. Finally, Tucker's Theorem of the Alternative

is derived using results from linear algebra.

Glossary of Symbols

$A_{i\cdot}$ denotes the *ith* row of the matrix A.

$A_{\cdot j}$ denotes the *jth* column of the matrix A.

$0 \neq x \geq 0$ denotes a vector that is semipositive, i.e., all of the entries are nonnegative and at least one entry is not 0.

$x \geq 0$ denotes a vector that is nonnegative.

I denotes the identity matrix (dimensions are inferred from context).

The set $H = \{x : c^T x = \alpha\}$, where $c \in \mathbf{R}_{n\times 1}, \alpha \in \mathbf{R}$, is called a hyperplane. The sets $H^+ = \{x : c^T x \geq \alpha\}$ and $H^- = \{c^T x \leq \alpha\}$ are called closed upper and lower half spaces, respectively.

CHAPTER I

FARKAS' LEMMA

One area where the Theorems of the Alternative are found is linear programming. This branch of mathematics deals with minimization and maximization problems. The basic mathematical programming problem is written in the form:

$$\text{Minimize (or Maximize) } f(x_1, ..., x_n) \text{ subject to } (x_1, ..., x_n) \in F \qquad (1)$$

where F is a subset of the domain of f. If $F = \{(x_1, ..., x_n) : \text{ each } x_i \in \mathbf{R}\}$, Problem (1) is said to be unconstrained ($\mathbf{R}$ denotes the set of real numbers). Problem (1) is constrained when F is a proper subset of $\{(x_1, ...x_n) : x_i \in \mathbf{R}\}$. When Problem (1) is constrained, the set F is usually defined by a system of constraints. We call the function $f(x_1, ..., x_n)$ the cost or objective function. The set F is frequently referred to as the set of feasible solutions. In other words, x is a feasible solution of (1) if and only if $x \in F$. Problem (1) is a linear programming problem when $f(x_1, ..., x_n)$ is linear and when F is completely determined by a system of linear equalities and inequalities.

Consider the two problems:

$$\text{Maximize } f(x_1, ..., x_n) \text{ subject to } (x_1, ..., x_n) \in F$$

and

$$\text{Minimize } -f(x_1, ..., x_n) \text{ subject to } (x_1, ..., x_n) \in F.$$

We can see that

$$f(x_1^*, ...x_n^*) = \max\ \{f(x_1, ..., x_n) : (x_1, ..., x_n) \in F\}$$

if and only if

$$-f(x_1^*, ..., x_n^*) = \min\ \{-f(x_1, ..., x_n) \in F\}.$$

Thus it is obvious that

$$\max\{f(x_1, ..., x_n) : (x_1, ..., x_n) \in F\} = \text{-min}\ \{-f(x_1, ...x_n) : (x_1, ..., x_n) \in F\}$$

Hence, without loss of generality, any algorithm that can be used to minimize $f(x_1, ..., x_n)$ over F can also be used to maximize $f(x_1, ..., x_n)$ over F.

Consider the system of m linear equations in n unknowns

$$Ax = b,$$

where $A \in \mathbf{R}_{m \times n}$, $\text{rank}A = m, x \in \mathbf{R}_{n \times 1}$, and $b \in \mathbf{R}_{m \times 1}$.

A basic solution of $Ax = b$ occurs when $x_B = B^{-1}b, x_D = [0 \cdots 0]^T$ and $x = [x_B^T \; x_D^T]^T$ (where B is a nonsingular submatrix of A). The matrix B is called a basis matrix and the set of columns of B (which are the columns of A) is termed an admissible basis of the basic solution. Any basic solution, x, of $Ax = b$ that is also a feasible solution of the linear programming problem,

$$\text{Minimize } c^T x \text{ subject to } Ax = b \text{ and } x \geq 0$$

is termed a basic feasible solution, a BFS (see Appendix for further details on generating an optimal solution of the linear programming problem).

We will now look at a key idea in linear programming. We will show that each linear programming problem has a special correspondence to another "dual" linear programming problem. The relationship between these two problems, one of which is a minimization while the other is a maximization problem, is such that a solution of one problem yields a solution of the other. Also, the number of variables of one problem will equal the number of constraints of the other. This enables us to select which form we shall choose to solve when one problem is more economical to solve than the other.

The primal problem, often referred to as the unsymmetric primal problem is expressed as

$$\text{Minimize } c^T x \text{ subject to } Ax = b \text{ and } x \geq 0 \qquad (2)$$

(here A is $m \times n$). The dual problem is defined to be

$$\text{Maximize } b^T w \text{ subject to } A^T w \leq c \qquad (3).$$

The *set of feasible solutions* of problem (2) is defined and denoted by $F = \{x : Ax = b$ and $x \geq 0\}$. When F is not empty then (2) is said to be feasible. A similar definition holds for (3).

From the definitions of the primal and dual problems, several important observations can be made. The first proposition is sometimes referred to as the Weak Duality Theorem.

Proposition 1 (Weak Duality Theorem) *If x and w are feasible solutions of the primal and dual problems, respectively, then $b^T w \leq c^T x$.*

PROOF:

> Clearly $A^T w \leq c$ implies that $w^T A \leq c^T$. Since $x \geq 0$ and $Ax = b$, then $w^T b = w^T A x \leq c^T x$.

Hence, when both problems are feasible, the value of the objective function of the primal problem is bounded below by any value of the dual objective function. Similarly, the value of the objective function of the dual problem is bounded above by any value of the primal objective function. When the dual problem has an unbounded maximum, the primal problem fails to be feasible.

Proposition 2 *If $\bar{x}$ and $\bar{w}$ are feasible solutions of the primal and dual problems, respectively, such that $c^T\bar{x} = b^T\bar{w}$, then $\bar{x}$ and $\bar{w}$ solve their respective problems. Moreover, the optimal minimum value of the primal objective function is equal to the optimal maximum value of the dual function.*

PROOF:

Since $\bar{x}$ and $\bar{w}$ are feasible solutions of the primal and dual problems such that $c^T\bar{x} = b^T\bar{w}$, then for any other feasible solution, x, of the primal problem, it follows that $c^Tx \geq b^T\bar{w} = c^T\bar{x}$. This implies that $\bar{x}$ gives an optimal minimum solution to the primal problem. Likewise, $\bar{w}$ gives an optimal maximum solution.

The aforementioned results lead us to the fundamental result concerning dual linear programming problems.

Theorem 3 (Fundamental Duality Theorem) *If either the primal or the dual problem has a finite optimal solution, then the other problem has a finite optimal solution as well. Moreover, the optimal values of the two objective functions are equal, i.e.,*

$$\min\{c^Tx : Ax = b \textit{ and } x \geq 0\} = \max\{b^Tw : A^Tw \leq c\}$$

PROOF:

Without loss of generality, assume that the BFS $x = [x_B^T\ x_D^T]^T$ where $x_B = B^{-1}b$, $x_D = [0\cdots 0]^T$, and $B = [A^{(1)}\cdots A^{(m)}]$, solves (2) (See appendix). The final simplex tableau is then

$$\begin{bmatrix} I_m & B^{-1}D & 0 & B^{-1}b \\ 0 & c_B^TB^{-1}D - c_D^T & 1 & c_B^TB^{-1}b \end{bmatrix} \qquad (4)$$

where $c_B^TB^{-1}D - c_D^T \leq 0$ and $z_o = c_B^TB^{-1}b$. Recall that $\begin{bmatrix} I_m & B^{-1}D \end{bmatrix} = B^{-1}A$. Define $w \in \mathbf{R}_{m\times 1}$ by

$$w = (c_B^TB^{-1})^T$$

Then

$$A^Tw = A^T(c_B^TB^{-1})^T = (c_B^TB^{-1}A)^T = (c_B^T\begin{bmatrix} I_m & B^{-1}D \end{bmatrix})^T$$
$$= \begin{bmatrix} c_B^T & c_B^TB^{-1}D \end{bmatrix}^T \leq \begin{bmatrix} c_B^T & c_D^T \end{bmatrix}^T = c \qquad (5)$$

and w is a feasible solution of (3) (See Appendix). Also,

$$b^Tw = b^T(c_B^TB^{-1})^T = (c_B^TB^{-1}b)^T = c_B^Tx_B = c^Tx.$$

Thus Proposition 2 implies that w solves the dual problem We have now verified the theorem for the case when the primal problem has a finite optimal solution.

What remains to be proven is that the result holds whenever the dual problem has a finite optimal solution.

Suppose that w solves the dual problem. We need to rewrite the dual problem in a form analagous to that of the primal problem. Let $w = w_1 - w_2$, where $w_1 = [w_{11} \cdots w_{1m}]^T \geq 0$ and $w_2 = [w_{21} \cdots w_{2m}]^T \geq 0$. Next add slack variables $w_3 = [w_{31} \cdots w_{3m}]^T \geq 0$ to the constraints $A^T w \leq c$. Replace the dual problem with the following problem

$$\begin{aligned} &\text{-Minimize } -b^T w_1 + b^T w_2 + 0^T w_3 \text{ subject to} \\ &\quad A^T w_1 - A^T w_2 + I w_3 = c, \\ &\quad w_1 \geq 0, w_2 \geq 0, \text{ and } w_3 \geq 0 \end{aligned}$$

i.e.,

$$\text{-Minimize } -b^T w_1 + b^T w_2 + 0^T w_3 \text{ subject to}$$

$$\begin{bmatrix} -A^T & A^T & -I \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = -c,$$

$$w_1 \geq 0, w_2 \geq 0, \text{ and } w_3 \geq 0.$$

We know from the first part of our proof that the dual of this last problem has a finite optimal solution and the two objective functions for these problems are equal at their optimal values. But the dual of this last problem is

$$\text{-Maximize } -c^T x \text{ subject to}$$

$$[\ -A^T \quad A^T \quad -I\ ]^T x \leq \begin{bmatrix} -b \\ b \\ 0 \end{bmatrix}$$

which is equivalent to

$$\text{Minimize } c^T x \text{ subject to } -Ax \leq -b, Ax \leq b, \text{ and } -x \leq 0$$

or

$$\text{Minimize } c^T x \text{ subject to } Ax = b \text{ and } x \geq 0.$$

Now that we understand the duality of linear programming, we are able to examine the result that we have been aiming for, Farkas' lemma. This is frequently referred to as Farkas' Theorem. We will prove it using the Fundamental Duality Theorem.

Theorem 4 (Farkas' Lemma) *Let $A \in \mathbf{R}_{m\times n}$, and $b = [b_1 \cdots b_m]^T \in \mathbf{R}_{m\times 1}$. Then there exists an $x = [x_1 \cdots x_n]^T \in \mathbf{R}_{n\times 1}$, such that $Ax = b$, $x \geq 0$ if and only if $A^T w \geq 0$, where $w = [w_1 \cdots w_m]^T \in \mathbf{R}_{m\times 1}$, implies that $b^T w \geq 0$.*

PROOF:

Suppose that there exists an $x \geq 0$ such that $Ax = b$. Let $w \in \mathbf{R}_{m\times 1}$ such that $A^T w \geq 0$. Then we have

$$b^T w = (Ax)^T w = x^T (A^T w) \geq 0.$$

Hence, if there exists $x \geq 0$ such that $Ax = b$, then $A^T w \geq 0$ implies that $b^T w \geq 0$.

Next assume that $A^T w \geq 0$ always implies that $b^T w \geq 0$. Now consider the problem

$$\text{Minimize } 0^T x \text{ subject to } Ax = b, x \geq 0.$$

The dual is

$$\text{Maximize } b^T w \text{ subject to } A^T w \leq 0.$$

Whenever $A^T w \leq 0$, then $A^T(-w) \geq 0$. So, $b^T(-w) \geq 0$ and $b^T w \leq 0$. Thus, $\{b^T w : A^T w \leq 0\} \leq 0 = b^T 0$, where $0 \in \{w : A^T w \leq 0\}$. Hence the dual has the optimal solution of 0 which happens at 0 which belongs to the set of feasible solutions of the dual problem. Therefore, the fundamental duality theorem implies that there exists $x \in F = \{x : Ax = b, x \geq 0\} \ni 0 = 0^T x = min\{0^T x : Ax = b, x \geq 0\}$, i.e. there exists $x \geq 0$ such that $Ax = b$.

Recall that the form given above for Farkas' lemma is not standard for a Theorem of the Alternative. As was stated earlier, the most commonly found form of these theorems is as follows,

Given a matrix A, one and only one of Problems I and II has a solution.

The more common form of Farkas' lemma is as follows:

Given a matrix A, one and only one of the following hold:
I. There exists $x \geq 0$ such that $Ax = b$,
II. There exists w such that $A^T w \leq 0$ and $b^T w > 0$.

What remains to be shown now is that these two forms of Farkas' lemma are equivalent. In other words, we ask ourselves the following question:
Are the following equivalent?

(V_1) There exists $x \geq 0$ such that $Ax = b$ if and only if $A^T w \geq 0$ implies that $b^T w \geq 0$

(V_2) One and only one of the following hold:
(1) There exists $x \geq 0$ such that $Ax = b$,
(2) There exists w such that $A^T w \leq 0$ and $b^T w > 0$.

PROOF:

Assume that (V_1) holds. We want to establish (V_2). We must do two things to prove this: 1) show that at least one of (1) and (2) must hold; 2) show that both cannot hold.

First, suppose that (1) does not hold, i.e., suppose that there does not exist $x \geq 0$ such that $Ax = b$. Since (V_1) is given, $A^T w \geq 0$ does not imply that $b^T w \geq 0$, i.e., there exists w such that $A^T w \geq 0$ and $b^T w < 0$. But this means that $A^T(-w) \leq 0$ and $b^T(-w) > 0$. Hence, (2) is true. Therefore either (1) or (2) must hold.

Next we will show that (1) and (2) cannot simultaneously be true. Suppose that there exists $x \geq 0$ such that $Ax = b$ and that there exists $w \in \mathbf{R}_{m\times 1}$ such that $A^T w \leq 0$ and $b^T w > 0$. In other words,

$$A^T w \leq 0 \Rightarrow w^T A = (A^T w)^T \leq 0^T$$

and

$$0 \geq (w^T A)x = w^T(Ax) = w^T b = b^T w > 0$$

which is a contradiction. Therefore, (1) and (2) cannot hold simultaneously. We have thus proven that $(V_1) \Rightarrow (V_2)$.

Now assume that (V_2) holds and suppose that there exists $x \geq 0$ such that $Ax = b$. Then there does not exist w such that $A^T w \leq 0$ and $b^T w > 0$. Consider w such that $A^T w \geq 0$. Then $A^T(-w) \leq 0$. Hence $b^T(-w) \not> 0$. So $b^T(-w) \leq 0$ and $b^T w \geq 0$, i.e., when there exists $x \geq 0$ such that $Ax = b$, then $A^T w \geq 0$ implies that $b^T w \geq 0$.

Next suppose that $A^T w \geq 0$ implies that $b^T w \geq 0$,i.e., $A^T w \leq 0$ implies that $A^T(-w) \geq 0$ which implies that $b^T(-w) \geq 0$. This in fact implies that $b^T w \leq 0$ implies that (2) never holds. Therefore (1) holds, i.e., there exists $x \geq 0$ such that $Ax = b$; $(V_2) \Rightarrow (V_1)$.

Geometrically, Farkas' lemma has some very nice interpretations. Farkas' lemma implies that either b lies in the convex cone generated by the columns of A or there exists a hyperplane that separates b and the columns of A^T and this hyperplane can be denoted by

$$H = \{z : z^T w = 0\}$$

(see figure 1, parts I and II).

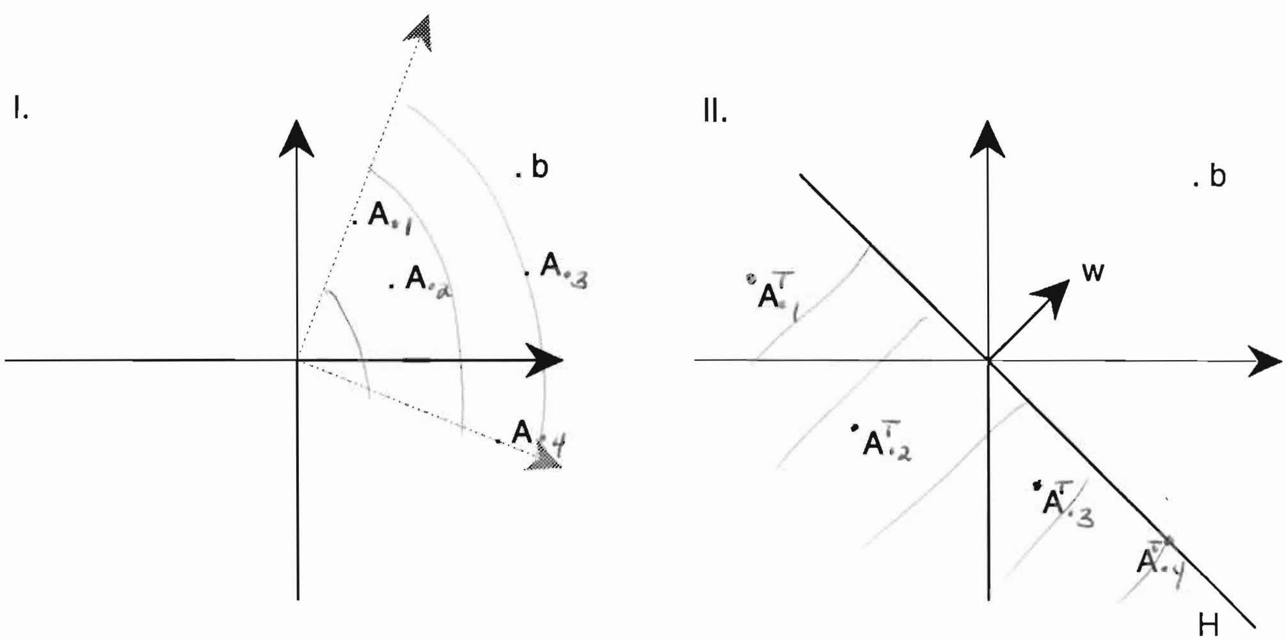

Figure 1:

CHAPTER II

PYE - GORDAN THEOREM OF THE ALTERNATIVE

Our next approach to the Theorems of the Alternative uses a projection theorem from convexity theory and is more geometrically motivated than the previous result. In their paper [6], Pye and Webster develop a Theorem of the Alternative from which they claim that the remainder of the Theorems of the Alternative can be easily generated. The Theorem that Pye and Webster generate is actually the geometric form of Gordan's Theorem of the Alternative as found in Marlow [3]. Before considering the Pye - Gordan Theorem of the Alternative, we need to examine the ideas leading up to the result.

Recall the vector space $\mathbf{R}_{n\times 1}$ consisting of all column vectors with n real components. Within the space, we can add any two vectors, and we can multiply vectors by scalars. For example, the vector space $\mathbf{R}_{2\times 1}$ represents the xy-plane. A nonempty subspace, S, of $\mathbf{R}_{n\times 1}$ is any nonempty set of vectors for which $x, y, \in S, \alpha, \beta \in \mathbf{R}$ imply that $\alpha x + \beta y \in S$. A linear combination of the vectors $x_1, ..., x_k \in \mathbf{R}_{n\times 1}$ is an expression of the form

$$\sum_{i=1}^{k} \alpha_i x_i = \alpha_1 x_1 + \cdots + \alpha_k x_k,$$

where $\alpha_1, ..., \alpha_k \in \mathbf{R}$. If $B \subseteq \mathbf{R}_{n\times 1}$, then we denote the set of all linear combinations of elements of B by $L(B)$, i.e.,

$$L(B) = \{\sum_{i=1}^{k} \alpha_i x_i : \text{ each } \alpha_i \in \mathbf{R}, x_i \in B \text{ and } k \in \mathbf{N}\}$$

Moreover, if a vector space V consists of all linear combinations of the particular vectors $x_1, ... x_k$, then these vectors *span* the space. Also, if L is a subspace in $\mathbf{R}_{n\times 1}$, then the *orthogonal complement* of L, denoted by $L^{\perp}$, is the set of vectors which are orthogonal to each point in L, i.e.,

$$L^{\perp} = \{y :< x, y >= 0 \text{ for each } x \in L\}.$$

where $< x, y >$ represents the *inner product* of x and y, i.e., $< x, y >= x^T y$. The *column space* of a matrix A consists of all linear combinations of the columns of A and is denoted by $R(A)$. The *null space* of a matrix A consists of all vectors x such that $Ax = 0$ and is denoted by$N(A)$. The column space of a matrix $A \subseteq \mathbf{R}_{m\times n}$ is a subspace of $\mathbf{R}_{m\times 1}$ while the null space is a subspace of $\mathbf{R}_{n\times 1}$.

The *norm* of x is

$$\|x\| = \sqrt{< x, x >} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

.

Proposition 5 *If L is a nonempty closed set in $\mathbf{R}_{n\times 1}$ and x is any point in $\mathbf{R}_{n\times 1}$, then the infimum, $inf\{\|x-l\| : l \in L\}$, is attained at some point $y \in L$.*

PROOF:

We begin by making the following definition: $d(x, L) := inf\{\|x-y\| : y \in L\}$. Now, we must consider two cases, $x \in L$ and $x \notin L$, in order to complete our proof.

Case 1: If $x \in L$, then $\|x-x\| = 0 = d(x, L)$.

Case 2: If $x \notin L$, let $B = d(x, L)$. For $n \in \mathbf{N}, B + \frac{1}{n} > B$. Hence, $B + \frac{1}{n}$ is not a lower bound for $\{\|x-y\| : y \in L\}$. Therefore, there exists a $y_n \in L$ such that $B + \frac{1}{n} > \|x - y_n\| \geq B$. Let a sequence $\{y_n\}$ be generated in that fashion and consider

$$B_1(x) := \{y \in L : \|y - x\| \leq B + 1\}.$$

Then $\{y_n\} \subseteq L \cap B_1(x)$. So $\{y_n\}$ is contained in a closed and bounded set. Therefore there exists $\{y_{nk}\}$ which is a subsequence of $\{y_n\}$ and y such that

$$\lim_{n_k \to \infty} y_{n_k} = y.$$

We know that

$$|\ \|y_{n_k} - x\| - \|y - x\|\ | \leq \|y_{n_k} - x - y + x\| = \|y_{n_k} - y\|.$$

It follows that as $n_k \to \infty, \|y_{n_k} - x\| \to \|y - x\|$. But we also have that as $n_k \to \infty, \|y_{n_k} - x\| \to B$. Hence,

$$d(x, L) = B = \lim_{n_k \to \infty} \|y_{nk} - x\| = \|y - x\|.$$

Therefore, there exists $y \in L$ such that $\|x - y\| = d(x, L)$ when $x \in \mathbf{R}_{n\times 1}$.

The point y is said to be $\|\ \|$-closest to x in L. This does not mean tht y is unique, there may be many such points y which are $\|\ \|$-closest to x. An example of this is when L is a circle in the plane with x as the center of L. In that case, every point of L is $\|\ \|$-closest to x. (See figure 2)

A set L is a *convex* set if and only if $l_1, l_2 \in L$ and $\alpha, \beta \in \mathbf{R}^+$, where $\mathbf{R}^+$ is the set of nonnegative real numbers, such that $\alpha + \beta = 1$ implies that $\alpha l_1 + \beta l_2 \in L$. Intuitively this means that a set L is convex if and only if any straight line connecting any two points of L is itself a subset of L. (See figure 3)

Theorem 6 *Let L be a convex set in $\mathbf{R}_{n\times 1}$ with $x \notin L$ and $y \in L$. Then the following statements are equivalent:*

(a) y is $\|\ \|$-closest to x in L.

(b) $l \in L \Rightarrow < x - y, l - y > \leq 0$.

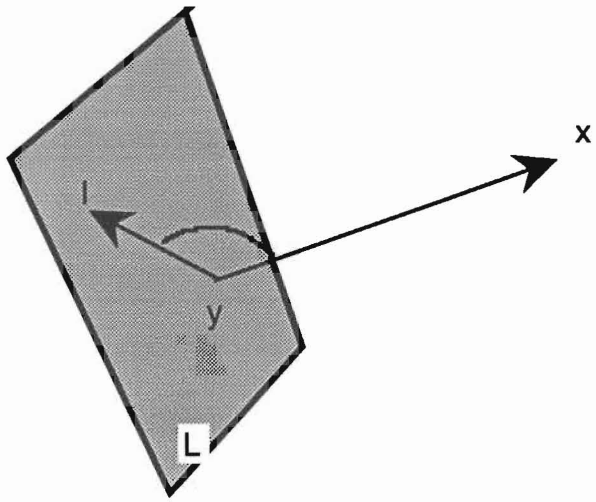

Figure 2: y is $\| \|$–closest to x in L

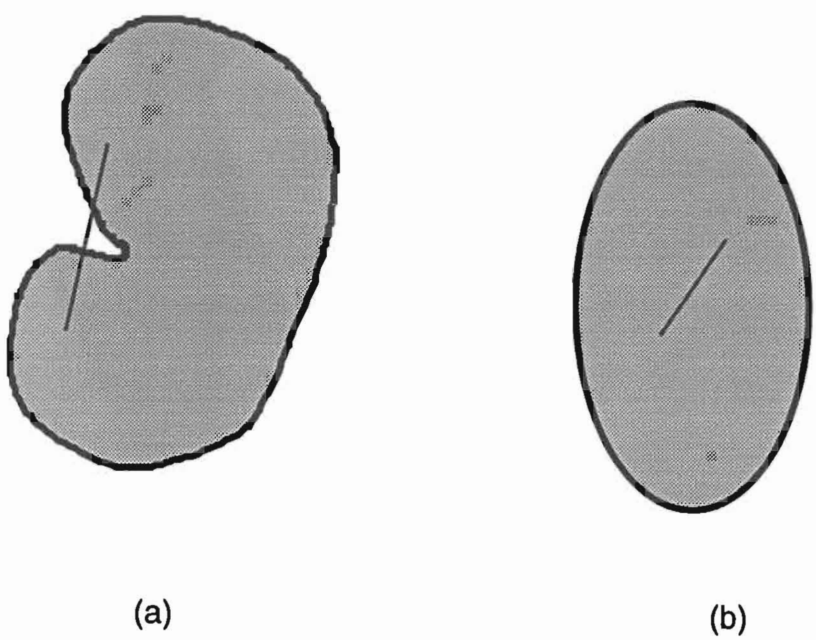

Figure 3: : (a) is not a convex set while Figure (b) is a convex set.

PROOF:

First we shall assume that (b) holds. Let $l \in L$. Then $< x - y, l - y > \leq 0$. We plan to show that

$$\|x - l\|^2 - \|x - y\|^2 = \|l - y\|^2 - 2 < x - y, l - y > . \qquad (1)$$

Since the RHS of (1) is greater than or equal to zero, that forces the LHS of (1) to be greater than of equal to zero as well. This tells us that $\|x - l\| \geq \|x - y\|$ which means that y is $\| \;\|$–closest to x in L. Recall that $\|x - l\|^2 =< x - l, x - l >$. So, we shall rewrite the above equation as

$$< x - l, x - l >=< x - y, x - y > + < l - y, l - y > -2 < x - y, l - y > .$$

We now want to show that the RHS of the equation does in fact equal the LHS. The linearity of the function allows us to do the following operations:

$$< x - y, x - y > + < l - y, l - y > -2 < x - y, l - y >$$

$$=< x - y, x - y > + < x - y, y - l > + < l - y, l - y > + < y - x, l - y >$$

$$=< x - y, x - y + y - l > + < l - y + y - x, l - y >$$

$$=< x - y, x - l > + < l - x, l - y >$$

$$=< x - y - l + y, x - l >$$

$$=< x - l, x - l > .$$

Hence, the RHS of (1) does in fact equal the LHS, so we can infer that (b) $\Rightarrow$ (a).

Next we will show that (a) $\Rightarrow$ (b). Since L is convex and $y \in L$, then $y + \alpha(l - y) = (1 - \alpha)y + \alpha l \in L$ whenever $l \in L$ and $\alpha \in [0, 1]$. We first will show that

$$\|x - [y + \alpha(l - y)]\|^2 - \|x - y\|^2 = \alpha[-2 < x - y, l - y > +\alpha\|l - y\|^2]. \qquad (2)$$

Since the LHS of (2) is greater than or equal to zero, the RHS of (2) is greater than or equal to zero as well. This tells us that $(\alpha/2)\|l - y\|^2 \geq< x - y, l - y >$. Taking the limit as α approaches zero forces $< x - y, l - y >\leq 0$.

Begin by adding $\|x - y\|^2$ to both sides of the equation, to get

$$< x - [y + \alpha(l - y)], x - [y + \alpha(l - y)] >$$

$$=< x - y, x - y > -2\alpha < x - y, l - y > +\alpha^2 < l - y, l - y > .$$

By the linearity of the inner product, we can rewrite the RHS of this equation as:

$$< x - y, x - y > -2\alpha < x - y, l - y > +\alpha^2 < l - y, l - y >$$

$$=< x-y, x-y > -\alpha < x-y, l-y > +\alpha^2 < l-y, l-y > -\alpha < x-y, l-y >$$
$$=< x-y, x-y-\alpha l+\alpha y > + < \alpha^2(l-y)-\alpha(x-y), l-y >$$
$$=< x-y, x-y-\alpha l+\alpha y > + < -\alpha l+\alpha y+x-y, \alpha y-\alpha l >$$
$$=< x-y-\alpha l+\alpha y, x-y-\alpha l+\alpha y >$$
$$=< x-[y+\alpha(l-y)], x-[y+\alpha(l-y)] > .$$

Thus the RHS does in fact equal the LHS. We can now infer that by choosing α sufficiently small but not 0, (a) $\Rightarrow$ (b).

We are now ready to state what Pye and Webster claim to be their main result.

Theorem 7 *Let S be a subspace of $\mathbf{R}_{n\times 1}$. If S does not contain a positive vector, then $S^{\perp}$ contains a semipositive vector.*

PROOF:

Assume S does not contain a positive vector. Furthermore, define C as

$$C = S - \mathbf{R}^{+}_{n\times 1} = \{s-p : s \in S \text{ and } p \in \mathbf{R}^{+}_{n\times 1}\}$$

Clearly C does not contain any positive vector so $C \neq \mathbf{R}_{n\times 1}$. C can be shown to be convex. In order to do so, let $\lambda \in [0,1], s_1 - p_1 \in S - \mathbf{R}^{+}_{n\times 1}$, and $s_2 - p_2 \in S - \mathbf{R}^{+}_{n\times 1}$. Then we have

$$\lambda(s_1 - p_1) + (1-\lambda)(s_2 - p_2)$$
$$= \lambda s_1 - \lambda p_1 + (1-\lambda)s_2 - (1-\lambda)p_2$$
$$= [\lambda s_1 + (1-\lambda)s_2] - [\lambda p_1 + (1-\lambda)p_2] \in S - \mathbf{R}^{+}_{n\times 1} = C.$$

C is also closed because it is an intersection of closed lower halfspaces.

Now choose any $x \notin C$. Then there is a point $y \in C$ which is $\|\ \|$-closest to x in C (by Theorem 6). This is equivalent to

$$\forall c \in C, < x-y, c-y > \leq 0.$$

The fact that $c+y \in C$ whenever $c \in C$ ensures that

$$\forall c \in C, < x-y, c > \leq 0. \qquad (5)$$

From (5) we infer that $0 \neq x-y \geq 0$ as $c \in C$ can be any vector in $\mathbf{R}^{-}_{n\times 1}$. From (5) we can also see that since $c \in C$ can be any vector in the subspace S,

$$< x-y, s > \leq 0 \text{ and } < x-y, -s > \leq 0, \ \forall s \in S$$
$$i.e., < x-y, s >= 0 \ \forall s \in S.$$

Hence, $x - y \in S^{\perp}$.

The next step is to state Theorem 7 in an equivalent form that is analogous to our standard form for Theorems of the Alternative.

Theorem 8 (Geometric form of Gordan's theorem) *Let S be a subspace of $\mathbf{R}_{n\times 1}$. Then one and only one of the following is true:*
I. S contains a positive vector.
II. $S^{\perp}$ contains a semipositive vector.

PROOF:

We need to show that I and II cannot hold simultaneously and that either I or II must be true. Let $x \in S$ and $y \in S^{\perp}$. Then $< x, y >= 0$. If $x > 0$ and $0 \neq y \geq 0$, then $< x, y > > 0$, which is a contradiction. Therefore I and II cannot hold simultaneously.

Now assume that S does not contain a positive vector, then by Theorem 7, $S^{\perp}$ must contain a semipositive vector. Hence if I does not hold, II must.

Recall that the four fundamental subspaces $R(A), N(A), R(A^T)$, and $N(A^T)$ satisfy

$$R(A)^{\perp} = N(A^T) \qquad R(A^T)^{\perp} = N(A)$$

$$R(A) = N(A^T)^{\perp} \qquad R(A^T) = N(A)^{\perp}.$$

This enables us to more clearly see how the Geometric form of Gordan's theorem can be used to quickly prove Stiemke's Theorem of the Alternative.

Theorem 9 (Stiemke's theorem) *Let A be a real matrix. Then one and only one of the following is true:*
I. There exists an $x > 0$ such that $Ax = 0$.
II. There exists a y such that $0 \neq A^T y \geq 0$.

PROOF:

From Theorem 8 it follows that exactly one of the following must be true:
I. $N(A)$ has a positive vector.
II. $N(A)^{\perp} = R(A^T)$ has a semipositive vector.

Pye and Webster further assert that from their main theorem, all of the other Theorems of the Alternative can be easily derived in this fashion, i.e., by properly selecting the subspace. This author has devoted much effort towards this end but has failed to show this to be true. However, this author would never attempt to state that it is impossible to do so. Perhaps the subspaces which are needed simply have not been discovered as of yet. Other mathematicians claim that these theorems can only be derived if a few of them are first proved as fundamental theorems. Once these fundamental theorems are proved, all of the remaining theorems will easily follow as corollaries.

CHAPTER III

TUCKER'S THEOREM OF THE ALTERNATIVE

The standard form of Tucker's Theorem of the Alternative is as follows:

Theorem 10 *Let A be a real matrix. Then one and only one of the following is true:*
(1) There exists $x > 0$ such that $Ax \leq 0$,
(2) There exists $y \geq 0$ such that $0 \neq A^T y \geq 0$.

Rather than prove Tucker's Theorem of the Alternative directly from this standard form, we shall take the approach found in [4]. This approach uses a variant form of Tucker's Theorem of the Alternative after first proving another lemma by Tucker. First, we will begin with the following Theorem of the Alternative.

Theorem 11 *Let $A = [a_{ij}] \in \mathbf{R}_{m\times n}, b \in \mathbf{R}_{m\times 1}$. Let $x = [x_1 \cdots x_n]^T$ and $\pi = [\pi_1 \cdots \pi_m]^T$. Then one and only one of the following systems has a solution.*
I. $Ax = b$
II. $\pi^T A = 0, \pi^T b = 1$

PROOF:

If I has a solution $\bar{x}$ and II has a solution $\bar{\pi}$, then $A\bar{x} = b$. Thus, $\bar{\pi}^T A\bar{x} = \bar{\pi}^T b$, but $(\bar{\pi}^T A)\bar{x} = 0, \bar{\pi}^T b = 1$. Hence it is impossible for both I and II to have solutions.

Put I into partitioned matrix form with $I \in \mathbf{R}_{m\times m}$ on the LHS of the tableau.

$$\left[\begin{array}{c|c|c} I & A & b \end{array}\right]$$

Perform Gauss-Jordan pivoting to get A in row echelon normal form, pivoting in rows 1 to m respectively. Consider the case when the ith row is the pivot row. Let the entries in this row be

$$\left[\begin{array}{c|c|c} d_{i1} \cdots d_{im} & \bar{a}_{i1} \cdots \bar{a}_{in} & \bar{b}_i \end{array}\right].$$

Let $D_{i\cdot} = [d_{i1} \cdots d_{im}]$. Then $[\bar{a}_{i1} \cdots \bar{a}_{in}] = D_{i\cdot}A$ and $\bar{b}_i = D_{i\cdot}b$. If $[\bar{a}_{i1} \cdots \bar{a}_{in}] = 0$ and $\bar{b}_i = 0$, this row at this stage represents a redundant constraint, simply move it to the bottom of the tableau and continue. Alternatively, if $[\bar{a}_{i1} \cdots \bar{a}_{in}] = 0$ and $\bar{b}_i \neq 0$, define

$$\bar{\pi}^T = \frac{D_{i\cdot}}{\bar{b}_i}.$$

Then we have $\bar{\pi}^T A = 0$ and $\bar{\pi}^T b = 1$, so $\bar{\pi}$ is a feasible solution of II and hence I has no feasible solution. If $[\bar{a}_{i1} \cdots \bar{a}_{in}] \neq 0$, select the first j such that $\bar{a}_{ij} \neq 0$, and perform a Gauss-Jordan pivot on $\bar{a}_{ij}$. Thus x_j becomes a basic variable. If the conclusion that I is infeasible is never made, then assigning the basic variables the corresponding values in the top right hand column and assigning all nonbasic variables the value zero gives a solution to I. Hence II has no solution.

As with Farkas' lemma, Tucker's Theorem of the Alternative can be proven by using the Fundamental Duality Theorem found in linear programming. In order to broaden our horizons, however, and to discover other areas of mathematics where these Theorems of the Alternative are found, we shall prove Tucker's Theorem of the Alternatve via Tucker's lemma. The following proof of Tucker's lemma involves a very interesting induction proof on the number of rows of the matrix A.

Lemma 12 (Tucker's lemma) *Given $A \in \mathbf{R}_{m\times n}$, there exists $x = [x_1 \cdots x_n]^T \in \mathbf{R}_{n\times 1}$, $\pi = [\pi_1 \cdots \pi_m]^T \in \mathbf{R}_{m\times 1}$ such that*

$$Ax \geq 0 \qquad (1)$$

$$\pi^T A = 0, \pi \geq 0 \qquad (2)$$

$$\pi + Ax > 0. \qquad (3)$$

PROOF:

We will first prove that there exists feasible solutions $x = [x_1 \cdots x_n]^T \in \mathbf{R}_{n\times 1}, \pi = [\pi_1 \cdots \pi_m]^T \in \mathbf{R}_{m\times 1}$ to (1) and(2) respectively satisfying

$$\pi_1 + A_1.x > 0. \qquad (4)$$

The proof is by induction on the number of rows in A.

Step 1: $m = 1$, i.e., $A \in \mathbf{R}_{1\times n}$

In this case, we let $\pi = \pi_1 = 1$ and $x = 0$ if $A_1. = 0$. Likewise, let $\pi = 0$, and $x = [A_1.]^T$ if $A_1. \neq 0$. Consider the instance when $A_1. = 0$, clearly (1) and (2) are satisfied. We see that (4) is also satisfied because $\pi_1 + A_1.x = 1 + 0x = 1 > 0$. Now consider the case when $A_1. \neq 0$. Again (1) and (2) are satisfied as is (4) because $\pi_1 + A_1.x = 0 + A_1.(A_1.)^T > 0$. So the theorem holds when $m = 1$.

Step 2:

INDUCTION HYPOTHESIS: If $D \in \mathbf{R}_{m-1\times n}$, there exist vectors $x = [x_1 \cdots x_n]^T \in \mathbf{R}_{n\times 1}, u = [u_1 \cdots u_{m-1}]^T \in \mathbf{R}_{m-1\times 1}$ satisfying:

$$Dx \geq 0, u^T D = 0, u \geq 0, u_1 + D_1.x > 0.$$

We now want to show that this result holds for $A \in \mathbf{R}_{m \times n}$.
Let $\bar{A} \in \mathbf{R}_{m-1 \times n}$ be obtained by deleting the last row, $A_{m\cdot}$, from A. Applying the induction hypothesis on $\bar{A}$, we know that there exists $x' = [x'_1 \cdots x'_n]^T \in \mathbf{R}_{n \times 1}, u' = [u'_1 \cdots u'_{m-1}]^T \in \mathbf{R}_{(m-1) \times 1}$ satisfying

$$\bar{A}x' \geq 0, u'^T \bar{A} = 0, u' \geq 0, u'_1 + A_{1\cdot}x' > 0 \qquad (5)$$

We now have two subcases to consider: either $A_{m\cdot}x' \geq 0$, or $A_{m\cdot}x' < 0$. Begin by considering the case when $A_{m\cdot}x' \geq 0$. Define $x = x', \pi = \begin{bmatrix} u'^T & 0 \end{bmatrix}^T$. Obviously,

$$Ax = \begin{bmatrix} \bar{A} \\ A_{m\cdot} \end{bmatrix} x = \begin{bmatrix} \bar{A}x \\ A_{m\cdot}x \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \geq 0.$$

Furthermore, π is feasible to (2), easily seen by

$$\pi^T A = \begin{bmatrix} u'^T & 0 \end{bmatrix} \begin{bmatrix} \bar{A} \\ A_{m\cdot} \end{bmatrix} = u'^T \bar{A} + 0 \cdot A_{m\cdot} = u'^T \bar{A} = 0.$$

Also, $\pi = \begin{bmatrix} u'^T & 0 \end{bmatrix}^T \geq 0$ because $u'^T \geq 0$ from (5). Finally, we must show that $\pi_1 + A_{1\cdot}x' > 0$. This is clear from (5).

Now consider the case where $A_{m\cdot}x' < 0$. First, for all $i = 1$ to m, define

$$\lambda_i := \frac{-A_{i\cdot}x'}{A_{m\cdot}x'}.$$

Then define $B \in \mathbf{R}_{(m-1) \times n}$ row wise by

$$B_{i\cdot} := A_{i\cdot} + \lambda_i A_{m\cdot}.$$

Applying the induction hypothesis to B there exists $x'' = [x''_1 \cdots x''_n]^T \in \mathbf{R}_{n \times 1}$, $u'' = [u''_1 \cdots u''_{m-1}]^T \in \mathbf{R}_{(m-1) \times 1}$ such that

$$Bx'' \geq 0, u''^T B = 0, u'' \geq 0, u''_1 + B_{1\cdot}x'' > 0.$$

Define

$$\alpha := \frac{-A_{m\cdot}x''}{A_{m\cdot}x'}.$$

Finally define

$$x := x'' + \alpha x' \in \mathbf{R}_{n \times 1} \text{ and}$$

$$\pi = \begin{bmatrix} u''^T & \sum_{i=1}^{m-1} \left(\frac{-A_{i\cdot}x'}{A_{m\cdot}x'} \right) u''_i \end{bmatrix}^T \in \mathbf{R}_{m \times 1}.$$

we shall now verify that x and π satisfy (1), (2), and (3). First notice that

$$A_{m\cdot}(x'' + \alpha x') = A_{m\cdot}x'' + \alpha A_{m\cdot}x' = A_{m\cdot}x'' - \frac{A_{m\cdot}x''}{A_{m\cdot}x'} A_{m\cdot}x' = 0.$$

Next notice that

$$\bar{A}_{i\cdot}(x'' + \alpha x') = \bar{A}x'' - \frac{A_{m\cdot}x''}{A_{m\cdot}x'}A_{i\cdot}x' = A_{i\cdot}x'' - \frac{A_{i\cdot}x'}{A_{m\cdot}x'}A_{m\cdot}x''$$

$$= A_{i\cdot}x'' + \lambda_i A_{m\cdot}x'' = (A_{i\cdot} + \lambda_i A_{m\cdot})x'' = B_{i\cdot}x'' \geq 0.$$

Thus, $\bar{A}(x'' + \alpha x') \geq 0$. Hence,

$$Ax = \begin{bmatrix} \bar{A} \\ A_{m\cdot} \end{bmatrix} [x'' + \alpha x'] = \begin{bmatrix} \bar{A}(x'' + \alpha x') \\ A_{m\cdot}(x'' + \alpha x') \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Now consider the following,

$$\pi^T A = \begin{bmatrix} u_1'' \cdots u_{m-1}'' & \sum_{i=1}^{m-1} \left(\frac{-A_{i\cdot}x'}{A_{m\cdot}x'}\right) u_i'' \end{bmatrix} \begin{bmatrix} A_{1\cdot} \\ \vdots \\ A_{m-1\cdot} \\ A_{m\cdot} \end{bmatrix}$$

$$= \sum_{i=1}^{m-1} u_i'' A_{i\cdot} + \sum_{i=1}^{m-1} \frac{-A_{i\cdot}x'}{A_{m\cdot}x'} u_i'' A_{m\cdot}$$

$$= \sum_{i=1}^{m-1} u_i'' \left(A_{i\cdot} - \frac{A_{i\cdot}x'}{A_{m\cdot}x'} A_{m\cdot}\right)$$

$$= \sum_{i=1}^{m-1} u_i'' (A_{i\cdot} + \lambda_i A_{m\cdot})$$

$$= \sum_{i=1}^{m-1} u_i'' B_{i\cdot}$$

$$= u''^T B = 0.$$

Clearly $\pi \geq 0$. Thus (2) is satisfied.

Finally, notice that

$$u_1'' + A_{1\cdot}(x'' + \alpha x')$$

$$= u_1'' + A_{1\cdot}x'' + A_{1\cdot}\frac{-A_{m\cdot}x''}{A_{m\cdot}x'}x'$$

$$= u_1'' + A_{1\cdot}x'' - \frac{A_{m\cdot}x''}{A_{m\cdot}x'}A_{1\cdot}x'$$

$$= u_1'' + A_{1\cdot}x'' - \frac{A_{1\cdot}x'}{A_{m\cdot}x'}A_{m\cdot}x''$$

$$= u_1'' + A_{1\cdot}x'' + \lambda_1 A_{m\cdot}x''$$

$$= u_1'' + (A_{1\cdot} + \lambda_1 A_{m\cdot})x''$$

$$= u_1'' + B_1.x'' > 0.$$

Hence, (3) is satisfied.

Therefore, when $A \in \mathbf{R}_{m\times n}$, there exists $x \in \mathbf{R}_{n\times 1}$ and $\pi \in \mathbf{R}_{m\times 1}$ such that

$$Ax \geq 0, \pi^T A = 0, \pi \geq 0, \text{ and } \pi_1 + A_1.x > 0.$$

It is now clear that for any $i = 1$ to m, we can use the above argument to show that there exist feasible solutions $x^i, \pi^i = [\pi_1^i \cdots \pi_m^i]^T$ to (1) and (2) respectively satisfying

$$\pi_i^i + A_i.x^i > 0.$$

Now all that remains is to define x and π such that there exist feasible solutions to (1), (2), and (3). Define

$$\bar{x} := \sum_{i=1}^{m} x^i$$

$$\bar{\pi} := \sum_{i=1}^{m} \pi^i.$$

Clearly, $\bar{\pi} \geq 0$. Note that from the above argument, each $Ax^i \geq 0$. Thus we have

$$A\bar{x} = A\left(\sum_{i=1}^{m} x^i\right) = Ax^1 + Ax^2 + \cdots + Ax^m \geq 0.$$

Likewise, since each ${\pi^i}^T A = 0$, we have

$$\bar{\pi}^T A = \left(\sum_{i=1}^{m} \pi^i\right)^T A = {\pi^1}^T A + {\pi^2}^T A + \cdots + {\pi^m}^T A = 0.$$

Also, we can see that (3) also holds,

$$\bar{\pi} + A\bar{x} = \sum_{i=1}^{m} \pi^i + A\left(\sum_{i=1}^{m} x^i\right)$$

$$= \pi^1 + \pi^2 + \cdots + \pi^m + Ax^1 + Ax^2 + \cdots + Ax^m$$

$$= \left(\pi^1 + Ax^1\right) + \left(\pi^2 + Ax^2\right) + \cdots + \left(\pi^m + Ax^m\right) > 0.$$

since each column, $\pi^j + Ax^j$ is nonnegative ($Ax \geq 0$ and $\pi \geq 0$) and the jth entry in the jth column is positive.

An immediate corollary to Tucker's lemma is the following:

Corollary 13 *Let* $A \in \mathbf{R}_{m_1 \times n}, D \in \mathbf{R}_{m_2 \times n}$ *respectively with* $n \geq 1$. *Then there exists* $x = [x_1 \cdots x_n]^T \in \mathbf{R}_{n \times 1}, \pi = [\pi_1 \cdots \pi_{m_1}]^T \in \mathbf{R}_{m_1 \times 1}, \mu = [\mu_1 \cdots \mu_{m_2}]^T \in \mathbf{R}_{m_2 \times 1}$ *satisfying*

$$Ax \geq 0, Dx = 0 \tag{6}$$

$$\pi^T A + \mu^T D = 0, \pi \geq 0 \tag{7}$$

$$\pi + Ax > 0. \tag{8}$$

PROOF:

Let $A \in \mathbf{R}_{m_1 \times n}, D \in \mathbf{R}_{m_2 \times n}$. Then $\begin{bmatrix} A^T & D^T & -D^T \end{bmatrix}^T \in \mathbf{R}_{(m_1+2m_2) \times n}$. Tucker's lemma implies that there exists $x \in \mathbf{R}_{n \times 1}$, and $\begin{bmatrix} \pi^T & \gamma^T & \nu^T \end{bmatrix}^T \in \mathbf{R}_{(m_1+2m_2) \times 1}$, so that

$$\begin{bmatrix} A \\ D \\ -D \end{bmatrix} x \geq 0,$$

$$\begin{bmatrix} \pi^T & \gamma^T & \nu^T \end{bmatrix} \begin{bmatrix} A \\ D \\ -D \end{bmatrix} = 0, \begin{bmatrix} \pi^T & \gamma^T & \nu^T \end{bmatrix}^T \geq 0,$$

$$\begin{bmatrix} \pi \\ \gamma \\ \nu \end{bmatrix} + \begin{bmatrix} A \\ D \\ -D \end{bmatrix} x > 0.$$

But,

$$\begin{bmatrix} A \\ D \\ -D \end{bmatrix} x \geq 0 \Leftrightarrow \begin{matrix} Ax \geq 0 \\ Dx \geq 0 \\ -Dx \geq 0 \end{matrix} .$$

which means that $Ax \geq 0$, and $Dx = 0$. So (6) is satisfied. Also,

$$0 = \begin{bmatrix} \pi^T & \gamma^T & \nu^T \end{bmatrix} \begin{bmatrix} A \\ D \\ -D \end{bmatrix} \Leftrightarrow \pi^T A + \gamma^T D - \nu^T D = 0 \Leftrightarrow \pi^T A + (\gamma^T - \nu^T) D = 0.$$

Let $\mu^T := \gamma^T - \nu^T \in \mathbf{R}_{m_2 \times 1}$. Then $\pi^T A + \mu^T D = 0$. Also, $\begin{bmatrix} \pi^T & \gamma^T & \nu^T \end{bmatrix} \geq 0$ if and only if $\pi^T \geq 0, \gamma^T \geq 0, \nu^T \geq 0$. Therefore $\pi^T \geq 0$ and (7) is satisfied. Finally,

$$0 < \begin{bmatrix} \pi \\ \gamma \\ \nu \end{bmatrix} + \begin{bmatrix} A \\ D \\ -D \end{bmatrix} x = \begin{bmatrix} \pi + Ax \\ \gamma + Dx \\ \nu - Dx \end{bmatrix}.$$

Therefore, $\pi + Ax > 0$ and (8) is satisfied. Therefore, Corollary I holds.

The complexity of the induction proof used in proving Tucker's lemma may cause fears of a long and complex proof of Tucker's Theorem of the Alternative. This is not the case,

Tucker's lemma and the corresponding corollary enable the proof of Tucker's Theorem of the Alternative to be done very simply and directly.

Theorem 14 (Tucker's Theorem of the Alternative) *Let* $m \geq 1$, $A \in \mathbf{R}_{m \times n}$, $B \in \mathbf{R}_{m_1 \times n}$, *and* $C \in \mathbf{R}_{m_2 \times n}$, *respectively. Let* $x = [x_1 \cdots x_n]^T \in \mathbf{R}_{n \times 1}$,
$\pi = [\pi_1 \cdots \pi_m]^T \in \mathbf{R}_{m \times 1}, \mu = [\mu_1 \cdots \mu_{m_1}]^T \in \mathbf{R}_{m_1 \times 1}, \gamma = [\gamma_1 \cdots \gamma_{m_2}]^T \in \mathbf{R}_{m_2 \times 1}$.
Then one and only one of the following is feasible.

I. There exists x *such that* $0 \neq Ax \geq 0, Bx \geq 0, Cx = 0$.

II. There exists π, μ, *and* γ *such that* $\pi^T A + \mu^T B + \gamma^T C = 0, \pi > 0, \mu \geq \mathbf{0}$.

PROOF:

We must first show that both I and II cannot hold simultaneously. To do this, assume that I and II are both feasible. Then consider the following expression

$$\begin{bmatrix} \pi^T & \mu^T & \gamma^T \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} x.$$

When we begin by postmultiplying $\begin{bmatrix} A^T & B^T & C^T \end{bmatrix}^T$ by x, we have

$$\begin{bmatrix} \pi^T & \mu^T & \gamma^T \end{bmatrix} \begin{bmatrix} Ax \\ Bx \\ Cx \end{bmatrix} = \pi^T(Ax) + \mu^T(Bx) + \gamma^T(Cx).$$

This yields a positive real number because $\pi^T(Ax)$ is a positive real number, $\mu^T(Bx)$ is nonnegative, and $\gamma^T(Cx)$ is 0.

If, however, we begin by premultiplying $\begin{bmatrix} A^T & B^T & C^T \end{bmatrix}^T$ by $\begin{bmatrix} \pi^T & \mu^T & \gamma^T \end{bmatrix}$, we have

$$\left[\pi^T A + \mu^T B + \gamma^T C\right] x = 0 \cdot x = 0.$$

Thus we have a contradiction and have shown that I and II cannot hold simultaneously.

Now suppose that I is infeasible. This means that every solution of

$$\begin{array}{l} Ax \geq 0 \\ Bx \geq 0 \\ Cx = 0 \end{array} \quad \text{or} \quad \begin{array}{c} \begin{bmatrix} A \\ B \end{bmatrix} x \geq 0 \\ Cx = 0 \end{array} \qquad (7)$$

must satisfy $Ax = 0$. By Corollary 13, there exists $\bar{x}$ feasible to (7) and $\bar{\pi}, \bar{\mu}, \bar{\gamma}$ feasible to

$$\bar{\pi}^T A + \bar{\mu}^T B + \bar{\gamma}^T C = 0 = \begin{bmatrix} \bar{\pi}^T & \bar{\mu}^T \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} + \bar{\gamma}^T C = 0$$

$$\bar{\pi} \geq 0 \text{ and } \bar{\mu} \geq 0 \qquad (8)$$

satisfying $\bar{\pi} + A\bar{x} > 0$. But since $\bar{x}$ is feasible to (7), $A\bar{x} = 0$ as discussed above; so $\bar{\pi} > 0$. Hence $\begin{bmatrix} \bar{\pi}^T & \bar{\mu}^T & \bar{\gamma}^T \end{bmatrix}^T$ satisfies II. Thus if I is infeasible, II is feasible. Therefore one and only one of the two systems I and II is feasible.

As stated earlier, this is not the standard form of Tucker's Theorem of the Alternative. We are now prepared to prove the standard form of Tucker's Theorem of the Alternative as a result of the version that we have just proven.

Theorem10 (Tucker's Theorem of the Alternative) Let A be a real matrix. Then one and only one of the following is true:

(1) There exists $x > 0$ such that $Ax \leq 0$,
(2) There exists $y \geq 0$ such that $0 \neq A^T y \geq 0$.

PROOF:

We shall apply Theorem 14 to the matrices A^T, I, and 0. Then one and only one of the following hold:

(I) There exists an x such that $0 \neq A^T x \geq 0, Ix \geq 0$, and $0x = 0$.

(II) There exists π^T, μ^T, and ν^T such that $\pi^T A^T + \mu^T I + \nu^T 0 = 0, \pi > 0$, and $\mu \geq 0$.

Conditions (1) and (2) then follow readily with a minor change of variables.

APPENDIX

OBTAINING AN OPTIMAL SOLUTION TO A LINEAR PROGRAMMING PROBLEM

Consider a linear programming problem of the form

Minimize $c^T x$ subject to $Ax = b, x \geq 0$ (1)

where $A \in \mathbf{R}_{m \times n}, x \in \mathbf{R}_{n \times 1}$. Without loss of generality we shall assume that $\text{rank}A = m$ (hence, $m \leq n$).

Define z_0 by $z_0 = c^T x$. then Problem (1) is equivalent to finding a solution $x \in \mathbf{R}_{n \times 1}$ and $z_0 \in \mathbf{R}$ of

$$\begin{aligned} Ax &= b \\ -c^T x + z_0 &= 0 \end{aligned} \qquad (2)$$

for which $x \geq 0$ and z_0 is minimal. These constraints (2) can be represented by the matrix equation

$$\begin{bmatrix} A & 0 \\ -c^T & 1 \end{bmatrix} \begin{bmatrix} x \\ z_0 \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \qquad (3)$$

or by the augmented matrix

$$\begin{bmatrix} A & 0 & b \\ -c^T & 1 & 0 \end{bmatrix} \qquad (4)$$

A contains m linearly independent columns because $\text{rank}A = m$. We shall assume that it is the first m columns of A that are linearly independent, denote them by $B = \left[A^{(1)} \cdots A^{(m)}\right]$. Denote the last $n - m$ columns of A by $D = \left[A^{(m+1)} \dots A^{(n)}\right]$. Then $A = [B \mid D]$. Let $x_B = [x_1 \cdots x_m]^T, x_D = [x_{m+1} \cdots x_n]^T, c_B = [c_1 \cdots c_m]^T$, and $c_D = [c_{m+1} \cdots c_n]^T$. Then $x = [x_B^T \, x_D^T]^T$ and $c = [c_B^T \, c_D^T]^T$. In addition, (3) becomes

$$\begin{bmatrix} B & D & 0 \\ -c_B^T & -c_D^T & 1 \end{bmatrix} \begin{bmatrix} x_B \\ x_D \\ z_0 \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \qquad (5)$$

and (4) becomes

$$\begin{bmatrix} B & D & 0 & b \\ -c_B^T & -c_D^T & 1 & 0 \end{bmatrix} \qquad (6)$$

We will call matrix (6) the initial matrix for Problem (1).

Multiplying (6) on the left by

$$\begin{bmatrix} B & 0 \\ -c_B^T & 1 \end{bmatrix}^{-1} = \begin{bmatrix} B^{-1} & 0 \\ c_B^T B^{-1} & 1 \end{bmatrix}$$

yields

$$\begin{bmatrix} I_m & B^{-1}D & 0 & B^{-1}b \\ 0 & c_B^T B^{-1} D - c_D^T & 1 & c_B^T B^{-1} b \end{bmatrix} \qquad (7)$$

This matrix, matrix (7) can also be obtained from matrix (6) by performing a modified Gauss-Jordan prodcedure on the rows, i.e., by performing a sequence of elementary row operation as indicated below

$$\begin{bmatrix} B & D & 0 & b \\ -c_B^T & -c_D^T & 1 & 0 \end{bmatrix} \rightarrow \cdots \rightarrow \begin{bmatrix} I_m & B^{-1}D & 0 & B^{-1}b \\ -c_B^T & -c_D^T & 1 & 0 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} I_m & B^{-1}D & 0 & B^{-1}b \\ 0 & c_B^T B^{-1} D - c_D^T & 1 & c_B^T B^{-1} b \end{bmatrix}$$

where the final operation consists of replacing row $m+1$ by itself plus each component of c_B^T times the corresponding row of $\begin{bmatrix} I_m & B^{-1}D & 0 & B^{-1}b \end{bmatrix}$.

Since $z_0 = c_B^T B^{-1} b - (c_B^T B^{-1} D - c_D^T) x_D$, it follows that $x_B = B^{-1}b$ and $x_D = [0 \cdots 0]^T$ is an optimal solution of (1) whenever $B^{-1}b \geq 0$ and $c_B^T B^{-1} D - c_D^T \leq 0$. This establishes the following theorem.

Theorem 15 *If $B^{-1}b \geq 0$ and $c_B^T B^{-1} D - c_D^T \leq 0$, then $x = [x_B^T \; x_D^T]^T$, where $x_B = B^{-1}b$ and $x_D = [0 \cdots 0]^T$ solves (1) and in fact, the minimum value of the objective function is $z_0 = c_B^T B^{-1} b$.*

A basic solution of $Ax = b$ occurs when $x_B = B^{-1}b, x_D = [0 \cdots 0]^T$ and $x = [x_B^T \; x_D^T]^T$ (B nonsingular). The matrix B is called a basis matrix and the set of columns $\{A^{(1)} \cdots A^{(m)}\}$ is termed an admissible basis of the basic solution. A basic feasible solution, a BFS, is any basic solution, x, of $Ax = b$ that is also a feasible solution of (1), i.e., $x \geq 0$. Here, $F = \{x : Ax = b \text{ and } x \geq 0\}$ is the set of feasible solutions of (1). From the preceding theorem, we see that a sufficient condition for a BFS to be an optimal solution of (1), is that $c_B^T B^{-1} D - c_D^T <= 0, x_B = B^{-1}b$ and $x_D = [0 \cdots 0]^T$.

REFERENCES

1. Bazarra, M.S. and Shetty, C.M. (1976). *Lecture Notes in Economics and Mathematical systems.*Berlin, Germany: Springer - Verlag.

2. Jeter, M. (1986). *Mathematical Programming: An Introduction to Optimization.* New York: Marcel Dekker, Inc.

3. Marlow, W.H. (1978). *Mathematics for Operations Research.*New York: Wiley and Sons.

4. Murty, K.G. (1988). *Linear Complementarity, Linear and Nonlinear Programming.* Berlin, Germany: Heldermann Verlag.

5. Nikaido, H. (1968). *Convex Structures and Economic Theory.*New York: Academic Press.

6. Pye, W.C. and Webster, P.G. (1989). Theorems of the Alternative. *Mathematics and Computer Education. 35 (2),*113 - 116.