
Fundamental Theorem of Linear Inequalities

Yohan Song

Abstract

The purpose of this paper is to explain the fundamental theorem of linear inequalities by borrowing Schrijver's approach. We first describe how it is analogous to a famous result in linear algebra. Then show how the Schrijver's approach, which we will call *Schrijver's algorithm* shortly, through examples. The focus is then shifted to the simplex method, and in particular, how its variant Bland's rule is analogous to Schrijver's algorithm.

1. Introduction

1.1. Notations

We briefly go over the notations that we are going to use throughout.

- We write \mathbb{N} to denote the *set of positive integers*.
- We write \mathbb{F} to denote \mathbb{Q} , the *field of rational numbers*, or \mathbb{R} , the *field of real numbers*. That is, any statement stated with \mathbb{F} holds for both \mathbb{Q} and \mathbb{R} .
- We write $M_{m \times n}(\mathbb{F})$, where $m, n \in \mathbb{N}$, to denote the *set of m -by- n matrices* whose entries are elements of \mathbb{F} .
- Given $n, m \in \mathbb{N}$, $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{F}^m$, we write

$$\text{cone}\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$$

to denote the *convex cone*

$$\left\{ \sum_{i=1}^n \lambda_i \mathbf{v}_i : \lambda_1, \dots, \lambda_n \geq 0 \right\}$$

generated by $\mathbf{v}_1, \dots, \mathbf{v}_n$.

- Given a subset $S \subseteq \mathbb{F}^n$, where $n \in \mathbb{N}$, we write S^\perp to denote the *orthogonal complement* of S .

2. The Theorem

2.1. The Theorem

We begin by showing the fundamental theorem of linear inequalities (Schrijver, 1986).

Theorem 1 (Fundamental Theorem of Linear Inequalities). *Let $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b} \in \mathbb{F}^m$, where $n, m \in \mathbb{N}$. Then exactly one of the following holds.*

1. $\mathbf{b} \in \text{cone}\{\mathbf{a}_i\}_{i=1}^n$.
2. There exists $\mathbf{c} \in \mathbb{F}^m$ such that the hyperplane

$$P = \{\mathbf{x} \in \mathbb{F}^m : \mathbf{c}^T \mathbf{x} = 0\}$$

has $t - 1$ linearly independent vectors from $\mathbf{a}_1, \dots, \mathbf{a}_n$ and that $\mathbf{c}^T \mathbf{b} < 0$ and $\mathbf{c}^T \mathbf{a}_i \geq 0$ for all $i \in \{1, \dots, n\}$, where $t = \text{rank}\{\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b}\}$.

The theorem itself can be quite difficult to sink in at the first glance, so before showing Schrijver's proof, we first examine similar statement in linear algebra as well as a 3-dimensional example for Theorem 1.

2.2. Systems of Linear Equalities

One big part of linear algebra is about solving *systems of linear equalities*, which can be written as

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

for some $A \in M_{m \times n}(\mathbb{F})$, $\mathbf{x} \in \mathbb{F}^n$, $\mathbf{b} \in \mathbb{F}^m$, where $m, n \in \mathbb{N}$. If (1) has a solution, then what (1) is essentially saying is that \mathbf{b} is a *linear combination of the columns of A* . In other words, if we write

$$A = [\mathbf{a}_1 \cdots \mathbf{a}_n], \mathbf{x} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix},$$

then

$$\mathbf{b} \stackrel{(1)}{=} A\mathbf{x} = [\mathbf{a}_1 \cdots \mathbf{a}_n] \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \sum_{i=1}^n \lambda_i \mathbf{a}_i. \tag{2}$$

Conversely, if (1) does not have a solution, then \mathbf{b} is not a linear combination of $\mathbf{a}_1, \dots, \mathbf{a}_n$, as (2) suggests. This means \mathbf{b} is linearly independent of $\mathbf{a}_1, \dots, \mathbf{a}_n$. We thus obtain the following characterization of *consistent* (i.e. having a solution) systems of linear equalities.

Theorem 2. Let $A \in M_{m \times n}(\mathbb{F})$, $\mathbf{b} \in \mathbb{F}^m$. Then exactly one of the following holds.

1. There exists $\mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \mathbf{b}$.
2. $\text{rank}(A) + 1 = \text{rank}[A|\mathbf{b}]$.

Observe that 1 of Theorem 2 is the condition which $A\mathbf{x} = \mathbf{b}$ is consistent, and it can be phrased as follows: $\mathbf{b} \in \text{span}\{\mathbf{a}_i\}_{i=1}^m$. This is where the similarities of the two theorems are revealed: Theorem 1 is concerned about systems of linear inequalities of the form

$$\begin{cases} A\mathbf{x} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{cases} \quad (3)$$

Saying that (3) is *feasible* is equivalent to 1 of Theorem 1, similar to how the consistency of $A\mathbf{x} = \mathbf{b}$ is equivalent to $\mathbf{b} \in \text{span}\{\mathbf{a}_i\}_{i=1}^m$. We also note that the competing conditions – 2's of Theorem 1, 2 – are also similar, but this similarity is different from how 1's of Theorem 1, 2 are similar to each other.

Say a student is asked to solve a system $A\mathbf{x} = \mathbf{b}$ and report if it does not have a solution. Suppose, after going through Gauss-Jordan elimination on $[A|\mathbf{b}]$, the student finds that $A\mathbf{x} = \mathbf{b}$ is inconsistent. Then, the student have two options to prove the conclusion:

1. show the whole elimination process, which could be quite long; or
2. use Theorem 2: calculate $\text{rank}[A|\mathbf{b}]$ and show that it is 1 greater than $\text{rank}(A)$.

Arguably the latter approach is simpler for checking if the result is correct or not. That is, 2 of Theorem 2 is useful when we desire to provide a *certificate* for inconsistency of a system of linear equalities.

In a similar manner, 2 of Theorem 1 provides *what* can be a certificate, namely \mathbf{c} , for the case which a system of linear inequalities of the form (3) is infeasible, although it does not say *how* we can obtain one. We demonstrate the *how* part with an example in the following subsection.

2.3. A Three-dimensional Example of The Theorem

Say we have the following:

$$A = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 3 & 3 & 3 & 3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}. \quad (4)$$

Figure 1. Column vectors $\mathbf{a}_1, \dots, \mathbf{a}_4$ of A and \mathbf{b}

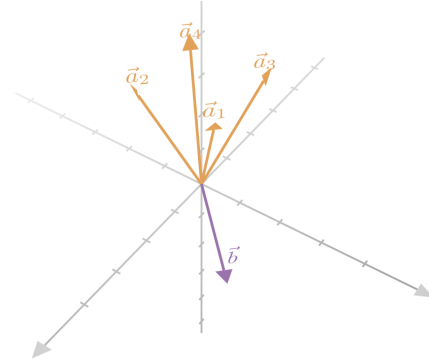


Figure 1 shows the column vectors

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}, \mathbf{a}_4 = \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix}$$

of A and \mathbf{b} .

From Figure 1, it *seems clear* that $\mathbf{b} \notin \text{cone}\{\mathbf{a}_i\}_{i=1}^4$. To concretely see this, we demonstrate *Schrijver's Algorithm*, an algorithm provided by Schrijver (Schrijver, 1986) in his proof of Theorem 1. Note that the following algorithm assumes that $\text{span}\{\mathbf{a}_i\}_{i=1}^n = \mathbb{F}^m$; cases which $\mathbf{a}_1, \dots, \mathbf{a}_n$ do not span the whole space will be dealt shortly.

Algorithm 1 Schrijver's Algorithm

Input: Column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b} \in \mathbb{F}^m$, where $n, m \in \mathbb{N}$ and $\text{span}\{\mathbf{a}_i\}_{i=1}^n = \mathbb{F}^m$.

Choose linearly independent $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_m} \in \{\mathbf{a}_i\}_{i=1}^n$.
repeat

Write \mathbf{b} as a linear combination $\mathbf{b} = \sum_{j=1}^m \lambda_{i_j} \mathbf{a}_{i_j}$ of $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_m}$, where $\lambda_{i_1}, \dots, \lambda_{i_m} \in \mathbb{F}$.

if $\forall j \in \{1, \dots, m\} [\lambda_{i_j} \geq 0]$ **then**

We are at case 1 of Theorem 1.

Terminate the loop.

end if

Let $h = \min \{i \in \{i_1, \dots, i_m\} : \lambda_{i_i} < 0\}$.

Consider $P = \text{span}\left(\{\mathbf{a}_{i_j}\}_{j=1, i_j \neq h}^n\right)$.

Choose nonzero $\mathbf{c} \in P^\perp$ with $\mathbf{c}^T \mathbf{a}_h > 0$.

if $\forall j \in \{1, \dots, m\} [\mathbf{c}^T \mathbf{a}_{i_j} \geq 0]$ **then**

We are at case 2 of Theorem 1.

Terminate the loop and report \mathbf{c} .

end if

Choose $s = \min \{i \in \{1, \dots, n\} : \mathbf{c}^T \mathbf{a}_i < 0\}$.

Replace \mathbf{a}_h with \mathbf{a}_s .

until

Let us see how Schrijver's algorithm works on our ex-

ample. We first choose $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, which forms a basis, say β , for \mathbb{F}^n . Observe that solving

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \mathbf{x} = \mathbf{b}$$

gives $\mathbf{x} = (2, -1, -1)$, so

$$\mathbf{b} = 2\mathbf{a}_1 - \mathbf{a}_2 - \mathbf{a}_3 = \underbrace{(2)}_{=\lambda_1} \mathbf{a}_1 + \underbrace{(-1)}_{=\lambda_2} \mathbf{a}_2 + \underbrace{(-1)}_{=\lambda_3} \mathbf{a}_3$$

is the unique way of writing \mathbf{b} as a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$. Moreover, we see that $\lambda_2, \lambda_3 < 0$, so the body of the first if in the loop is not being executed, and we proceed to find h , which is 2 for this case. This means

$$P = \text{span}\{\mathbf{a}_1, \mathbf{a}_3\},$$

and observe that

$$\mathbf{v} = \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}$$

is orthogonal to both $\mathbf{a}_1, \mathbf{a}_3$, which means $\mathbf{v} \in P^\perp$. But note that

$$\mathbf{v}^T \mathbf{a}_2 = \begin{bmatrix} 0 & 3 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} = -6 < 0,$$

so we scale \mathbf{v} by (-1) : let $\mathbf{c} = -\mathbf{v}$. Then we know that

$$\mathbf{c}^T \mathbf{a}_1 = \mathbf{c}^T \mathbf{a}_3 = 0, \mathbf{c}^T \mathbf{a}_2 \geq 0$$

from above, and also note that

$$\mathbf{c}^T \mathbf{a}_4 = \begin{bmatrix} 0 & -3 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix} = 6 \geq 0.$$

So indeed Schrijver's algorithm successfully produces a certificate \mathbf{c} for $\mathbf{b} \notin \text{cone}\{\mathbf{a}_i\}_{i=1}^4$.

In fact, this example is simple enough that we can directly verify $\mathbf{b} \notin \text{cone}\{\mathbf{a}_i\}_{i=1}^4$: suppose that

$$\mathbf{b} = \sum_{i=1}^4 \lambda_i \mathbf{a}_i$$

for some nonnegative $\lambda_1, \dots, \lambda_4$ for the sake of contradiction. If $\lambda_1 = \dots = \lambda_4 = 0$, then $\sum_{i=1}^4 \lambda_i \mathbf{a}_i = \mathbf{0}$, so we face a contradiction. On the other hand, if some $\lambda_i \neq 0$, then the last component of $\sum_{i=1}^4 \lambda_i \mathbf{a}_i$ is nonzero, so again, a contradiction occurs.

2.4. Correctness of Schrijver's Algorithm

It is clear from Schrijver's algorithm that if the loop terminates, then Schrijver's algorithm is correctly reports at which case we are. So let us prove this fact.

Lemma 3. *Given $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b} \in \mathbb{F}^m$, where $n, m \in \mathbb{N}$, such that $\text{span}\{\mathbf{a}_i\}_{i=1}^n = \mathbb{F}^m$, Schrijver's algorithm terminates.*

Proof. (Schrijver, 1986) Suppose, for the sake of contradiction, that the loop does not terminate. For all $r \in \mathbb{N}$, let β_r to be the set of m linearly independent \mathbf{a}_i 's used during the r th iteration of the loop. Since we have only finitely many choices for β_r , there exist $l, k \in \mathbb{N}$ such that $\beta_l = \beta_k$. We may assume that $l < k$.

This means we have $k - l$ h 's, say $h_l, \dots, h_{k-1} \in \{1, \dots, n\}$, that are the indices of the \mathbf{a}_i 's *popped off* from the corresponding β_r . Let

$$H = \max\{h_r\}_{r=l}^{k-1}$$

and let $p \in \{l, \dots, k-1\}$ be such that $\beta_p \setminus \beta_{p+1} = \{\mathbf{a}_H\}$ (i.e. \mathbf{a}_H is *popped off* from the basis β_p for \mathbb{F}^m that we have at p th iteration). Since $\beta_k = \beta_l$, there exists some other index $q \in \{l, \dots, k-1\}$ such that \mathbf{a}_H is *pushed in* to the basis at the q th iteration (i.e. $\beta_{q+1} \setminus \beta_q = \{\mathbf{a}_H\}$). Let $\mathbf{c}_q \in \mathbb{F}^m$ be the vector \mathbf{c} that we obtain at q th iteration, that

$$\mathbf{c}_q^T \mathbf{a}_i = 0$$

for all $\mathbf{a}_i \in \beta_q, \mathbf{a}_i \neq \mathbf{a}_{h_q}$ and that

$$\mathbf{c}_q^T \mathbf{a}_{h_q} > 0.$$

This means, in particular, that

$$\mathbf{c}_q^T \mathbf{b} = \underbrace{\lambda_{h_q}}_{<0} \underbrace{\mathbf{c}_q^T \mathbf{a}_{h_q}}_{>0} < 0, \quad (5)$$

since h_q is the smallest index at q th iteration such that the corresponding coefficient λ_{h_q} is negative.

On the other hand, if we write

$$\beta_p = \{\mathbf{a}_{i_j}\}_{j=1}^n$$

and

$$\mathbf{b} = \sum_{j=1}^n \lambda_{i_j} \mathbf{a}_{i_j}$$

for convenience, then observe that we can split $\mathbf{c}_q^T \mathbf{b}$ into three parts:

$$\mathbf{c}_q^T \mathbf{b} = \lambda_H \mathbf{c}_q^T \mathbf{a}_H + \sum_{\substack{1 \leq j \leq n \\ i_j < H}} \lambda_{i_j} \mathbf{c}_q^T \mathbf{a}_{i_j} + \sum_{\substack{1 \leq j \leq n \\ i_j > H}} \lambda_{i_j} \mathbf{c}_q^T \mathbf{a}_{i_j}. \quad (6)$$

Now observe the following.

1. \mathbf{a}_H plays the role of \mathbf{a}_s at q th iteration (i.e. replacing \mathbf{a}_{h_q}), so it follows that

$$\forall j \in \{1, \dots, n\} [i_j < H \implies \mathbf{c}_q^T \mathbf{a}_{i_j} \geq 0], \quad (7)$$

and that

$$\mathbf{c}_q^T \mathbf{a}_H < 0. \quad (8)$$

2. Since $H = \max \{h_r\}_{r=l}^{k-1}$, it follows that, for every $i > H$, \mathbf{a}_i is neither added to nor removed from the bases $\beta_l, \dots, \beta_{k-1}$. So in particular,

$$\forall j \in \{1, \dots, n\} [i_j > H \implies \mathbf{a}_{i_j} \in \beta_q],$$

which means

$$\forall j \in \{1, \dots, n\} [i_j > H \implies \mathbf{c}_q^T \mathbf{a}_{i_j} = 0], \quad (9)$$

since $i_j > H \geq h_q$ so $i_j \neq h_q$.

3. On the other hand, at p th iteration, \mathbf{a}_H is *popped off*, so

$$\forall j \in \{1, \dots, n\} [i_j < H \implies \lambda_{i_j} \geq 0], \quad (10)$$

and that

$$\lambda_H < 0. \quad (11)$$

Combining (6), ..., (11) gives

$$\mathbf{c}_q^T \mathbf{b} > 0 \quad (12)$$

after some cleanup. But now note that (5) and (12) are contradicting each other, as desired. \square

Corollary 4. *Fundamental theorem of linear inequalities is true when $\mathbf{a}_1, \dots, \mathbf{a}_n$ span the whole space.*

Proof. To see that 1, 2 of Theorem 1 exclude each other, suppose, for the sake of contradiction, that both hold. Let $\lambda_1, \dots, \lambda_n \geq 0$ be such that

$$\mathbf{b} = \sum_{i=1}^n \lambda_i \mathbf{a}_i$$

and let \mathbf{c} be the vector mentioned in 2. Then we have

$$0 \stackrel{2}{>} \mathbf{c}^T \mathbf{b} = \mathbf{c}^T \sum_{i=1}^n \lambda_i \mathbf{a}_i \stackrel{2}{\geq} 0,$$

so we have a contradiction. To see why at least one of 1, 2 occurs when the premise of the theorem is met, see the proof of Schrijver's algorithm. \square

After his proof, Schrijver mentions that (Schrijver, 1986) *the above proof of this fundamental theorem also gives a fundamental algorithm: it is a disguised form of the famous simplex method, with Bland's rule incorporated.*

Let us see what this means shortly. But before this, we briefly consider the case which $\mathbf{a}_1, \dots, \mathbf{a}_n$ do not span the whole space. Suppose that

$$\underbrace{\text{span} \{\mathbf{a}_i\}_{i=1}^n}_{=W} \neq \mathbb{F}^m.$$

If $\mathbf{b} \notin W$, then $\mathbf{b} \notin \text{cone} \{\mathbf{a}_i\}_{i=1}^n \subseteq W$. Since $\dim(W) < m$, W^\perp is nonempty so by choosing any $\mathbf{c} \in W^\perp$ with $\mathbf{c}^T \mathbf{b} < 0$,¹ we are done. Otherwise, note that we can restrict Schrijver's algorithm to W and the proof of termination is essentially the same.

3. Schrijver's Algorithm and the Simplex Method

3.1. Turning Inequalities to Equalities with Nonnegativity Constraints

Suppose that we have a system of linear inequalities

$$\begin{aligned} a_{11}z_1 + \dots + a_{1k}z_k &\leq b_1 \\ &\vdots \\ a_{m1}z_1 + \dots + a_{mk}z_k &\leq b_m \end{aligned} \quad (13)$$

with an objective function

$$\sum_{j=1}^k c_j z_j$$

which we desire to minimize, where $a_{ij}, b_i, c_j \in \mathbb{F}$. To turn (13) to a system of linear *equalities* with positive constraints (i.e. a system of the form (3)), we have two tricks.

1. *slack variables*: Define

$$s_i = b_i - \sum_{j=1}^k a_{ij} z_j \quad (14)$$

for all $i \in \{1, \dots, m\}$. s_i 's are called *slack variables*.

2. *difference of two positive numbers*: Since any real number can be written as a difference of two non-negative real numbers, define z_j^+, z_j^- such that

$$z_j = z_j^+ - z_j^- \quad (15)$$

for all $j \in \{1, \dots, k\}$.

¹Concretely speaking, we choose any $\mathbf{v} \in W^\perp$ and if $\mathbf{v}^T \mathbf{b} < 0$, we take $\mathbf{c} = \mathbf{v}$; otherwise, we take $\mathbf{c} = -\mathbf{v}$

Observe that s_i 's and z_j^+, z_j^- 's are nonnegative by definition, so (13) becomes

$$\begin{aligned} a_{11}(z_1^+ - z_1^-) + \cdots + a_{1k}(z_k^+ - z_k^-) + s_1 &= b_1 \\ &\vdots \\ a_{m1}(z_1^+ - z_1^-) + \cdots + a_{mk}(z_k^+ - z_k^-) + s_m &= b_m \end{aligned}$$

with a nonnegativity constraint

$$s_i, z_j^+, z_j^- \geq 0$$

for all $i \in \{1, \dots, m\}, j \in \{1, \dots, k\}$. Also the objective function becomes

$$\sum_{j=1}^k c_j (z_j^+ - z_j^-).$$

More compactly, we may define

$$A = \begin{bmatrix} a_{11} & a_{11} & \cdots & a_{1k} & a_{1k} & 1 & 0 & \cdots & 0 \\ a_{21} & a_{21} & \cdots & a_{2k} & a_{2k} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m1} & \cdots & a_{mk} & a_{mk} & 0 & 0 & \cdots & 1 \end{bmatrix}$$

and

$$\mathbf{x} = \begin{bmatrix} z_1^+ \\ z_1^- \\ \vdots \\ z_k^+ \\ z_k^- \\ s_1 \\ \vdots \\ s_m \end{bmatrix}, \mathbf{c} = \begin{bmatrix} c_1 \\ c_1 \\ \vdots \\ c_k \\ c_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

to write a system equivalent to (13)

$$A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0} \quad (16)$$

with the objective function

$$\mathbf{c}^T \mathbf{x}.$$

Observe that any solution of (16) is associated with a solution of (13) by means of (14) and (15) and vice versa, which means if we minimize $\mathbf{c}^T \mathbf{x}$ with constraints (16), then we obtain an optimal solution for (13) as well.

For reference, denote $n = 2k + m$ so that $\mathbf{x}, \mathbf{c} \in \mathbb{F}^n$ and $A \in M_{m \times n}(\mathbb{F})$.

3.2. Phase II

Suppose that we are given a polytope P in some (rational or Euclidean) space, which represents the *feasible*

region of an LP problem. A key insight is that at least one of the vertices of P is an optimal solution for the LP problem. Given this, the simplex algorithm operates as follows (Dantzig, August 1963):

- *phase I*: Find a vertex \mathbf{v} of P .
- *phase II*: While we have a neighboring vertex \mathbf{u} of \mathbf{v} has an objective value at least that of \mathbf{v} , we move to \mathbf{u} (i.e. update \mathbf{v} with \mathbf{u}).

Of course, the above description is purely geometrical, and the algebraic description is as follows.

Observe that, if we set $x_1 = \cdots = x_{2k} = 0$ (i.e. the first $2k$ entries of \mathbf{x}) and $x_{2k+i} = b_i$ for all $i \in \{1, \dots, m\}$ (i.e. the last m entries of \mathbf{x}), then (16) tells us that \mathbf{x} at least satisfies the first constraint: $A\mathbf{x} = \mathbf{b}$. If \mathbf{x} also satisfies the nonnegativity constraint $\mathbf{x} \geq \mathbf{0}$, then we have found a basic feasible solution for (16). In general, given any solution \mathbf{x} , if there are m indices $j_1, \dots, j_m \in \{1, \dots, n\}$ such that

1. $x_j = 0$ for all $j \in \{1, \dots, n\} \setminus \{j_1, \dots, j_m\}$; and
2. $A_{\{j_1, \dots, j_m\}}$ (i.e. the square submatrix of A whose columns are j_1, \dots, j_m th columns of A , respectively) is invertible;

then \mathbf{x} is called a *basic feasible solution*. If we remove the nonnegativity constraint, then \mathbf{x} is called a *basic solution*.

We accept the following result without proof (Jiri Matousek, 2007).

Lemma 5. *Let P be the feasible region of an LP problem in equational form (i.e. a system of linear equalities with nonnegativity constraints). Then the following are equivalent about $\mathbf{v} \in P$.*

1. \mathbf{v} is a vertex.
2. \mathbf{v} is a basic feasible solution.

So, by Lemma 5, the two phases of the simplex algorithm can be equivalently written as:

- *phase I*: Find a basic feasible solution \mathbf{v} of the LP problem.
- *phase II*: While we have a neighboring basic feasible solution \mathbf{u} of \mathbf{v} has an objective value at least that of \mathbf{v} , we move to \mathbf{u} (i.e. update \mathbf{v} with \mathbf{u}).

Throughout the rest of this subsection, let us fix an LP problem

$$\begin{aligned} & \text{minimize } \mathbf{c}^T \mathbf{x} \\ & \text{subject to } A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \quad (17)$$

where $A \in M_{m \times n}(\mathbb{F})$, $\mathbf{c} \in \mathbb{F}^n$, $\mathbf{b} \in \mathbb{F}^m$. We also denote

$$P = \{\mathbf{x} \in \mathbb{F}^n : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\},$$

the feasible region for (17). Since we are at phase II, we are also assuming that we already obtained a basic feasible solution $\mathbf{v} \in P$ from phase I. Let

$$\mathcal{B}^{(0)} = \{j_1^{(0)}, \dots, j_m^{(0)}\} \subseteq \{1, \dots, n\}$$

be the set of indices such that

$$v_j = 0$$

for all $j \in \{1, \dots, n\} \setminus \mathcal{B}$ and that

$$A_{\mathcal{B}^{(0)}} = \begin{bmatrix} \mathbf{a}_{j_1^{(0)}} & \cdots & \mathbf{a}_{j_m^{(0)}} \end{bmatrix} \quad (18)$$

is invertible, where each \mathbf{a}_j represents the j th column of A . We may also order the set of indices, say $\mathcal{N}^{(0)}$, that are not included in $\mathcal{B}^{(0)}$ in some way and define $A_{\mathcal{N}^{(0)}}$ analogous to (18). Then, the matrix

$$\begin{bmatrix} A_{\mathcal{B}^{(0)}} & A_{\mathcal{N}^{(0)}} \end{bmatrix}$$

is obtained by *swapping rows* of A . Since we assumed $A_{\mathcal{B}^{(0)}}$ is invertible, we may perform Gauss-Jordan elimination on it:

$$A_{\mathcal{B}^{(0)}}^{-1} \begin{bmatrix} A_{\mathcal{B}^{(0)}} & A_{\mathcal{N}^{(0)}} | \mathbf{b} \end{bmatrix} = \begin{bmatrix} I & A_{\mathcal{B}^{(0)}}^{-1} A_{\mathcal{N}^{(0)}} | A_{\mathcal{B}^{(0)}}^{-1} \mathbf{b} \end{bmatrix}.$$

Note that we are utilizing the fact that swapping rows does not affect the overall sum, as long as we make the same changes for the order which the entries of the vector that is right-multiplied occur. That is, if we define $\mathbf{v}_{\mathcal{B}^{(0)}}$, $\mathbf{v}_{\mathcal{N}^{(0)}}$ analogous to (18), then

$$\begin{bmatrix} I & A_{\mathcal{B}^{(0)}}^{-1} A_{\mathcal{N}^{(0)}} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{\mathcal{B}^{(0)}} \\ \mathbf{v}_{\mathcal{N}^{(0)}} \end{bmatrix} = A_{\mathcal{B}^{(0)}}^{-1} \mathbf{b},$$

meaning that

$$\mathbf{v}_{\mathcal{B}^{(0)}} = A_{\mathcal{B}^{(0)}}^{-1} \mathbf{b},$$

since we know $\mathbf{v}_{\mathcal{N}^{(0)}} = \mathbf{0}$.

To make our discussion little easier, let us think this process in terms of *variables*: that is, we are going to use $\mathbf{x} = (x_1, \dots, x_n)$ as placeholder for a feasible solution for (17) (or, equivalently, a point of P), and we are going to treat the entries x_1, \dots, x_n to be variables. Then the *basic variables* associated with $\mathcal{B}^{(0)}$

are x_{j_1}, \dots, x_{j_m} , and other variables are called *non-basic* (this shows why we denoted the sets with \mathcal{B}, \mathcal{N} , respectively).

After the elimination, there is a convenient test for an optimality. First, denote

$$z = \mathbf{c}^T \mathbf{x} = \begin{bmatrix} \mathbf{c}_{\mathcal{B}^{(0)}}^T & \mathbf{c}_{\mathcal{N}^{(0)}}^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\mathcal{B}^{(0)}} \\ \mathbf{x}_{\mathcal{N}^{(0)}} \end{bmatrix},$$

the objective function. If we use the augmented matrix notation, then we would instead write:

$$\begin{bmatrix} \mathbf{c}_{\mathcal{B}^{(0)}}^T & \mathbf{c}_{\mathcal{N}^{(0)}}^T | z \end{bmatrix}.$$

This viewpoint allows us to combine the (eliminated) system of linear equalities and the objective function into a single augmented matrix:

$$\left[\begin{array}{cc|c} I & A_{\mathcal{B}^{(0)}}^{-1} A_{\mathcal{N}^{(0)}} & A_{\mathcal{B}^{(0)}}^{-1} \mathbf{b} \\ \mathbf{c}_{\mathcal{B}^{(0)}}^T & \mathbf{c}_{\mathcal{N}^{(0)}}^T & z \end{array} \right].$$

By utilizing the fact that we have an identity block, we can *eliminate* basic variables from the objective function:

$$\left[\begin{array}{cc|c} I & A_{\mathcal{B}^{(0)}}^{-1} A_{\mathcal{N}^{(0)}} & A_{\mathcal{B}^{(0)}}^{-1} \mathbf{b} \\ \mathbf{0}^T & \mathbf{c}'^T & z - d \end{array} \right], \quad (19)$$

where

$$\mathbf{c}' = \mathbf{c}_{\mathcal{N}^{(0)}} - \mathbf{c}_{\mathcal{B}^{(0)}}^T A_{\mathcal{B}^{(0)}}^{-1} A_{\mathcal{N}^{(0)}}$$

and

$$d = \mathbf{c}_{\mathcal{B}^{(0)}}^T A_{\mathcal{B}^{(0)}}^{-1} \mathbf{b}.$$

This means the objective function can be described entirely by using nonbasic variables only, namely

$$z = \mathbf{c}'^T \mathbf{x}_{\mathcal{N}^{(0)}} + d.$$

When we substitute $\mathbf{v}_{\mathcal{N}^{(0)}}$ in place of $\mathbf{x}_{\mathcal{N}^{(0)}}$,

$$z = \mathbf{c}'^T \mathbf{v}_{\mathcal{N}^{(0)}} + d.$$

Now the interesting part is $\mathbf{c}' = (c'_i)_{i \in \mathcal{N}^{(0)}}$ but not the constant term afterwards. The optimality test then utilizes the fact that, when we change any variable, we have to change at least one nonbasic variable, and that we may change a nonbasic variable only in a way that it increases.

1. If each $c'_i > 0$, then changing (i.e. increasing) any nonbasic variables results in the increase in z . Since z is completely described as a linear combination of nonbasic variables (up to a constant), it follows that \mathbf{v} is optimal.
2. If there is some index $i \in \mathcal{N}^{(0)}$ such that $c'_i = 0$, then increasing x_i does not affect the value of z . In this case, the test is inconclusive.

3. If there is some index $i \in \mathcal{N}^{(0)}$ such that $c'_i < 0$, then increasing x_i reduces the value of z . Hence \mathbf{v} is not optimal.

If 1 occurs, then we can simply report \mathbf{v} ; otherwise, in both cases 2, 3 (say $i = s$ is such that $c'_i \leq 0$), we try to increase x_s as much as possible.² That is, each row (except the last row) of (19) describes how much x_s can be increased: suppose that the h th row of (19) is

$$\begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & a'_{h(m+1)} & \cdots & a'_{hn} & | & b'_h \end{bmatrix}.$$

Then by fixing other nonbasic variables as constants and assuming that $a'_{hs} \neq 0$, the largest magnitude for x_s is $\left| \frac{b'_h}{a'_{hs}} \right|$ (i.e. by setting $x_{j_h} = 0, x_s = \frac{b'_h}{a'_{hs}}$, the equality for the h th row still holds). So we consider the ratios $\frac{b'_1}{a'_{1s}}, \dots, \frac{b'_m}{a'_{ms}}$. Since we want x_s to be non-negative at the end of the day, we better exclude $\frac{b'_h}{a'_{hs}}$'s that are negative. Moreover, among the nonnegative $\frac{b'_h}{a'_{hs}}$'s, we have to choose the smallest one, since otherwise the nonnegativity constraint for every x_{j_k} 's with $\frac{b'_k}{a'_{ki}} < \frac{b'_h}{a'_{hs}}$ is broken.³

After we choose the h th row, by using elementary operations, we eliminate x_s from every other row, including the row that describes the objective function. Then by swapping h th column and i th column (which can be thought as interchanging the h th entry of $\mathcal{B}^{(0)}$ and the place of i in $\mathcal{N}^{(0)}$, resulting in new sets of basic and nonbasic indices $\mathcal{B}^{(1)}, \mathcal{N}^{(1)}$), we again obtain a system of the form (19), where every occurrence of $^{(0)}$ is replaced with $^{(1)}$.

The above process of choosing which variable to *leave* (i.e. x_h) and which to *enter* (i.e. x_s) the set of basic variables is known as *pivoting*. Also note that the new basic feasible solution \mathbf{v}' that results from pivoting is again a vertex of P by Lemma 5. In other words, what we described so far in this subsection is more-or-less an algebraic machinery of checking optimality and choosing which neighboring vertex to move to, in case the current vertex is not optimal.

3.3. Bland's Rule

Observe that there could be more than one choice for the entering variable, and the same happens for the leaving variable. To deal with this, there are various *pivot rules* used to systematically choose the variables

²Note that there could be more than one choices for s here; we will come back to this shortly when we discuss Bland's rule.

³A subtlety occurs when there is a tie; we will come back to this shortly when we discuss Bland's rule.

from the indices. Unfortunately, under certain circumstances, known as *degeneracy*,⁴ bad choices of the pivot rule result in *cycling* (Jiri Matousek, 2007). Although it has known that cycling happens very rarely in practice (Jiri Matousek, 2007), we nevertheless have certain pivot rules that guarantees to terminate without cycling, albeit inefficient. One of them is called *Bland's rule*.

In short, *Bland's rule always chooses smallest available index*. That is, whenever there is a tie for s (i.e. the index for the leaving variable), out of the candidates, Bland's rule chooses the smallest one. It does so for h (i.e. the index for the entering variable). More detailed explanation can be found in Bland's original paper (Bland, 1977).

For us, it is more interesting to figure out how Bland's rule relates to Schrijver's algorithm. So suppose that we are given the inputs

$$\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b} \in \mathbb{F}^m$$

for the algorithm, where we also know that $n, m \in \mathbb{N}$ and $\text{span}\{\mathbf{a}_i\}_{i=1}^n = \mathbb{F}^m$. The step that we choose linearly independent $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_m}$ that spans the whole space before the main iteration corresponds to choosing the initial basic variables, and the index set $\{i_1, \dots, i_m\}$ corresponds to $\mathcal{B}^{(0)}$ that we used in the description of the simplex method.

Then writing \mathbf{b} as a linear combination of $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_m}$ is equivalent to row-reducing the matrix $[A_{\mathcal{B}^{(0)}} \ A_{\mathcal{N}^{(0)}} | \mathbf{b}]$, and the coefficients $\lambda_{i_1}, \dots, \lambda_{i_m}$ are read-off from the rightmost column after the row-reduction. That is,

$$\begin{bmatrix} \lambda_{i_1} \\ \vdots \\ \lambda_{i_m} \end{bmatrix} = A_{\mathcal{B}^{(0)}}^{-1} \mathbf{b}.$$

Note that $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$.

After this is where things get interesting. Bland's rule operates in phase II: it tells us how to move from a basic feasible solution to a better one (or one at least as good as the current) without cycling. But we are not there yet; we are struggling to figure out if the system $A\mathbf{x} = \mathbf{b}$ with $\mathbf{x} \geq \mathbf{0}$ is feasible.

Fortunately, we can use the following trick: instead of moving from a basic feasible solution to another basic feasible solution, we move from a basic solution to another basic solution.

⁴We call a linear program *degenerate* if there exists a basic feasible solution that has at least one basic variable equal to 0.

In the description of phase II, note that we chose the entering variable first and then the leaving variable. However, for Schrijver's algorithm, note that we choose the leaving variable first. The reason is clear when we consider $A_{\mathcal{B}(0)}^{-1}\mathbf{b}$: we want this vector to have nonnegative entries only. That is, we desire to have an index set, say $\mathcal{B}^{(p)}$, at some p th iteration such that $A_{\mathcal{B}(0)}^{-1}\mathbf{b}$ to be nonnegative. For this reason, the appropriate basic variables to be removed are the ones associated the rows of the row reduced matrix with negative entries at the end. These are exactly the candidates that Schrijver's algorithm has for h (i.e. the leaving index), and it chooses the smallest index as Bland's rule does.

After choosing h , Schrijver's algorithm also chooses a vector \mathbf{c} perpendicular to the linear hyperplane spanned by $\{\mathbf{a}_{i_j}\}_{j=1, i_j \neq h}^n$ and lies at the same side with \mathbf{a}_h .⁵ Observe that, by definition, $\mathbf{c}^T A$ is the h th row of $A_{\mathcal{B}(0)}^{-1} A = [I \quad A_{\mathcal{B}(0)}^{-1} A_{\mathcal{N}(0)}]$ up to multiplying a positive scalar.⁶ This means the indices j 's such that $\mathbf{c}^T \mathbf{a}_j < 0$ are the indices such that the j th entry of the h th row of row-reduced matrix is negative. Since the h th entry of $A_{\mathcal{B}(0)}^{-1}\mathbf{b}$ is also negative, it follows that we have to choose one of j 's in order to replace x_h by x_j . So we choose the smallest j as Bland's algorithm does, which we denoted as s , and replace x_h by x_s .

Although what we are doing here is slightly different from the actual Bland's rule (which operates at phase II as remarked), proof presented by Schrijver, as shown in the previous section, shows that the analogous approach works.

4. Further Topics

4.1. Consequences of the Fundamental Theorem

A significant part of the fundamental theorem of linear inequalities is that it serves as a more-or-less *fundamental lemma* to many theoretical results in linear programming. For instance, the famous *Farkas' lemma* is a direct consequence of the theorem (Schrijver, 1986).

Theorem 6 (Farkas' Lemma). *Let $A \in M_{m \times n}(\mathbb{F})$ and let $\mathbf{b} \in \mathbb{F}^m$. Then there exists $\mathbf{x} \in \mathbb{F}^n$ with $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$ if and only if $\mathbf{y}^T \mathbf{b} \geq 0$ for each $\mathbf{y} \in \mathbb{F}^m$ with $\mathbf{y}^T A \geq \mathbf{0}$.*

Proof. For the forward direction, suppose that there

⁵Any $n - 1$ dimensional linear hyperplane in \mathbb{F}^n separates the space into two disjoint sides.

⁶Equivalently, \mathbf{c}^T is the h th row of $A_{\mathcal{B}(0)}^{-1}$ up to a constant.

exists $\mathbf{x} \in \mathbb{F}^n$ with $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$. Then given any $\mathbf{y} \in \mathbb{F}^m$ such that $\mathbf{y}^T A \geq \mathbf{0}$, we have

$$\mathbf{y}^T \mathbf{b} = \mathbf{y}^T A \underbrace{\mathbf{x}}_{\geq \mathbf{0}^T \geq \mathbf{0}} \geq 0.$$

Conversely, suppose that $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$ is infeasible. Let $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{F}^m$ be the columns of A . Then $\mathbf{b} \notin \text{cone}\{\mathbf{a}_j\}_{j=1}^n$ by the fundamental theorem of linear inequalities, and the theorem provides a certificate \mathbf{c} . This \mathbf{c} is exactly what we want for \mathbf{y} . \square

We also present other important consequences of the theorem without proofs. They are all from *Schrijver's Theory of Linear and Integer Programming* (Schrijver, 1986).

Theorem 7 (Minkowski-Weyl Theorem). *A convex cone in an Euclidean space is polyhedral if and only if finitely generated.*

Corollary 8 (Decomposition Theorem). *$P \subseteq \mathbb{R}^n$ is a polyhedron if and only if $P = Q + C$ for some polytope $Q \subseteq \mathbb{R}^n$ and polyhedral cone $C \subseteq \mathbb{R}^n$.*

Theorem 9 (Strong Duality Theorem). *Let $A \in M_{m \times n}(\mathbb{F}), \mathbf{b} \in \mathbb{F}^m, \mathbf{c} \in \mathbb{F}^n$. Then*

$$\max\{\mathbf{c}^T \mathbf{x} : A\mathbf{x} \leq \mathbf{b}\} = \min\{\mathbf{y}^T \mathbf{b} : \mathbf{y} \geq \mathbf{0}, \mathbf{y}^T A = \mathbf{c}^T\}$$

provided that the sets are nonempty.

4.2. Phase I: More Efficient Feasibility Check

As we noted in the previous section, cycling almost never happens in practice. For this reason, Bland's rule is not popular in practice (Jiri Matousek, 2007), and we use different algorithms to check the feasibility of an LP problem. We demonstrate the one that is introduced by Dantzig (Dantzig, August 1963). Suppose that we are given an LP problem

$$\begin{aligned} &\text{minimize } \mathbf{c}^T \mathbf{x} \\ &\text{subject to } A\mathbf{x} = \mathbf{b} \quad , \\ &\quad \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{20}$$

where $\mathbf{c} \in \mathbb{F}^n, \mathbf{b} \in \mathbb{F}^m$. We also assume that $\mathbf{b} \geq \mathbf{0}$, since we may multiply -1 to both sides of any equality. We then introduce *artificial variables* (or *error variables*) x_{n+1}, \dots, x_{n+m} . That is, we consider an LP problem

$$\begin{aligned} &\text{minimize } \sum_{j=1}^m x_{n+j} \\ &\text{subject to } [A \quad I] \mathbf{z} \quad , \\ &\quad \mathbf{z} \geq \mathbf{0} \end{aligned} \tag{21}$$

where $\mathbf{z} = (\mathbf{x}, x_{n+1}, \dots, x_{n+m})$. Note that (21) is feasible doubtlessly, by setting $\mathbf{x} = \mathbf{0}$ and $x_{n+j} = b_j$ for all

$j \in \{1, \dots, m\}$, where b_j is the j th entry of \mathbf{b} . Then, instead of using Bland's rule, we use Dantzig's rule (see (Dantzig, August 1963)) for pivoting.

After obtaining an optimal solution for (21), if the optimal objective value is 0, then the optimal solution we have is nothing more than a basic feasible solution for (20). Hence we proceed to phase II. Otherwise, if the optimal objective value is strictly greater than 0, then we know (20) is not feasible.

4.3. Implementing Schrijver's Algorithm

Implementing Schrijver's algorithm is as straightforward as its description. An implementation can be found at <https://github.com/snochi/schrijver>.

References

- Bland, R. G. New finite pivoting rules for the simplex method. *Mathematics of Operations Research*, 2(2): 103–107, 1977. ISSN 0364765X, 15265471. URL <http://www.jstor.org/stable/3689647>.
- Dantzig, G. B. *Linear Programming and Extensions*. Princeton University Press, August 1963.
- Jiri Matousek, B. G. *Understanding and Using Linear Programming*. Springer Science & Business Media, 2007.
- Schrijver, A. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1986.