TWO PROOFS OF FARKAS' LEMMA

DAVID RAHMAN

This note presents an algebraic proof of Farkas' Lemma based on Gauss-Jordan elimination, following Gale (1960), and a geometric proof based on the separating hyperplane theorem. $\mathbb{R}^n$ is our ambient space, with x, y, b, c denoting vectors, A, B denoting matrices, λ, μ scalars, and m, n natural numbers.

1. Linear Independence, Bases and Rank

The vectors $x_1, \ldots, x_k \in \mathbb{R}^n$ are *linearly independent* if given $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$,

$$\alpha_1 x_1 + \cdots + \alpha_k x_k = \mathbf{0} \quad \Rightarrow \quad \alpha_1 = \cdots = \alpha_k = 0.$$

Lemma 1. *If each of the vectors $y_0, y_1, \ldots, y_m \in \mathbb{R}^n$ is a linear combination of the vectors $x_1, \ldots, x_m \in \mathbb{R}^n$ then the y_j's are linearly dependent.*[1]

Proof. Proceed by induction on m. Starting with $m = 1$, let $y_0 = \alpha_0 x_1$ and $y_1 = \alpha_1 x_1$. If both α_0 and α_1 are equal to zero then $y_0 = y_1 = 0$, so the y_j's are dependent. Otherwise, without loss say $\alpha_0 \neq 0$. Now, $\alpha_0 y_1 - \alpha_1 y_0 = \alpha_0 \alpha_1 x_1 - \alpha_1 \alpha_0 x_1 = 0 x_1 = 0$, implying dependence.

Assume that the claim holds for $m = k - 1$ and that

$$y_j = \sum_{i=1}^{k} \alpha_{ij} x_i, \qquad j = 0, 1, \ldots, k.$$

In other words, each y_j is a linear combination of the x_i's. We must show that the y_j's are linearly dependent. Again, if all the α_{ij}'s equal zero then all the y_j's are zero, implying linear dependence. Now assume that not all the α_{ij}'s are zero. Without loss of generality, say $\alpha_{10} \neq 0$ and define, for $j = 1, \ldots, k$,

$$z_j = y_j - \frac{\alpha_{1j}}{\alpha_{10}} y_0 = \sum_{i=1}^{k} \alpha_{ij} x_i - \frac{\alpha_{1j}}{\alpha_{10}} \sum_{i=1}^{k} \alpha_{i0} x_i = \sum_{i=2}^{k} \left(\alpha_{ij} - \frac{\alpha_{1j}}{\alpha_{10}} \alpha_{i0} \right) x_i.$$

But now each of the k vectors $z_1, \ldots, z_k$ is a linear combination of the $k-1$ vectors $x_2, \ldots, x_k$. By the induction hypothesis, the z_j's are linearly dependent, that is, there exist numbers $\beta_1, \ldots, \beta_k$, not all zero, such that $\beta_1 z_1 + \cdots + \beta_k z_k = 0$. But since

$$\mathbf{0} = \sum_{j=1}^{k} \beta_j z_j = \sum_{j=1}^{k} \beta_j y_j - \frac{y_0}{\alpha_{10}} \sum_{j=1}^{k} \beta_j \alpha_{1j},$$

it follows that the y_j's are linearly dependent. □

Date: January 29, 2025.

Department of Economics, University of Minnesota.

[1] That is, not linearly independent: $\alpha_0 y_0 + \alpha_1 y_1 + \cdots + \alpha_m y_m = 0$ for some $\alpha_0, \alpha_1, \ldots, \alpha_m$, not all zero.

Every vector in $\mathbb{R}^n$ is a linear combination of $e_1, \ldots, e_n$—where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ has a 1 in the ith place and zeros elsewhere for every i—so every collection of $n+1$ vectors is linearly dependent by Lemma 1. The following corollary rephrases this observation.

Corollary 1. *Every system of n homogeneous linear equations in $n+1$ unknowns has a nonzero solution.*

Proof. Consider the system of homogeneous equations

$$\sum_{j=0}^{n} \alpha_{ij}\xi_j = 0, \qquad i = 1, \ldots, n.$$

Let $a_j = (\alpha_{1j}, \ldots, \alpha_{nj})$ for $j = 0, \ldots, n$. These are $n+1$ vectors in $\mathbb{R}^n$, so by the previous paragraph they are linearly dependent. Therefore, there are numbers $\xi_0, \xi_1, \ldots, \xi_n$, not all zero, such that

$$\sum_{j=0}^{n} a_j \xi_j = \mathbf{0},$$

i.e., the numbers $\xi_0, \xi_1, \ldots, \xi_n$ are a nonzero solution of our n homogeneous equations. □

Clearly, Corollary 1 extends immediately to n equations in any number of unknowns greater than n. If $S \subset \mathbb{R}^n$ is a collection of vectors, the *rank* of S is the maximum number of linearly independent vectors that can be simultaneously selected from S. If the rank of S is r then any set of r linearly independent vectors from S is called a *basis* of S.

Lemma 2. *Let $S \subset \mathbb{R}^n$. The vectors $x_1, \ldots, x_r \in S$ are a basis of S if and only if every vector in S is a linear combination of the x_i's.*

Proof. For necessity, suppose that every vector $y \in S$ is a linear combination of the x_i's. By Lemma 1, S contains no larger set of linearly independent vectors, since any set of more than r vectors must be linearly dependent by virtue of being linear combinations of the x_i's. Therefore, S has rank r and the x_i's form a basis of S.

For sufficiency, suppose that the x_i's form a basis of S. By definition of basis, for any other vector $y \in S$, the collection $\{y, x_1, \ldots, x_r\}$ is linearly dependent, implying

$$\alpha_0 y + \sum_{i=1}^{r} \alpha_i x_i = \mathbf{0}$$

for some scalars $\alpha_0, \alpha_1, \ldots, \alpha_r$, not all zero. Moreover, $\alpha_0 \neq 0$, since otherwise the x_i's would be linearly dependent. Therefore,

$$y = -\frac{1}{\alpha_0} \sum_{i=1}^{r} \alpha_i x_i$$

and y is a linear combination of the x_i's, as claimed. □

By Lemma 2, the vectors $e_1, \dots, e_n$ form a basis of $\mathbb{R}^n$, as does any collection of n linearly independent vectors. If $A \in \mathbb{R}^{m\times n}$ is a matrix, its *row rank* is the rank of its collection of row vectors, and its *column rank* is the rank of its collection of column vectors.

Lemma 3. *The row rank and column rank of a matrix coincide.*

Proof. Let r be the row rank and s the column rank of a given matrix

$$A = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{bmatrix}.$$

For a contradiction, suppose that $r < s$. Choose a row basis for A. Reordering rows if necessary, assume that the row basis consists of the first r rows of A, labeled $a^1, \dots, a^r$. Likewise, choose a column basis for A, which, by reordering columns if necessary, we may assume consists of the first s columns of A, labeled $a_1, \dots, a_s$.

Let $\hat{a}^i = (\alpha_{i1}, \dots, \alpha_{is})$ be the vector consisting of the first s entries of the ith row of A. Since $r < s$, by Corollary 1, the system of equations

$$\hat{a}^i y = 0, \qquad i = 1, \dots, r,$$

has a nonzero solution vector $\bar{y}$. Furthermore, since $a^1, \dots, a^r$ form a row basis for A, by Lemma 2, for every row a^k there exist numbers $\beta_{1k}, \dots, \beta_{rk}$ such that

$$a^k = \sum_{i=1}^{r} \beta_{ik} a^i, \qquad k = 1, \dots, m.$$

Truncating the last $n - s - 1$ entries from each vector in the equations above trivially implies

$$\hat{a}^k = \sum_{i=1}^{r} \beta_{ik} \hat{a}^i, \qquad k = 1, \dots, m,$$

therefore

$$\hat{a}^k \bar{y} = \left(\sum_{i=1}^{r} \beta_{ik} \hat{a}^i \right) \bar{y} = \sum_{i=1}^{r} \beta_{ik} \left(\hat{a}^i \bar{y} \right) = 0, \qquad k = 1, \dots, m.$$

Letting $\bar{y} = (\bar{\eta}_1, \dots, \bar{\eta}_s)$, we may write the k equations above in vector form as follows:

$$\sum_{j=1}^{s} \bar{\eta}_j a_j = \mathbf{0}.$$

But this implies that the column vectors $a_1, \dots, a_s$ are linearly dependent, which contradicts the assumption that they are a basis. Therefore, our assumption that $r < s$ must have been flawed, and $r \geq s$. A symmetric argument exchanging the roles of rows and columns yields the inequality $s \geq r$, whence $r = s$, as was claimed. $\square$

By Lemma 3, we may simply refer to the *rank* of a matrix A without having to distinguish between row rank and column rank.

Corollary 2. *If the vectors $a_1, \dots, a_n \in \mathbb{R}^m$ are linearly independent then for any numbers $\gamma_1, \dots, \gamma_n$ there is a vector $y \in \mathbb{R}^m$ such that*

$$ya_j = \gamma_j, \qquad j = 1, \dots, n.$$

Proof. Let A be the matrix whose columns are $a_1, \dots, a_n$. By Lemma 3, the rows of A have rank n, so if the vectors $a^1, \dots, a^n$ form a row basis of A (after reshuffling of rows, if necessary) then they constitute n linearly independent vectors. By Lemma 2, therefore, every vector $c = (\gamma_1, \dots, \gamma_n)$ is a linear combination of $a^1, \dots, a^n$, that is,

$$c = \sum_{i=1}^{n} \eta_i a^i$$

for some numbers $\eta_1, \dots, \eta_n$. Reorganizing this system of equations yields precisely $\gamma_j = ya_j$ for $j = 1, \dots, n$, as required, with the vector $y = (\eta_1, \dots, \eta_n, 0, \dots, 0)$. □

2. Linear Equations and Inequalities

Theorem 1. *Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, either* (i) *there exists $x \in \mathbb{R}^n$ such that $Ax = b$, or* (ii) *there exists $y \in \mathbb{R}^m$ such that $yA = \mathbf{0}$ and $yb = 1$, but not both.*

Proof. First, let us see that (i) and (ii) cannot both hold simultaneously. Otherwise, there would exist vectors x and y such that $Ax = b$, $yA = \mathbf{0}$ and $yb = 1$. Multiplying by y gives, though, $0 = \mathbf{0}x = yAx = yb = 1$, a contradiction if there ever was one. Next, let us show that if (i) fails then (ii) must hold.[2] To this end, let $a_1, \dots, a_s$ be a column basis for A. Since, by hypothesis, there is no x such that $Ax = b$, it must be the case that the vectors $a_1, \dots, a_s, b$ are together linearly independent, otherwise there would be a solution x to $Ax = b$. By Corollary 2, there exists a vector y such that $ya_j = 0$ for $j = 1, \dots, s$ and $yb = 1$. But since the vectors $a_1, \dots, a_s$ form a column basis for A, every column vector a_k can be written as a linear combination $a_k = \lambda_1 a_1 + \cdots + \lambda_s a_s$, therefore $ya_k = y(\lambda_1 a_1 + \cdots + \lambda_s a_s) = \lambda_1 ya_1 + \cdots + \lambda_s ya_s = 0$ for $k = 1, \dots, n$. In other words, $yA = \mathbf{0}$ and $yb = 1$, i.e., (ii) holds, as required. □

Example 1. Consider the following system of two equations in three unknowns:

$$\begin{bmatrix} -4 & 2 & -5 \\ 2 & -1 & 2.5 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Multiplying both sides by the vector $(1, 2)$, we obtain the implication

$$0\xi_1 + 0\xi_2 + 0\xi_3 \;=\; 3,$$

which is clearly impossible. Hence, there is no vector (ξ_1, ξ_2, ξ_3) that solves the system of equations above.

[2] This is logically equivalent to the assertion that if (ii) fails then (i) must hold.

Our next result, known as Farkas' Lemma, extends Theorem 1 above by including linear inequalities. It is a crucial first result on which our duality framework will be based.

Let us begin with some preliminary background for its proof. Suppose that $x = (\xi_1, \ldots, \xi_n)$ solves the system of equations $Ax = b$. Equivalently, $yAx = yb$ for every vector y. Solving out the variable ξ_n from this system of equations could involve the following calculations:

$$yAx = yb \quad \Leftrightarrow \quad y\sum_{j=1}^{n} a_j\xi_j = yb \quad \Leftrightarrow \quad ya_n\xi_n + \sum_{j=1}^{n-1} ya_j\xi_j = yb.$$

Therefore, $ya_n \neq 0$ implies the following expression for ξ_n:

$$\xi_n = \frac{1}{(ya_n)}\left[(yb) - \sum_{j=1}^{n-1}(ya_j)\xi_j\right].$$

Of course, $\xi_n \geq 0$ if $ya_j \geq 0$, $\xi_j \geq 0$ for $j = 1, \ldots, n-1$, $yb < 0$ and $ya_n < 0$. We will find such y and ξ_j in the proof below when looking for a nonnegative solution to $Ax = b$. Substituting this expression for ξ_n back into the original system of equations yields

$$\begin{aligned}
\sum_{j=1}^{n-1} a_j\xi_j + a_n\xi_n = b \quad &\Leftrightarrow \quad \sum_{j=1}^{n-1} a_j\xi_j + \frac{1}{(ya_n)}\left[(yb) - \sum_{j=1}^{n-1}(ya_j)\xi_j\right] a_n = b \\
\text{multiply both sides by } (ya_n) \quad &\Leftrightarrow \quad \sum_{j=1}^{n-1}(ya_n)a_j\xi_j + \sum_{j=1}^{n-1}\left[(yb) - (ya_j)\xi_j\right] a_n = (ya_n)b \\
\text{collect terms and rearrange} \quad &\Leftrightarrow \quad \sum_{j=1}^{n-1}\left[(ya_n)a_j - (ya_j)a_n\right]\xi_j = (ya_n)b - (yb)a_n.
\end{aligned}$$

This is precisely the system of equations in $n-1$ unknowns that is used in the proof of Farkas' Lemma below. The vectors $(ya_n)b - (yb)a_n$ and $(ya_n)a_j - (ya_j)a_n$ are what is left to satisfy the system of equations with unknowns $(\xi_1, \ldots, \xi_{n-1})$ after substituting for ξ_n above.

Theorem 2. *Given $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, either* (i) *there exists $x \in \mathbb{R}^n$ such that $x \geq \mathbf{0}$ and $Ax = b$, or* (ii) *there exists $y \in \mathbb{R}^m$ such that $yA \geq \mathbf{0}$ and $yb < 0$, but not both.*

Proof. Notice first that (i) and (ii) cannot hold simultaneously: otherwise, $Ax = b$ implies $yAx = yb$, but $yA \geq \mathbf{0}$ together with $x \geq \mathbf{0}$ imply $yAx \geq 0$, contradicting $yb < 0$. Therefore, if either (i) or (ii) holds then only that one alternative holds. It remains to show that if (i) fails then (ii) holds.[3] To this end, assume that there does not exist a vector $x \geq \mathbf{0}$ such that $Ax = b$. This could be for two reasons: (a) the equation $Ax = b$ has no solution, even when the restriction that x be nonnegative is rescinded, or (b) there is a solution to the equation $Ax = b$ but not one that also satisfies $x \geq \mathbf{0}$. If (a) is the reason, so $Ax = b$ has no solution at all, then by Theorem 1 there exists a vector y such that $Ay = \mathbf{0}$ and $yb = -1$, implying (ii). Now suppose that (b) is the reason, so there exists a solution x to $Ax = b$ but no nonnegative solution. We will proceed by induction on n, the number of columns of A.

[3]Of course, showing that if (ii) fails then (i) must hold is logically equivalent.

If $n = 1$ then $Ax = b$ becomes $a_1\xi_1 = b$, where $\xi_1 \in \mathbb{R}$ and $a_1, b \in \mathbb{R}^m$. With the previous paragraph's logic, assume that this equation does have a solution but no nonnegative one: $\xi_1 < 0$. Letting $y = -b$, clearly $yb = -b \cdot b < 0$ and $ya_1 = yb/\xi_1 = -b \cdot b/\xi_1 > 0$, so (ii) holds.

Pursuing induction, assume now that the theorem holds for the first $n-1$ columns of A. Clearly, if $Ax = b$ has no nonnegative solution then neither does the equation

$$\sum_{j=1}^{n-1} a_j\xi_j = b,$$

which imposes $\xi_n = 0$. By the induction hypothesis, there exists a vector $\hat{y} \in \mathbb{R}^m$ such that $\hat{y}a_j \geq 0$ for $j = 1, \ldots, n-1$ and $\hat{y}b < 0$. If $\hat{y}$ also satisfies $\hat{y}a_n \geq 0$ then $\hat{y}$ satisfies (ii) and we are done, so suppose that $\hat{y}a_n < 0$ and let

$$\hat{a}_j = (\hat{y}a_n)a_j - (\hat{y}a_j)a_n, \qquad j = 1, \ldots, n-1, \qquad \hat{b} = (\hat{y}a_n)b - (\hat{y}b)a_n.$$

The vectors $\hat{a}_j$ and $\hat{b}$ satisfy $\hat{y}\hat{a}_j = \hat{y}\hat{b} = 0$. Consider the following system of equations:

$$\sum_{j=1}^{n-1} \hat{a}_j\hat{\xi}_j = \hat{b}.$$

This system of equations can have no nonnegative solution. To see why, if $\hat{\xi}_1, \ldots, \hat{\xi}_{n-1}$ were such a nonnegative solution then substituting for $\hat{a}_j$ and $\hat{b}$ would yield

$$\begin{aligned}\sum_{j=1}^{n-1} \hat{a}_j\hat{\xi}_j = \hat{b} \quad &\Leftrightarrow \quad \sum_{j=1}^{n-1} \left[(\hat{y}a_n)a_j - (\hat{y}a_j)a_n\right]\hat{\xi}_j = (\hat{y}a_n)b - (\hat{y}b)a_n \\ &\Leftrightarrow \quad \sum_{j=1}^{n-1} a_j\hat{\xi}_j + \frac{1}{(\hat{y}a_n)}\left[(\hat{y}b) - \sum_{j=1}^{n-1}(\hat{y}a_j)\hat{\xi}_j\right] a_n = b.\end{aligned}$$

But $\hat{y}a_n < 0$, $\hat{y}a_j \geq 0$ and $\hat{\xi}_j \geq 0$ for $j = 1, \ldots, n-1$, and $\hat{y}b < 0$, therefore

$$\hat{\xi}_n = \frac{1}{(\hat{y}a_n)}\left[(\hat{y}b) - \sum_{j=1}^{n-1}(\hat{y}a_j)\hat{\xi}_j\right] \geq 0,$$

so the vector $(\hat{\xi}_1, \ldots, \hat{\xi}_n)$ would constitute a nonnegative solution of the system $Ax = b$, contradicting our assumption that it has no nonnegative solution. Hence, by induction, there exists a vector $\hat{z} \in \mathbb{R}^m$ such that $\hat{z}\hat{a}_j \geq 0$ for $j = 1, \ldots, n-1$ and $\hat{z}\hat{b} < 0$.

Finally, let

$$y = (\hat{y}a_n)\hat{z} - (\hat{z}a_n)\hat{y}.$$

It is easy to see that this vector y satisfies

$$\begin{aligned} ya_n &= \left[(\hat{y}a_n)\hat{z} - (\hat{z}a_n)\hat{y}\right]a_n = (\hat{y}a_n)(\hat{z}a_n) - (\hat{z}a_n)(\hat{y}a_n) = 0, \\ yb &= \left[(\hat{y}a_n)\hat{z} - (\hat{z}a_n)\hat{y}\right]b = (\hat{y}a_n)\hat{z}b - (\hat{z}a_n)\hat{y}b = \hat{z}\left[(\hat{y}a_n)b - (\hat{y}b)a_n\right] = \hat{z}\hat{b} < 0, \text{ and} \\ ya_j &= \left[(\hat{y}a_n)\hat{z} - (\hat{z}a_n)\hat{y}\right]a_j = (\hat{y}a_n)\hat{z}a_j - (\hat{z}a_n)\hat{y}a_j = \hat{z}\left[(\hat{y}a_n)a_j - (\hat{y}a_j)a_n\right] = \hat{z}\hat{a}_j \geq 0 \end{aligned}$$

for $j = 1, \ldots, n-1$. Therefore, the vector y satisfies (ii) and the theorem is proved. □

Example 2. Consider the following system of two equations in three unknowns:

$$\begin{bmatrix} 4 & 1 & -5 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The alternative to existence of a nonnegative solution to the system above is

$$4\eta_1 + \eta_2 \geq 0, \qquad \eta_1 \geq 0, \qquad -5\eta_1 + 2\eta_2 \geq 0, \qquad \eta_1 + \eta_2 < 0.$$

This alternative system of inequalities is impossible to satisfy. Indeed, $\eta_1 \geq 0$ and $\eta_1 + \eta_2 < 0$ imply $\eta_2 < 0$, which together with $\eta_1 \geq 0$ contradicts $-5\eta_1 + 2\eta_2 \geq 0$. By Farkas' Lemma there exists a nonnegative vector (ξ_1, ξ_2, ξ_3) that solves the system of equations above.

Given a matrix A, consider its column vectors $a_1, \ldots, a_n \in \mathbb{R}^m$. The set of all nonnegative linear combinations of these column vectors forms a cone-shaped region surrounded by the column vectors themselves. That the equation $Ax = b$ has a nonnegative solution x can be interpreted as the vector b lying inside this cone. Conversely, the statement that there is no nonnegative solution to $Ax = b$ means that the vector b does not lie in the cone generated by the column vectors of A. In this case, Farkas' Lemma asserts that there is a vector y that makes an acute angle with a_j for $j = 1, \ldots, n$ and an obtuse angle with b. Dually, there exists a hyperplane that separates b and the cone generated by $a_1, \ldots, a_n$. See Figure 1 below.

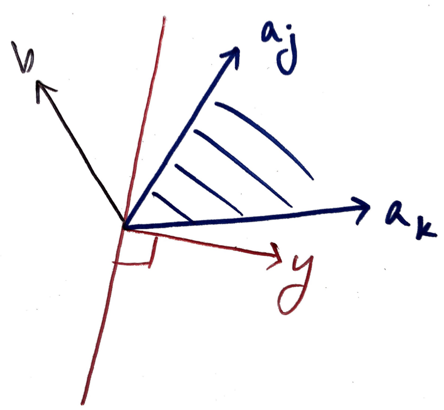

FIGURE 1. Illustration of Farkas' Lemma

3. A Separating Hyperplane Theorem

A subset $C \subset \mathbb{R}^n$ is *convex* if $\lambda x + (1 - \lambda) y \in C$ for every $x, y \in C$ and $\lambda \in [0, 1]$. A subset $H \subset \mathbb{R}^n$ is called a *hyperplane* if $H = \{x \in \mathbb{R}^n : px = \alpha\}$ for some $p \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, where $px = \sum_i p_i x_i$. A closed *halfspace* is any subset of the form $\{x \in \mathbb{R}^n : px \leq \alpha\}$, and an open halfspace any subset of the form $\{x \in \mathbb{R}^n : px < \alpha\}$. I will also use the more compact notation $[px = \alpha]$, $[px \leq \alpha]$, etc. A subset of $S \subset \mathbb{R}^n$ is *closed* if it contains all its limit points, and open if for every $y \in S$ there exists $\varepsilon > 0$ such that $B_\varepsilon(y) = [\|x - y\| < \varepsilon] \subset S$, where $\|z\| = (\sum_i z_i^2)^{1/2}$ is the usual Euclidean norm.

Lemma 4. *If $C \subset \mathbb{R}^n$ is a closed, convex set and $y \in \mathbb{R}^n \setminus C$ a point outside C then there exists $x^* \in C$ that minimizes the distance between y and C:*

$$x^* \in \arg\min_{x \in C} \|x - y\| .$$

Moreover, $\|x^* - y\| > 0$.

Proof. First, notice that $\delta = \inf\{\|x - y\| : x \in C\} > 0$. By definition of the infimum, there is a sequence $\{x_m\} \subset C$ such that $\|x_m - y\| \to \delta$ as $m \to \infty$. Taking a subsequence if necessary, for every $\varepsilon > 0$ there is a sequence $\{x_m\} \subset C$ such that $\delta \leq \|x_m - y\| \leq \delta + \varepsilon$. The closed set $\{x \in \mathbb{R}^n : \|x - y\| \in [\delta, \delta + \varepsilon]\}$ is contained in the closed ball with center y and radius $\delta + \varepsilon$, hence is bounded, too. Therefore, $\{x_m\}$ has a convergent subsequence by the Bolzano-Weierstrass Theorem. Let x^* be its limit. Continuity of Euclidean distance implies $\|x^* - y\| = \delta$. By hypothesis, C is closed, so $x^* \in C$. But $y \notin C$, so $x^* \neq y$ and $\delta > 0$. □

Proposition 1. *If $C \subset \mathbb{R}^n$ is a closed, convex set and $y \in \mathbb{R}^n \setminus C$ a point outside C then there exists a hyperplane $[px = \alpha]$ that separates y and C, that is, $py < \alpha < px$ for all $x \in C$.*

Proof. After a translation by $-y$, assume without loss that $y = 0$. Let $x^* \in C$ be a point of C that minimizes the distance between C and 0, that is, $x^* \in \arg\min\{\|x\| : x \in C\}$. Such x^* exists by Lemma 3.2. Let m be the midpoint on the line joining 0 and x^*, that is, $m = \frac{1}{2}x^*$. Let $[px = \alpha]$ be the hyperplane that passes through m and is perpendicular to the vector x^*. Specifically, let $p = x^*/\|x^*\|$ and $\alpha = pm = (x^*/\|x^*\|) \cdot x^*/2 = \frac{1}{2}(x^* \cdot x^*)/\|x^*\| = \frac{1}{2}\|x^*\| > 0$.

We will now show that the vector 0 lies on one side of this hyperplane and every $x \in C$ lies on the other side. First, $p0 = 0 < \alpha$, so 0 belongs to the open halfspace $[px < \alpha]$. Next, notice that $px^* = x^* \cdot x^*/\|x^*\| = \|x^*\| > \frac{1}{2}\|x^*\| = \alpha$, so x^* belongs instead to $[px > \alpha]$. Finally, consider any $x \in C$. Since C is a convex set, $\lambda x + (1-\lambda)x^* \in C$ for every $\lambda \in (0,1)$, and since x^* minimizes the norm in C, clearly $\|x^*\|^2 \leq \|\lambda x + (1-\lambda)x^*\|^2$. But $\|x\|^2 = x \cdot x$, therefore

$$\begin{aligned} x^* \cdot x^* &\leq (x^* + \lambda(x - x^*)) \cdot (x^* + \lambda(x - x^*)) \\ &= x^* \cdot x^* + 2\lambda x^* \cdot (x - x^*) + \lambda^2 (x - x^*) \cdot (x - x^*) \\ \Leftrightarrow \quad 0 &\leq x^* \cdot (x - x^*) + \tfrac{1}{2}\lambda(x - x^*) \cdot (x - x^*). \end{aligned}$$

Letting $\lambda > 0$ decrease to zero, it follows that $x^* \cdot (x - x^*) \geq 0$. But $x \in C$ was arbitrary, so $x^* \cdot (x - x^*) \geq 0$ for every $x \in C$. Since $x^* = 2m$, we may rewrite this inequality as $\|x^*\| p \cdot (x - 2m) \geq 0$, or equivalently $px \geq 2pm$. Since $pm > 0$, clearly $2pm > pm$, hence $px > pm = \alpha$, proving that indeed x lies on the other side of our hyperplane, as required. □

This separating hyperplane theorem leads to a geometric proof of Farkas' Lemma. First, a preliminary result. The *cone* generated by a set of vectors $a_1, \ldots, a_n$ is the set

$$\text{cone}\{a_1, \ldots, a_n\} = \{\lambda_1 a_1 + \cdots + \lambda_n a_n : (\lambda_1, \ldots, \lambda_n) \geq 0\},$$

For any matrix $A \in \mathbb{R}^{m \times n}$, the cone generated by its columns is denoted by cone(A).

Lemma 5. *If $A \in \mathbb{R}^{m\times n}$ is any matrix, cone(A) is a closed convex set.*

Proof. The following proof is based on Vohra (2005). For convexity, let $y, y' \in \text{cone}(A)$. By definition of cone(A), there exist $x, x' \geq 0$ such that $y = Ax$ and $y' = Ax'$. Given $\lambda \in [0,1]$,

$$\lambda y + (1-\lambda)y' = \lambda Ax + (1-\lambda)Ax' = A(\lambda x + (1-\lambda)x'),$$

so, since $\lambda x + (1-\lambda)x' \geq 0$, we must have $\lambda y + (1-\lambda)y' \in \text{cone}(A)$.

We prove that cone(A) is closed in two steps. First, suppose that all the columns of A are linearly independent. Let $\{w_t\} \subset \text{cone}(A)$ be a convergent sequence, with limit w. We will show that $w \in \text{cone}(A)$. Since each w_t belongs to cone(A), there exists $x_t \geq 0$ such that $w_t = Ax_t$. Because $A'A$ and A have equal rank, $A'A$ is invertible and the projection $(A'A)^{-1}A'$ is continuous. Therefore, $Ax_t \to w$ implies $x_t \to (A'A)^{-1}A'w = x$. Since A is continuous, $x_t \to x$ implies $Ax_t = w_t \to w = Ax$, and $w \in \text{cone}(A)$.

When not necessarily all of A's columns are linearly independent, we will show that cone(A) is the union of the cones generated by all the linearly independent subsets of columns of A. Let $b \in \text{cone}(A)$, so there is a vector $x = (\xi_1, \ldots, \xi_n) \geq 0$ such that $Ax = b$. Let $S = \{j : \xi_j > 0\}$ and B be the matrix consisting of the columns indexed by S. Although there may be many ways of expressing b as a nonnegative linear combination of the columns of A, assume that x uses the fewest number of columns, therefore b cannot be expressed as a nonnegative linear combination of fewer than $|S|-1$ columns of A. By construction, $b \in \text{cone}(B)$. If the columns of B are linearly independent, we are done. If not, there is a nonzero linear combination of the columns of B, call it $z = (\zeta_1, \ldots, \zeta_n)$ with $\{j : \zeta_j > 0\} \subset S$, such that $Az = 0$. For any $t \in \mathbb{R}$, therefore, $A(x - tz) = b$. We will find a scalar t such that (1) $\xi_j - t\zeta_j \geq 0$ for all $j \in S$ and (2) $\xi_j - t\zeta_j = 0$ for at least one $j \in S$. This leads to a contradiction, since with such t, the vector b can be written as a nonnegative linear combination of $|S|-1$ columns of B.

We may find such t as follows. If $\zeta_j > 0$ for all $j \in S$, let $t = \min_j\{\xi_j/\zeta_j : j \in S\}$ and j^* satisfy $\xi_{j^*}/\zeta_{j^*} = t$. Such t clearly satisfies (1) and (2), since $\xi_j - t\zeta_j \geq \xi_j - (\xi_j/\zeta_j)\zeta_j \geq 0$ for every j with equality at j^*. If on the other hand $\zeta_j < 0$ for at least one $j \in S$, set $t = -\max_j\{|\xi_j/\zeta_j| : \zeta_j < 0\}$ and let j^* satisfy $\xi_{j^*}/\zeta_{j^*} = t$. For any $j \in S$ such that $\zeta_j \geq 0$, clearly $\xi_j - t\zeta_j \geq 0$, since $t < 0$ and $\xi_j \geq 0$. For any $j \in S$ such that $\zeta_j < 0$, the definition of t gives $\xi_j - t\zeta_j \geq \xi_j - (\xi_j/\zeta_j)\zeta_j \geq 0$ with equality at j^*. Hence, the columns of B are linearly independent. Since b was arbitrary, every $b \in \text{cone}(A)$ can be written as nonnegative linear combination from a linearly independent subset of columns of A.

Our final step in proving that cone(A) is closed is the following. Since A has finitely many column vectors, it has finitely many linearly independent subsets of column vectors. Having proved that every $b \in \text{cone}(A)$ belongs to cone(B) for some matrix B whose columns are a linearly independent subset of the columns of A, and that cone(B) is closed, it follows that cone(A) is closed from the fact that cone(A) is the union of finitely many closed sets, namely the finitely many cone(B)'s, and the union of finitely many closed sets is closed. □

Theorem 3. *Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, either* (i) *there exists $x \in \mathbb{R}^n$ such that $x \geq 0$ and $Ax = b$, or* (ii) *there exists $y \in \mathbb{R}^m$ such that $yA \geq \mathbf{0}$ and $yb < 0$, but not both.*

Geometric Proof. We will use the separating hyperplane theorem to prove that if (i) fails then (ii) must hold. (The argument that (i) and (ii) cannot both hold is omitted here, see the previous algebraic proof.) If (i) fails then $b \notin \text{cone}(A) = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$. By Lemma 5, cone$(A)$ is a closed convex set, so by Proposition 1 there exists a hyperplane $[px = \alpha]$ that separates b from cone(A), that is, $pb < \alpha < py$ for all $y \in \text{cone}(A)$. Since $\mathbf{0} \in \text{cone}(A)$, it follows that $\alpha < 0$. If a_j is the jth column vector of A, we will prove that $pa_j \geq 0$, showing that the vector $y = p$ satisfies (ii). Since $\lambda a_j \in \text{cone}(A)$ for every $\lambda \geq 0$ and $px > \alpha$ for every $x \in \text{cone}(A)$, it follows that $p\lambda a_j > \alpha$. If $pa_j < 0$ then making λ arbitrarily large renders λpa_j less than α, a contradiction. Therefore, $pa_j \geq 0$ for all j. □

4. Exercises

Exercise 1. Show that for every system of linear inequalities and/or equations there exists a matrix A, a vector b and vector of variables x such that the original system has a solution if and only if there exists a nonnegative solution $x \geq 0$ to the system $Ax = b$. Show that the same conclusion holds for the system $Ax \leq b$, as well as the system $Ax \leq b$ for $x \geq 0$.

Exercise 2. Prove that given a matrix A and vector b, either there exists a vector x such that $Ax \leq b$ or there exists a vector $y \geq 0$ such that $yA = 0$ but $yb < 0$. Prove that given a matrix A and vector b, either there exists a vector $x \geq 0$ such that $Ax \leq b$ or there exists a vector $y \geq 0$ such that $yA \geq 0$ but $yb < 0$.

Exercise 3. A square matrix $A \in \mathbb{R}^{n \times n}$ is called *stochastic* if $\sum_i \alpha_{ij} = 1$ for $j = 1, \ldots, n$. An *invariant distribution* for A is a vector $\pi \in \mathbb{R}^n$ such that $\pi \geq 0$, $\sum_j \pi_j = 1$ and $A\pi = \pi$. Show that every stochastic matrix has an invariant distribution.

Exercise 4. Prove that the union of finitely many closed sets is closed. Prove that the arbitrary intersection of (i) closed sets is closed, and (ii) convex sets is convex.

Exercise 5. Given two subsets $S, T \subset \mathbb{R}^n$, their *Minkowski* sum is the set

$$S + T = \{y + z : y \in S, z \in T\}.$$

Show that the Minkowski sum of two convex sets is a convex set.

Exercise 6. Show that if C and D are two closed, convex sets such that $C \cap D = \emptyset$ and C is compact then there is a hyperplane $[px = \alpha]$ such that $px > \alpha > py$ for all $z \in C$ and $y \in D$. Find an example of two disjoint, noncompact, closed and convex sets for which no such hyperplane exists. (A subset of $\mathbb{R}^n$ is *compact* if it is closed and bounded.)