CHAPTER 4: Convex Sets and Separation Theorems

1 Convex sets

Convexity is often assumed in economic theory since it plays important roles in optimization. The convexity of preferences can be interpreted as capturing consumer's liking for variety, and the convexity of production set is related to the existence of nonincreasing returns to scale.

Definition 1 (Convex set) *A set X in $\mathbb{R}^n$ is convex if given any two points x' and x'' in X, the point*

$$x^{\lambda} = (1-\lambda)x' + \lambda x''$$

is also in X for every $\lambda \in [0, 1]$.

A vector of the form x^{λ} as defined above is called a *convex combination* of x' and x''. The set of all convex combinations of x' and x'' is the line segment connecting these two points, which is denoted by $[x', x'']$.

Example 1 Following are the examples of convex sets.

(a) Given a set X, and two points x' and x'' in X, the line segment $[x', x'']$ is a convex set.

(b) The disc $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq r\}$ is a convex set.

(c) The circle $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = r\}$ is not a convex set.

(d) The set $H(p, \alpha) = \{x \in \mathbb{R}^n \mid p \cdot x = \alpha$ for $p \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}\}$ is a convex set. ■

Lemma 1 *Following are some useful properties relating to convex sets.*

(a) An arbitrary intersection of convex sets is convex.

(b) If X and Y are convex sets in $\mathbb{R}^n$, and $\alpha \in \mathbb{R}$, then the sets

$$\begin{aligned} X + Y &= \{z \in \mathbb{R}^n \mid z = x + y \text{ for some } x \in X \text{ and } y \in Y\}, \text{ and} \\ \alpha X &= \{z \in \mathbb{R}^n \mid z = \alpha x \text{ for some } x \in X\} \end{aligned}$$

are convex.

(c) If X is a convex set in $\mathbb{R}^n$, then $int(X)$ and $cl(X)$ are also convex.

Proof. The proof is left as an exercise. □

2 Convex hull

First we introduce a generalization of the notion of convex combination to more than two vectors in $\mathbb{R}^n$.

Definition 2 (Convex combination) *A point $y \in \mathbb{R}^n$ is said to be a convex combination of m vectors $x^1, \dots, x^m \in \mathbb{R}^n$ if it can be written as*

$$y = \sum_{i=1}^{m} \lambda_i x^i, \;\; \textit{with } \lambda_i \in [0, 1] \textit{ for all } i \;\textit{ and } \sum_{i=1}^{m} \lambda_i = 1.$$

The following theorem gives a characterization of convexity in terms of convex combinations.

Theorem 1 *A set X is convex if and only if every convex combination of points in X lies in X.*

We are sometimes interested in extending a set X so that is becomes convex by adding as few points to it as possible. The resulting set is called the convex hull of X.

Definition 3 (Convex hull) *Let X be a set in $\mathbb{R}^n$. The convex hull of X, denoted $co(X)$, is the smallest convex set that contains X.*

Clearly, there is at least one convex set that contains X, namely, $\mathbb{R}^n$ itself. If there are more, $co(X)$ is the intersection of all such sets.

Theorem 2 *The convex hull of X is the set of all convex combinations of elements in X, i.e.,*

$$co(X) = \left\{ y \in \mathbb{R}^n \mid y = \sum_{i=1}^{m} \lambda_i x^i \textit{ for some } m, \textit{ with } x^i \in X, \; \lambda_i \in [0, 1] \textit{ for all } i, \; \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

Proof. Let Y be the set of all convex combinations of elements in X. Clearly, Y contains X since any x in X can be written as a trivial convex combination with itself. Next we show that Y is a convex set. Let y^1 and y^2 be two points in Y such that

$$\begin{aligned} y^1 &= \sum_{i=1}^{m} \lambda_i x^i, \;\; \text{with } \lambda_i \in [0, 1] \text{ for all } i \text{ and } \sum_{i=1}^{m} \lambda_i = 1, \\ y^2 &= \sum_{j=1}^{n} \mu_j x^j, \;\; \text{with } \mu_j \in [0, 1] \text{ for all } j \text{ and } \sum_{j=1}^{n} \mu_j = 1. \end{aligned}$$

Then for some $\alpha \in [0, 1]$,

$$y = (1-\alpha) y^1 + \alpha y^2 = \sum_{i=1}^{m} (1-\alpha)\lambda_i x^i + \sum_{j=1}^{n} \alpha \mu_j x^j \, .$$

Since $(1-\alpha)\lambda_i \in [0, 1]$ for each i, $\alpha\mu_j \in [0, 1]$ for each j, and $\sum_{i=1}^{m}(1-\alpha)\lambda_i + \sum_{j=1}^{n} \alpha\mu_j = 1$. Thus, y is a convex combination of points in X, and hence is in Y. Moreover, any convex set that contains X must include all convex combinations of points in X, and must therefore contains Y. Thus, Y is the smallest convex set containing X, i.e, $Y = co(X)$. $\square$

The previous theorem says that any point in the convex hull of X can be expressed as a convex combination of a finite number of points in X, but it does not tell us how many such points are required. The following theorem says that if X is a subset of the n-dimensional Euclidean space, then this convex combination can be created with at most $n+1$ points in X.

Theorem 3 (Caratheodory's theorem) *Let $X \subseteq \mathbb{R}^n$. If y is a convex combination of points in X, then y is a convex combination of $n+1$ or fewer points in X.*

The idea behind the above theorem is the following. Consider the set $X = \{(0,0), (0,1), (1,0), (1,1)\}$, a subset of $\mathbb{R}^2$. The convex hull of this set is the square with vertices $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$. Consider now the point $y = (1/4, 1/4)$, which is in $co(X)$. We can construct a set $X' = \{(0,0), (0,1), (1,0)\}$, the convex hull of which is the triangle with vertices $(0,0)$, $(0,1)$ and $(1,0)$ that contains y.

3 Separation theorems

In this section we analyze some important results regarding convex sets that have interesting applications in economic theory, especially in general equilibrium.

Definition 4 (Hyperplane and half-spaces) *Given $p \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, the hyperplane generated by p and α is the set of points in $\mathbb{R}^n$ given by*

$$H(p, \alpha) = \{x \in \mathbb{R}^n \mid p \cdot x = \alpha\}.$$

The above hyperplane $H(p, \alpha)$ divides $\mathbb{R}^n$ into two regions: an upper half-space and a lower half-space, which are respectively given by

$$\begin{aligned} H_U(p, \alpha) &= \{x \in \mathbb{R}^n \mid p \cdot x \geq \alpha\}, \\ H_L(p, \alpha) &= \{x \in \mathbb{R}^n \mid p \cdot x \leq \alpha\}. \end{aligned}$$

It is easy to show that $H(p, \alpha)$, $H_U(p, \alpha)$ and $H_L(p, \alpha)$ are convex sets.

Example 2 Following are the examples of hyperplanes.

(a) If $n = 1$, the hyperplane $H(p, \alpha) = \{x \in \mathbb{R} \mid px = \alpha\}$ is the set $\{\alpha/p\}$. The half-spaces are given by $H_U(p, \alpha) = [\alpha/p, \infty)$ and $H_L(p, \alpha) = (-\infty, \alpha/p]$.

(b) If $n = 2$, $p = (p_1, p_2) \in \mathbb{R}^2$ and $p_2 \neq 0$, then

$$H(p, \alpha) = \{(x_1, x_2) \in \mathbb{R}^2 \mid p_1 x_1 + p_2 x_2 = \alpha\}$$

is the set of points lying on the straight line with slope $-p_1/p_2$ that passes through the point $(0, \alpha/p_2)$. ■

Definition 5 (Separating hyperplane) *A hyperplane $H(p, \alpha) \subset \mathbb{R}^n$ separates two subsets X and Y of $\mathbb{R}^n$ if $X \subset H_U(p, \alpha)$ and $Y \subset H_L(p, \alpha)$, or vice-versa. Moreover, the separation is strict if $X \cap H(p, \alpha) = \varnothing$ and $Y \cap H(p, \alpha) = \varnothing$.*

Example 3 Let $X = \{(x, y) \in \mathbb{R}^2 \mid y \geq x^2\}$ and $Y = \{(x, y) \in \mathbb{R}^2 \mid y \geq -x^2 - 1\}$. Then the hyperplane $H((0, 1), -1/2)$, i.e., the line $y = -1/2$ separates X and Y. Notice that $H((0, 1), -1/2)$ is one of the infinitely many hyperplanes that separate X and Y. ■

Lemma 2 *Given $X \subset \mathbb{R}^n$ non-empty, closed and convex, and $x^* \notin X$, there exists a unique $y^* \in X$ such that*

$$d(x^*, y^*) = \min\{d(x^*, y) \mid y \in X\}. \tag{1}$$

Proof. Let $x^* \notin X$ and $\bar{B}_r(x^*)$ be the closed ball with radius $r > 0$ and center x^*. Pick r sufficiently large so that $Y = \bar{B}_r(x^*) \cap X \neq \varnothing$. Since $\bar{B}_r(x^*)$ and X both are closed set, so is Y. Since $\bar{B}_r(x^*)$ is bounded, and $Y \subset \bar{B}_r(x^*)$, Y is also bounded. Thus, Y is compact. The distance function $d(x^*, y) : Y \longrightarrow \mathbb{R}_+$ is a continuous function on a compact set Y. Therefore, by Weirstrass' theorem, there exists $y^* \in Y$ such that

$$d(x^*, y^*) = \min\{d(x^*, y) \mid y \in Y\}.$$

If $y \in X \setminus Y$, then we must have $y \notin \bar{B}_r(x^*)$, and hence $d(x^*, y) > r$. Therefore, we have established that $d(x^*, y^*) = \min\{d(x^*, y) \mid y \in X\}$. Finally, we show that y^* is unique. Suppose that there is $z^* \neq y^*$ such that $d(x^*, y^*) = d(x^*, z^*)$. Since X is convex, the point $w^* = (1/2)y^* + (1/2)z^*$ is in X. Then it must be the case that $d(x^*, w^*) < d(x^*, y^*)$, which is a contradiction to (1). □

Definition 6 (Supporting hyperplane) *A hyperplane $H(p, \alpha)$ is a supporting hyperplane for a set X if it contains a point on the boundary of X and the whole set lies on the same side of $H(p, \alpha)$. Equivalently, $H(p, \alpha)$ supports X if*

$$\alpha = \inf\{p \cdot x \text{ for } x \in X\} \text{ or } \alpha = \sup\{p \cdot x \text{ for } x \in X\}.$$

Intuition suggests that a convex set in $\mathbb{R}^n$ should have a supporting hyperplane through each point of its boundary, and also, given two disjoint convex sets, there should be a hyperplane that separates the two sets. The following theorems establish that both of the above intuitions are correct.

Theorem 4 *Let X be a non-empty, closed and convex subset of $\mathbb{R}^n$, and let x^* be a point in $\mathbb{R}^n \setminus X$. Then*

(a) (supporting hyperplane theorem) *there exists a point $y^* \in bd(X)$ and a hyperplane $H(p, \alpha)$ that passes through y^* that supports X and separates X from $\{x^*\}$, i.e., $H(p, \alpha)$ is such that*

$$p \cdot x^* < \alpha = p \cdot y^* = \inf\{p \cdot x \text{ for } x \in X\}.$$

(b) *There exists a second hyperplane $H(p, \beta)$ that separates X strictly from $\{x^*\}$, i.e.,*

$$p \cdot x^* < \beta < p \cdot x, \text{ for all } x \in X.$$

Proof. (a) Let $p = y^* - x^*$ where y^* is given by (1), and $\alpha = p \cdot y^*$. Then $H(p, \alpha)$ passes through y^* that is orthogonal to the line joining y^* and x^*. We claim that this is the desired hyperplane. First, we have

$$p \cdot x^* = p \cdot y^* - p \cdot (y^* - x^*) = \alpha - ||p||^2 < \alpha.$$

So $x^* \in H_L(p, \alpha)$. To show that X lies in $H_U(p, \alpha)$, we proceed by contradiction. Suppose there is a point $y \in X$ such that $p \cdot y < \alpha$, and let

$$x^\lambda = (1-\lambda)y^* + \lambda y, \text{ for } \lambda \in (0, 1).$$

Since X is convex, $x^\lambda \in X$. Now, notice that $x^* - x^\lambda = x^* - y^* - \lambda(y - y^*) = -p - \lambda(y - y^*)$. Therefore,

$$\begin{aligned} ||x^* - x^\lambda||^2 &= ||p||^2 + 2\lambda p \cdot (y - y^*) + \lambda^2 ||y - y^*||^2, \\ \Rightarrow \; ||x^* - y^*||^2 - ||x^* - x^\lambda||^2 &= -\lambda \left[2p \cdot (y - y^*) + \lambda ||y - y^*||^2\right]. \end{aligned}$$

Since by construction $p \cdot y^* = \alpha$ and by assumption $p \cdot y < \alpha$, we have $p \cdot (y - y^*) < 0$. Hence, the above equation implies that, for small values of λ, $||x^* - y^*|| - ||x^* - x^\lambda|| > 0$, which is a contradiction since y^* is chosen to minimize the distance $||x^* - y||$ for $y \in X$. (b) The proof of this part is identical (with the same p) to that of (a), except that we now choose β so as to $H(p, \beta)$ passes through the point $(1/2)x^* + (1/2)y^*$. □

The above theorem can easily be generalized to the case when X is a convex set, but not necessarily closed. The following theorem proves the existence of a hyperplane that separates two disjoint convex sets.

Theorem 5 (Minkowski's separating hyperplane theorem) *Let X and Y be two non-empty and disjoint convex sets in $\mathbb{R}^n$. Then there exists a hyperplane $H(p, \alpha) \subset \mathbb{R}^n$ that separates X and Y.*

Proof. Let $Z = X - Y = X + (-1)Y$, which is convex by Lemma 1(b). We claim that $0 \notin Z$. If we had $0 \in Z$, there would exist a point $x \in X$ and $y \in Y$ such that $x - y = 0$. But this implies $x = y$, and so $x \in X \cap Y$, which contradicts the assumption that X and Y are disjoint. Since $0 \notin Z$, by the general version of the previous theorem, there exists $p \in \mathbb{R}^n$ such that, for $z \in Z$, $x \in X$ and $y \in Y$,

$$0 = p \cdot 0 \le p \cdot z = p \cdot (x - y) \;\Rightarrow\; p \cdot y \le p \cdot x.$$

The set of real numbers of the form $\{p \cdot y \text{ for } y \in Y\}$ is bounded above by the number $p \cdot x$, and hence has a supremum, which we call α. Thus, we have that

$$p \cdot y \le \alpha \le p \cdot x, \text{ for } x \in X \text{ and } y \in Y.$$

Therefore, $H(p, \alpha)$ separates X and Y. □

Theorem 5 is used to prove the so-called *second theorem of welfare economics*. Let us introduce some necessary definitions prior to proving the theorem. The production plan of a firm is described by its production set $Y := \{y \in \mathbb{R}^n \mid F(y) \le 0\}$, where $F(\cdot)$ is called the transformation function. Obviously, $bd(Y) = \{y \in \mathbb{R}^n \mid F(y) = 0\}$. A production $y \in Y$ is *profit-maximizing* for some price vector $p \in \mathbb{R}^n_{++}$ if $p \cdot y \ge p \cdot y'$ for all $y' \in Y$. A production $y \in Y$ is *efficient* if there is no $y' \in Y$ such that $y' \succsim y$ and $y' \ne y$. It is obvious that every efficient y must lie on $bd(Y)$, but the converse is not necessarily true. Now we prove the following important theorem.

Theorem 6 (Second theorem of welfare economics) *Suppose the production set Y is convex. Then every production $y^* \in Eff(Y)$ is a profit-maximizing production for some price vector $p \in \mathbb{R}^n_+$.*

Proof. Suppose that $y^* \in Eff(Y)$, and defined the set $P_{y^*} := \{y' \in \mathbb{R}^n \mid y' > y^*\}$. The set P_{y^*} is convex, and because y is efficient, $Y \cap P_{y^*} = \varnothing$. Then by Theorem 5, there is a non-zero price vector such that that $p \cdot y' \geq p \cdot y$ for every $y' \in P_{y^*}$ and $y \in Y$. Note, in particular, that this implies $p \cdot y' \geq p \cdot y^*$ with $y' > y^*$. We must have $p \in \mathbb{R}^n_+$ because if $p_i < 0$ for some i, then we would have $p \cdot y' < p \cdot y$ for some y such that $y < y'$ with $y'_i - y_i$ sufficiently large. Since $y^* \in bd(Y)$, there is a hyperplane $H(p, \pi)$ that passes through y^*, and supports P_{y^*}, i.e., $\pi = p \cdot y^* = \inf\{p \cdot y' \text{ for } y' \in P_{y^*}\}$. Now take any $y \in Y$. Then $p \cdot y' \geq p \cdot y$ for every $y' \in P_y$. Since the set of real numbers of the form $\{p \cdot y \mid y \in Y\}$ is bounded above, by the supremum property we have $p \cdot y \leq \pi = p \cdot y^*$ for all $y \in Y$, i.e., y^* is profit-maximizing at p. □

4 A quick tour of linear algebra

Let $A = \{a^1, \ldots, a^m\}$ denote a finite set of vectors as well as the index set of the vectors.

Definition 7 (Linear combination and span) *A vector b can be expressed as a linear combination of vectors in $A = \{a^1, \ldots, a^m\}$ if there exist real numbers $\{x_j\}_{j \in A}$ such that*

$$b = \sum_{j \in A} x_j a^j.$$

The set of all such vectors that can be expressed as a linear combinations of vectors in A is called the span of A and is denoted $span(A)$.

Definition 8 (Linear independence) *A set $A = \{a^1, \ldots, a^m\}$ of vectors is linearly independent if for all sets of real numbers $\{x_j\}_{j \in A}$*

$$\sum_{j \in A} x_j a^j = 0 \quad \Longrightarrow \quad x_j = 0 \text{ for all } j \in A.$$

If the vectors in A are linearly dependent, then it is the case that there exist real numbers $\{x_j\}_{j \in A}$ not all zero such that

$$\sum_{j \in A} x_j a^j = 0.$$

Definition 9 (Rank) *The rank of a set A of vectors, denoted $\rho(A)$, is the size of the largest subset of linearly independent vectors in A.*

Example 4 The set $A = \{(0, 1, 0), (-2, 2, 0)\}$ is a set of linearly independent vectors, whereas the vectors in $A' = \{(0, 1, 0), (-2, 2, 0), (-2, 3, 0)\}$ are linearly dependent. Notice that $\rho(A) = \rho(A') = 2$. ■

It is also true that, given a set A of vectors, the dimension of $span(A)$ is equal to its rank. Let $A = \{(0, 1), (1, 0)\}$. Then $span(A) = \mathbb{R}^2$, and $dim[span(A)] = \rho(A) = 2$.

Definition 10 (Finite cone) *Given an $m \times n$ matrix A, the set of all non-negative linear combinations of the columns of A is called the finite cone generated by the columns of A. Formally, such cone is given by:*

$$cone(A) = \{y \in \mathbb{R}^m \mid y = Ax \text{ for some } x \in \mathbb{R}^n_+\}.$$

Notice the difference between $cone(A)$ and $span(A)$, which is given by:

$$span(A) = \{y \in \mathbb{R}^m \mid y = Ax \text{ for some } x \in \mathbb{R}^n\}.$$

For example, let the columns of a matrix $A_{2\times 2}$ be the vectors $(1,0)$ and $(0,1)$. Then $span(A) = \mathbb{R}^2$, whereas $cone(A)$ is the non-negative orthant. The following lemma states a useful property of finitely generated cones.

Lemma 3 *Let A be an $m \times n$ matrix, then $cone(A)$ is a closed convex set.*

We will skip the proof of the above lemma. The convexity of $cone(A)$ is easy to show, whereas the proof of closedness is a bit more complicated. Interested readers should refer to Vohra (2005).

5 The theorems of the alternative

Let us start with the problems of the following kind:

> Given a matrix $A_{m\times n}$ and $b \in \mathbb{R}^m$, find an $x \in \mathbb{R}^n$ such that $Ax = b$ or prove that no such x exists.

The above kind of problems motivates the *fundamental theorem of linear algebra* which is stated below.

Theorem 7 (Fundamental theorem of linear algebra) *Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$, $F = \{x \in \mathbb{R}^n \mid Ax = b\}$, and $G = \{y \in \mathbb{R}^m \mid \ yA = 0 \text{ and } yb \neq 0\}$. Then $F \neq \varnothing$ if and only if $G = \varnothing$.*

The proof of the above theorem is omitted. One can provide an easy geometric interpretation. The first alternative is equivalent to the fact that $b \in span(A)$. Let $span(A)$ be a plane is $\mathbb{R}^2$. If $b \notin span(A)$, then b must be a vector that has non-zero inner product ($yb \neq 0$) with a vector y that is orthogonal to the plane, i.e., orthogonal to each of the linearly independent column vectors of A ($yA = 0$).

Next, we modify the above question a bit, and look for the set of non-negative solutions to the system $Ax = b$. The well-known *Farkas lemma*, which can be derived from the separation theorems, analyzes such problems.

Theorem 8 (Farkas lemma) *Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$, $F = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$, and $G = \{y \in \mathbb{R}^m \mid \ yA \geq 0 \text{ and } yb < 0\}$. Then $F \neq \varnothing$ if and only if $G = \varnothing$.*

The proof of the above theorem may be omitted by the students since the graphical intuition is simple enough. Take a matrix A whose columns are $a_1 = (1,1)$ and $a_2 = (0,1)$. Clearly, $span(A) = \mathbb{R}^2$ and $cone(A)$ is triangle shaped area between the vectors $(1,1)$ and $(0,1)$. The first alternative implies that vector $b = (b_1, b_2)$ lies inside the triangular area. To see the intuition behind the second alternative, let $b = (1,0) \notin cone(A)$. We need to find a vector $y \in \mathbb{R}^2$ with the desired property. Let $y = (1,-1)$. Notice that $y^T \cdot a_1 = (1,-1)^T \cdot (1,1) = 0$ and $y^T \cdot a_2 = (1,-1)^T \cdot (1,0) = 1$, and hence $yA \in [0,1]$. Also, $y^T \cdot b = (1,-1)^T \cdot (0,1) = -1 < 0$. The idea is to make use of the separation theorem. Since $cone(A)$ is closed and convex, and $b \notin cone(A)$, then there exists a hyperplane that strictly separates b from $cone(A)$.

Convince yourselves that $z_2 = 1/2 + z_1$ where $(z_1, z_2) \in \mathbb{R}^2$ is such a hyperplane, and it is orthogonal to y. The set G are called the Farkas alternative of F.

The following theorem is a generalization of Farkas lemma.

Theorem 9 *Let $A_{m\times n}$, $B_{m\times t}$, $C_{k\times n}$ and $D_{k\times t}$ are matrices, and $b \in \mathbb{R}^m$ and $d \in \mathbb{R}^k$ are vectors. Define*

$$F := \{x \in \mathbb{R}^n, x' \in \mathbb{R}^t \mid Ax + Bx' = b,\ Cx + Dx' \le d, x \ge 0\},$$
$$G := \{y \in \mathbb{R}^m, y' \in \mathbb{R}^k \mid\ yA + y'C \ge 0,\ yB + y'D = 0,\ yb + y'd < 0 \text{ and } y' \ge 0\}$$

Then $F \ne \varnothing$ if and only if $G = \varnothing$.

Proof. Consider the system of inequalities $Cx + Dx' \le d$. This can be converted to a system of equations by introducing slack variables for every inequality, i.e., there exists $s \in \mathbb{R}^k_+$ such that

$$Cx + Dx' + s = d.$$

Next, the vector $x' \in \mathbb{R}^t$ can be written as $x' = z - z'$ where $z, z' \in \mathbb{R}^t$. This is because any real number x'_j can be written as the difference of two non-negative real numbers z_j and z'_j. So, F is given by:

$$F := \{x \in \mathbb{R}^n, z, z' \in \mathbb{R}^t, s \in \mathbb{R}^k \mid Ax + Bz - Bz' + 0 \cdot s = b,\ Cx + Dz - Dz' + Is = d, x, z, z', s \ge 0\},$$

where $I_{k\times k}$ is the identity matrix. In matrix notation, the above system looks

$$\begin{bmatrix} A & B & -B & 0 \\ C & D & -D & I \end{bmatrix} \begin{bmatrix} x \\ z \\ z' \\ s \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix} \qquad \Longleftrightarrow \qquad Qr = w.$$

The Farkas alternative of F is given by:

$$G := \{\tilde{y} \in \mathbb{R}^{m+k} \mid \tilde{y}Q \ge 0 \text{ and } \tilde{y}w < 0\} \quad \text{where} \quad \tilde{y} = \begin{bmatrix} y \\ y' \end{bmatrix}$$

Notice that $\tilde{y}w < 0$ is equivalent to $yb + y'd < 0$. And $\tilde{y}Q \ge 0$ is equivalent to the following system of inequalities:

$$\begin{aligned} & yA + y'C \ge 0, \\ & yB + y'D \ge 0, \\ & -yB - y'D \ge 0, \\ & y \cdot 0 + y'I \ge 0. \end{aligned}$$

The second and the third inequalities together imply $yB + y'D = 0$, and the last inequality implies $y' \ge 0$. Now apply Farkas lemma to complete the proof. ☐

Definition 11 (Linear and affine functions) *Let X be a convex subset of $\mathbb{R}^n$. A function $L : X \to \mathbb{R}$ is linear if (a) $L(x+y) = L(x) + L(y)$ for any vectors $x, y \in X$, and (b) $L(\alpha x) = \alpha L(x)$ for any vector $x \in X$ and any scalar α. A function $f : X \to \mathbb{R}$ is an affine function if there is a linear function $L : X \to \mathbb{R}$ and a real number b such that $f(x) = L(x) + b$ for all $x \in X$.*

Convince yourself that all linear functions are affine functions, but the converse is not true. The following theorem is another theorem of the alternatives, which is very useful in establishing the existence of the Lagrange multipliers for constrained optimization problems.

Theorem 10 (Gordan's theorem) *Let $f_i : X \longrightarrow \mathbb{R}$ for $i = 1, \ldots, m$ be affine functions, X be a convex subset of $\mathbb{R}^n$. Define*

$$F := \{x \in \mathbb{R}^n \mid f_i(x) < 0 \text{ for } i = 1, \ldots, m\},$$
$$G := \{y \in \mathbb{R}^m_+, y_i > 0 \text{ for at least one } i \mid y \cdot f(x) \geq 0 \text{ for all } x \in \mathbb{R}^n\},$$

where $f(x) = (f_1(x), \ldots, f_m(x))$. Then $F \neq \varnothing$ if and only if $G = \varnothing$.

Proof. We first show that if (a) holds then (b) cannot hold. Given any x satisfying (a) and any multipliers $y_1 \geq 0, \ldots, y_m \geq 0$, each term of the expression $\sum_{i=1}^m y_i f_i(x)$ is non-positive. Terms for which $y_i \neq 0$ are actually negative, so that the above expression must be negative if the multipliers are not all zero. Therefore (b) cannot hold.

Assume now that (a) does not hold. We must show that in this case (b) holds. Let

$$S = \{z \in \mathbb{R}^m \mid \text{there exists } x \in X \text{ such that } z_i > f_i(x) \text{ for all } i = 1, \ldots, m\}.$$

The above set is clearly non-empty since each z_i can be written as $z_i = f_i(x) + \alpha$ for $\alpha > 0$. Next, we show that S is a convex set. Let z' and z'' be two elements of S, and define $z^\mu = \mu z' + (1-\mu)z''$ and $x^\mu = \mu x' + (1-\mu)x''$ for $\mu \in (0, 1)$. The point z' is in S implies that there exists some x' in X such that $z'_i > f_i(x')$ for each $i = 1, \ldots, m$. Similarly, z'' is in S implies that there exists some x'' in X such that $z''_i > f_i(x'')$ for each $i = 1, \ldots, m$. Hence, it follows that

$$z_i^\mu = \mu z'_i + (1-\mu)z''_i > \mu f_i(x') + (1-\mu) f_i(x'') = f_i(x^\mu), \text{ for each } i = 1, \ldots, m.$$

Notice that $x^\mu \in X$ since X is convex, and the last equality holds because f_i is an affine function for each i. Thus, we have shown that that there exists some x $(= x^\mu)$ in X such that $z^\mu > f_i(x)$ for each i, and hence z^μ is in S. Next notice that, if (a) does not hold, then $0 \notin S$. If 0 were in S, then for all $i = 1, \ldots, m$ we would have $f_i(x) < 0$, contradicting the hypothesis that (a) does not hold good. So, by the Separating Hyperplane theorem we have that there exists a non-zero vector $y = (y_1, \ldots, y_m)$ such that

$$0 \leq y \cdot z = \sum_{i=1}^m y_i z_i, \text{ for all } z \in S.$$

Fix an $x^0 \in X$ and set $w^0 = (f_1(x^0), \ldots, f_m(x^0))$. Then for any $q \in \mathbb{R}^m_{++}$ and $r > 0$ we have that $w^0 + rq \in S$, and hence

$$y \cdot (w^0 + rq) \geq 0 \;\Rightarrow\; y \cdot q \geq -\frac{1}{r}(y \cdot w^0).$$

Taking the limit of the above expression as $r \to \infty$ and using the fact that $y \cdot q$ is continuous, we have that

$$y \cdot q \geq 0, \text{ for all } q \in \mathbb{R}^m_{++}$$

The above implies that $y_i \geq 0$ for all $i = 1, \ldots, m$. Now consider any $x \in X$, and any fixed $q \in \mathbb{R}^m_{++}$ and $\varepsilon > 0$. Thus we have that

$$y \cdot (f(x) + \varepsilon q) \geq 0.$$

Taking the limit of the above expression as $\varepsilon \to 0$, we get the desired result. □

6 Applications

Farkas lemma has been applied to many problems in economics and finance. In what follows we discuss two useful applications.

6.1 No arbitrage in financial markets

Suppose there are m financial assets and n states of nature. Let a_{ij} be the payoff from one share of asset i in state j. Thus the returns of m assets will be represented by a matrix $A_{m\times n}$. A portfolio of assets is represented by a vector $y \in \mathbb{R}^m$ where the i-th component, y_i represents the portfolio weight of asset i. Let $w \in \mathbb{R}^n$ be a vector whose j-th component denotes wealth in state j. We assume that wealth in a future state is related to the current portfolio by:

$$w_j = \sum_{i=1}^{m} y_i a_{ij} \quad \text{for } j = 1, \ldots, n \quad \text{with} \quad \sum_{i=1}^{m} y_i = 1.$$

Therefore, $w = Ay$. This assumes that assets are infinitely divisible, and the returns are linear in the quantities held.

The no arbitrage condition asserts that a portfolio that pays off non-negative amounts in every state must have a non-negative price. If $p \in \mathbb{R}^m_{++}$ is a vector of asset prices, the no arbitrage condition boils down to

$$yA \geq 0 \implies y \cdot p \geq 0.$$

Equivalently, the system $yA \geq 0$ and $y \cdot p < 0$ has no solution. From Farkas lemma we deduce the existence of vector $\hat{\pi} \in \mathbb{R}^n_+$ such that $p = A\hat{\pi}$. Since $p > 0$, it follows that $\hat{\pi} > 0$. Scale $\hat{\pi}$ by dividing through $\sum_j \hat{\pi}_j$. Let $p^* = p/\sum_j \hat{\pi}_j$ and $\pi = \hat{\pi}/\sum_j \hat{\pi}_j$. Notice that π is a probability vector. As long as relative prices are all that matter such scaling can be done without loss of generality. Now we have $p^* = A\pi$, i.e., there is a probability distribution under which the expected return of every asset is equal to its price. The probabilities are called the risk-neutral probabilities.

6.2 Core of a coalitional game

A coalitional game is denoted by (N, v) where $N = \{1, \ldots, n\}$ is the set of players, and $v : 2^N \to \mathbb{R}$, called the *characteristic function*, which assigns to each subset S of N a real number $v(S)$. The subset S is called a *coalition*. A vector $x \in \mathbb{R}^n$ is called a *feasible allocation* if $x_i \geq v(\{i\})$ and $\sum_{i=1}^{n} x_i = v(N)$. Let $I(N, v)$ denote the set of feasible allocations or *imputations* of the game (N, v).

Consider an economy with two buyers 1 and 2, and one seller s, i.e., $N = \{1, 2, s\}$. The seller has one item, say a car, and the buyers have nothing. We are interested in a feasible trade of this item. Suppose that the seller has no value for the object. Buyer 1 derives a utility of 5, and buyer 2 derives a utility of 10 form the object, respectively. Let S be a coalition. Note that there are $2^3 = 8$ possible coalitions, namely, $\varnothing$, $\{1\}$, $\{2\}$, $\{s\}$, $\{1, 2\}$, $\{1, s\}$, $\{2, s\}$ and $\{1, 2, s\}$. The characteristic function $v(S)$, which denotes the "worth" of a coalition S is given by:

$$v(\varnothing) = v(\{1\}) = v(\{2\}) = v(\{s\}) = v(\{1, 2\}) = 0,$$
$$v(\{1, s\}) = 5, \quad \text{and} \quad v(\{2, s\}) = v(\{1, 2, s\}) = 10.$$

We assume that the item is assigned to the highest-valuation buyer, and hence $v(N) = 10$.

The *core* of a game (N, v) is the set

$$C(N, v) := \left\{ x \in I(N, v) : \sum_{i \in S} x_i \geq v(S) \text{ for all } S \subseteq N \right\}.$$

Let us first compute the core allocations of the above buyer-seller economy. If $(x_1, x_2, x_s) \in C(N, v)$, then we must have

$$\begin{aligned} &x_1 + x_2 + x_s = 10, \\ &x_1 + x_2 \geq 0, \\ &x_1 + x_s \geq 5, \\ &x_2 + x_s \geq 10, \\ &x_i \geq 0 \quad \text{for } i \in \{1, 2, s\}. \end{aligned}$$

Notice that

$$x_2 + x_s \geq 10 \iff 10 - x_1 \geq 10 \iff x_1 \geq 0,$$

which together with $x_1 \geq 0$ imply $x_1 = 0$. Then $x_s \geq 5$. Similarly,

$$x_1 + x_s \geq 5 \iff 10 - x_2 \geq 5 \iff x_1 \geq 5.$$

Therefore, $x_1 = 0$, $x_2 \in [0, 5]$ and $x_s \in [5, 10]$ with $x_2 + x_s = 10$ constitute a core allocation of the buyer-seller game.

Unfortunately, not all coalitional games have non-empty cores. For example, consider (N, v) where $N = \{1, 2\}$, and $v(\varnothing) = 0$, $v(\{1\}) = v(\{2\}) = 0.75$ and $v(\{1, 2\}) = 1$. It is easy to see that $C(N, v) = \varnothing$. Therefore, we will look for conditions under which the core of a coalitional game is non-empty.

Definition 12 (Balanced collection) *Let $\mathscr{C} \subseteq 2^N \setminus \varnothing$ be a collection of non-empty coalitions of N. The collection $\mathscr{C}$ is balanced if there exist numbers $y_S \geq 0$ for $S \in \mathscr{C}$ such that for every player $i \in N$ we have*

$$\sum_{S: i \in S} y_S = 1. \tag{2}$$

The numbers $\{y_S \mid S \in \mathscr{C}\}$ in (2) are known as the balancing coefficients for the collection $\mathscr{C}$.

Let $B(N)$ be the set of feasible solutions to the following system:

$$\begin{aligned} y_S \geq& 0 \quad \text{for all } S \subseteq N, \\ \sum_{S: i \in S} y_S =& 1 \quad \text{for all } i \in N. \end{aligned}$$

The number y_S can be thought of as the weights given to coalition S. It is easy to verify that $B(N) \neq \varnothing$ since we can always find

$$y_S = \begin{cases} 1 & \text{for all } S \text{ with } |S| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

In our buyer-seller economy, the collection of singletons is given by $\mathscr{C} = \{\{1\}, \{2\}, \{s\}\}$. Then, $y_{\{1\}} = y_{\{2\}} = y_{\{s\}} = 1$, and hence $(1, 1, 1)$ is a vector of balancing coefficients. For the collection of coalitions of size 2, the vector of balancing coefficients is $(1/2, 1/2, 1/2)$.

Definition 13 (Balanced game) *A game (N, v) is balanced if*

$$\sum_{S \subseteq N} v(S) y_S \leq v(N)$$

for every balanced collection of weights, i.e., for all $y \in B(N)$.

Let us verify whether the buyer-seller game is balanced. Notice that the only coalitions, besides the grand coalition, having positive worths are $\{1, s\}$ and $\{2, s\}$. So, we need to show that for every $y \in B(N)$, we have

$$\begin{aligned} & v(\{1, s\}) y_{\{1,s\}} + v(\{2, s\}) y_{\{2,s\}} \leq v(N) \\ \iff & 5y_{\{1,s\}} + 10y_{\{2,s\}} \leq 10 \\ \iff & y_{\{1,s\}} + 2y_{\{2,s\}} \leq 2. \end{aligned}$$

Since $y_{\{1,s\}} + y_{\{2,s\}} = 1$, we have from the above that $y_{\{2,s\}} \leq 1$ because $y_{\{1,s\}}, y_{\{2,s\}} \geq 0$. Now,

$$\begin{aligned} & y_{\{1,s\}} + 2y_{\{2,s\}} \leq 2 \\ \iff & y_{\{1,s\}} \leq 2\left[1 - y_{\{2,s\}}\right] = 2y_{\{1,s\}}. \end{aligned}$$

The above is true since $y_{\{1,s\}} \geq 0$, and hence (N, v) is balanced.

Theorem 11 (Bondareva-Shapley) *Given a coalitional game (N, v), $C(N, v) \neq \varnothing$ if and only if (N, v) is balanced.*

Proof. If $x \in C(N, v)$, then

$$\begin{aligned} & \sum_{i \in N} x_i = v(N), \\ & \sum_{i \in S} x_i \geq v(S) \ \forall \ S \subseteq N, \\ & x_i \ \text{free} \ \forall \ i \in N. \end{aligned} \tag{CORE}$$

The Farkas alternative for (CORE) is

$$\begin{aligned} & v(N) y_N - \sum_{S \subseteq N} v(S) y_S < 0, \\ & y_N - \sum_{S : i \in S} y_S = 0 \ \forall \ i \in N, \\ & y_S \geq 0 \ \forall \ S \subseteq N, \\ & y_N \ \text{free}. \end{aligned} \tag{BAL}$$

Now suppose $C(N, v) \neq \varnothing$. Then (CORE) has a solution, and the Farkas alternative (BAL) has no solutions. Consider $y \in B(N)$ and let $y_N = 1$. Therefore, $B(N)$ is the set of solutions to the final two constraints of (BAL). Since (BAL) has no solutions, we must have

$$v(N) y_N - \sum_{S \subseteq N} v(S) y_S \geq 0,$$

i.e., (N, v) is balanced.

To prove the other direction, suppose that (N, v) is balanced, i.e., for all $y \in B(N)$ we have

$$v(N)\, y_N \geq \sum_{S \subseteq N} v(S)\, y_S.$$

We will show that (BAL) has no solutions, and hence (CORE) has a solution. Assume for contradiction that (BAL) has a solution y. Clearly, $y_N \neq 0$, else $y_S = 0$ for all S which would contradict the first inequality of (BAL). Define $y'_S = y_S / y_N$ for all $S \subseteq N$. Since y is a solution to (BAL), we get

$$\begin{aligned} & v(N) < \sum_{S \subseteq N} v(S)\, y'_S, \\ & \sum_{S:\, i \in S} y'_S = 1 \ \ \forall\ i \in N, \\ & y'_S \geq 0 \ \ \forall\ S \subseteq N. \end{aligned}$$

Therefore, $y' \in B(N)$. But the first inequality contradicts the fact that (N, v) is balanced. $\square$