Topic 2: Convex hulls

2.1 The convex hull of a set

2.1.1 Definition *The* ***convex hull*** *of A, denoted $\operatorname{co} A$, is the smallest convex set that includes A. Smallest means in the sense of inclusion: That is, $\operatorname{co} A$ is a convex set that includes A and if C is any convex set that includes A, then $\operatorname{co} A \subset C$.*

Equivalently,

$$\operatorname{co} A = \cap\{C : C \text{ is convex and } A \subset C\}.$$

This definition has a subtlety. It assumes that a smallest convex set that includes A exists. It takes a little work to show that the equivalent definition (in terms of intersection) is the smallest such set. We shall come back to this subtlety in Topic 4.

2.1.2 Definition *The convex hull of a finite set of points is called a* ***polytope****.*

By this definition, the empty set is considered to be a polytope.

2.1.3 Exercise The convex hull of A is the set of all convex combinations of points of A:

$$\operatorname{co} A = \left\{ \sum_{i=1}^{n} \alpha_i x_i : n \geqslant 1;\ \alpha_1, \ldots, \alpha_n \geqslant 0;\ x_1, \ldots, x_n \in A;\ \sum_{i=1}^{n} \alpha_i = 1 \right\}.$$

□

This is not hard, and is revisited in Topic 4, but try it for yourself. The key is to show that the set of convex combinations is itself a convex set. In Theorem 2.4.1 we find a bound on how big n needs to be.

2.1.4 Exercise For sets $A_1, \ldots, A_n$ in a vector space X,

$$\operatorname{co}\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \operatorname{co} A_i.$$

Note that if any A_i is empty, the result is still true, as both sides will be empty.

□

The next result is akin to the statement that if x_n is a linear combination of $x_1, \ldots, x_{n-1}$, then the span of $x_1, \ldots, x_n$ is the just the span of $x_1, \ldots, x_{n-1}$.

2.1.5 Exercise If $x_n \in \operatorname{co}\{x_1, \dots, x_{n-1}\}$, then $\operatorname{co}\{x_1, \dots, x_n\} = \operatorname{co}\{x_1, \dots, x_{n-1}\}$. □

Sample answer: Assume $x_n = \alpha_1 x_1 + \cdots + \alpha_{n-1} x_{n-1}$ is a convex combination, and let $y \in \operatorname{co}\{x_1, \dots, x_n\}$, say

$$y = \sum_{i=1}^{n} \lambda_i x_i = \sum_{i=1}^{n-1} (\lambda_i + \lambda_n \alpha_i) x_i.$$

But observe that $\sum_{i=1}^{n-1} (\lambda_i + \lambda_n \alpha_i) = 1$, so $y \in co\{x_1, \dots, x_{n-1}\}$. ■

2.1.6 Lemma *For nonempty convex sets $C_1, \dots, C_n$ we have:*

1. *The convex hull of the union $\bigcup_{i=1}^n C_i$ satisfies*

$$\operatorname{co}\Big(\bigcup_{i=1}^n C_i\Big) = \Big\{\sum_{i=1}^n \lambda_i x_i : \lambda_i \geqslant 0,\ x_i \in C_i,\ i = 1, \dots, n;\ \sum_{i=1}^n \lambda_i = 1\Big\}.$$

2. *If each C_i is also compact, then $\operatorname{co}\big(\bigcup_{i=1}^n C_i\big)$ is compact.*

Proof: (1.) This is just an exercise in rearranging convex combinations. Soon you will be expected to be able to do this in your sleep.

Let x belong to $\operatorname{co}\Big(\bigcup\limits_{i=1}^n C_i\Big)$. From Exercise 2.1.3, x is a convex combination of elements of $\bigcup\limits_{i=1}^n C_i$. We can number them as

$$x = \sum_{i=1}^n \sum_{j=1}^{k_i} \mu_{ij} x_{ij},$$

where each $\mu_{ij} \geqslant 0$, x_{ij} belongs to C_i and $\sum_{i=1}^n \sum_{j=1}^{k_i} \mu_{ij} = 1$. (This way of writing x need not be unique, but that doesn't matter.) Let $\lambda_i = \sum_{j=1}^{k_i} \mu_{ij}$. If $\lambda_i > 0$, then

$$\sum_{j=1}^{k_i} \frac{\mu_{ij}}{\lambda_i} = 1 \quad \text{and} \quad z_i = \sum_{j=1}^{k_i} \frac{\mu_{ij}}{\lambda_i} x_{ij} \text{ belongs to the convex set } C_i\ .$$

If $\lambda_i = 0$, let z_i be any point in C_i. Then observe that

$$x = \sum_{i=1}^n \sum_{j=1}^{k_i} \mu_{ij} x_{ij} = \sum_{i=1}^n \lambda_i z_i,$$

where each z_i belongs to C_i, each $\lambda_i \geqslant 0$, and $\sum_{i=1}^n \lambda_i = 1$.

(2.) Recall (Definition 1.1.5) that the **unit simplex**

$$\Delta_{n-1} = \Big\{(\lambda_1, \dots, \lambda_n) \in \boldsymbol{R}^{\mathrm{n}}_+ : \sum_{i=1}^n \lambda_i = 1\Big\}$$

is a closed, bounded, convex set. (Thus it is a compact subset of $\boldsymbol{R}^n$.) Define the function

$$f\colon \Delta_{n-1} \times C_1 \times \cdots \times C_n \to X$$

by

$$f\Big((\lambda_1, \dots, \lambda_n), x_1, \dots, x_n\Big) = \sum_{i=1}^n \lambda_i x_i.$$

Then f is continuous, and by part (1)

$$\operatorname{co}\left(\bigcup_{i=1}^n C_i\right) = f(\Delta_{n-1} \times C_1 \times \cdots \times C_n).$$

Since the continuous image of a compact set is compact (Lemma A.7.15), we are done. ■

Since every finite set is compact, we have the following.

2.1.7 Corollary *Every polytope in a tvs is compact.*

2.2 The closed convex hull of a set

2.2.1 Definition *The* ***closed convex hull*** *of A, denoted $\overline{\operatorname{co}}\,A$, is the smallest closed convex set that includes A.*

We shall see in Corollary 5.1.2 that the closed convex hull of A is the closure of the convex hull of A. This is not generally the same as the convex hull of the closure of A.

2.2.2 Exercise Give an example of a closed set A in the plane whose convex hull is not closed. This is also an example where $\overline{(\operatorname{co} A)} \neq \operatorname{co}(\overline{A})$. □

2.3 Digression: Basic nonnegative linear combinations

2.3.1 Definition *Let $A = \{x_1, \dots, x_n\}$ be a nonempty finite set of vectors in a vector space. Let y be a linear combination of vectors in A,*

$$y = \sum_{i=1}^n \lambda_i x_i, \text{ and let } B = \{x_i \in A : \lambda_i \neq 0\}.$$

Then we say that y ***depends*** *on the set B. We say that y is a* ***basic linear combination*** *of A if it depends on a linearly independent subset of A.*

2.3.2 Remark Recall that the empty set is linearly independent, so for $x_1 \in A$, writing $0 = \alpha_1 x_1$ with $\alpha_1 = 0$, which gives 0 as a linear combination of x_1 where $\{i : \alpha_1 > 0\} = \varnothing$, so 0 is a basic linear combination of A.

You should know from linear algebra that every linear combination can be replaced by a basic linear combination. The trick we want is to do it with nonnegative coefficients. The next result is true for general (not necessarily finite dimensional) vector spaces, and it should be more widely known.

2.3.3 Lemma (Nonnegative Basic Linear Combinations) *A nonnegative linear combination of a set of vectors can be replaced by a basic nonnegative linear combination of those vectors.*

That is, if $x_1, \dots, x_n$ *are vectors and* $y = \sum_{i=1}^n \lambda_i x_i$ *where each* λ_i *is nonnegative, then there exist nonnegative* $\beta_1, \dots, \beta_n$ *such that* $y = \sum_{i=1}^n \beta_i x_i$ *and* $\{x_i : \beta_i > 0\}$ *is independent.*

Proof: The case $y = 0$ is handled by Remark 2.3.2. We treat the remaining case by induction on the number of vectors x_i on which the nonzero y depends. So consider the proposition:

> $\mathbb{P}[n]$: A nonzero nonnegative linear combination of not more than n vectors can be replaced by a nonnegative linear combination that depends on a linearly independent subset.

The validity of $\mathbb{P}[1]$ is easy. If $y \neq 0$ and $y = \lambda_1 x_1$, where $\lambda_1 \geqslant 0$, then we must in fact have $\lambda_1 > 0$ and $x_1 \neq 0$. That is, y depends on the linearly independent subset $\{x_1\}$.

We now show that $\mathbb{P}[n-1] \implies \mathbb{P}[n]$. So assume $y = \sum_{i=1}^n \lambda_i x_i$ and that each $\lambda_i > 0$, $i = 1, \dots, n$. If $\{x_1, \dots, x_n\}$ itself is independent, there is nothing to prove, just set $\beta_i = \lambda_i$ for each i. On the other hand, if $\{x_1, \dots, x_n\}$ is dependent, then there exist numbers $\alpha_1, \dots, \alpha_n$, not all zero, such that

$$\sum_{i=1}^n \alpha_i x_i = 0.$$

We may assume that at least one $\alpha_i > 0$, for if not we simply replace each α_i by $-\alpha_i$.

Now consider the following expression

$$\begin{aligned} y &= \sum_{i=1}^n \lambda_i x_i - \gamma \underbrace{\sum_{i=1}^n \alpha_i x_i}_{=0} \\ &= \sum_{i=1}^n \Big(\lambda_i - \gamma\alpha_i\Big) x_i. \end{aligned}$$

When $\gamma = 0$, this reduces to our original expression. Whenever $\gamma > 0$ and $\alpha_i \leqslant 0$, then $\lambda_i - \gamma\alpha_i > 0$, so the only coefficients that we need to worry about are those with $\alpha_i > 0$. We will choose $\gamma > 0$ just large enough so that at least one of the coefficients $\lambda_i - \gamma\alpha_i$ becomes zero and none become negative. Now for $\alpha_i > 0$,

$$\lambda_i - \gamma\alpha_i \geqslant 0 \iff \gamma \leqslant \frac{\lambda_i}{\alpha_i}.$$

Thus by setting

$$\bar{\gamma} = \min\left\{\frac{\lambda_i}{\alpha_i} : \alpha_i > 0\right\}$$

we are assured that

$$\lambda_i - \bar{\gamma}\alpha_i \geqslant 0 \text{ for all } i = 1, \ldots, n \quad \text{and} \quad \lambda_i - \bar{\gamma}\alpha_i = 0 \text{ for at least one } i.$$

Thus

$$y = \sum_{i=1}^{n} \Big(\lambda_i - \bar{\gamma}\alpha_i\Big) x_i$$

expresses y as a linear combination depending on no more than $n-1$ of the x_i's. Thus by the induction hypothesis $\mathbb{P}[n-1]$, we can express y as a linear combination that depends on a linearly independent subset. ∎

2.3.4 Remark The above proof is highly instructive and is typical of the method we shall use in the study of inequalities. We started with two equalities in n variables

$$y = \sum_{i=1}^{n} \lambda_i x_i$$
$$0 = \sum_{i=1}^{n} \alpha_i x_i.$$

We then took a linear combination of the two equalities, namely

$$1y + \gamma 0 = 1 \sum_{i=1}^{n} \lambda_i x_i + \gamma \sum_{i=1}^{n} \alpha_i x_i,$$

where the coefficients 1 and γ were chosen to eliminate one of the variables, thus reducing a system of equalities in n variables to a system in no more than $n-1$ variables. Keep your eyes open for further examples of this technique. (If you want to be pedantic, you might remark as Kuhn [9] did, that we did not really "eliminate" a variable, we just set its coefficient to zero.)

2.4 Carathéodory's Theorem

The first application of Lemma 2.3.3 is Carathéodory's [5] theorem on convex hulls in finite dimensional spaces.

2.4.1 Carathéodory's Convexity Theorem *For an m-dimensional vector space, every vector in the convex hull of a set can be written as a convex combination of at most $m+1$ vectors from the set.*

Proof: Let A be a subset of an m-dimensional space, and let x belong to the convex hull of A. Then we can write x as a convex combination $x = \sum_{i=1}^{n} \lambda_i x_i$ of points x_i belonging to A. For any vector y in X consider the "augmented" vector $\hat{y}$ in $X \times \boldsymbol{R}$ defined by $\hat{y}_j = y_j$ for $j = 1, \dots, m$ and $\hat{y}_{m+1} = 1$. Then it follows that $\hat{x} = \sum_{i=1}^{n} \lambda_i \hat{x}_i$ since $\sum_{i=1}^{n} \lambda_i = 1$. Renumbering if necessary, by Lemma 2.3.3, we can write $\hat{x} = \sum_{i=1}^{k} \alpha_i \hat{x}_i$, where $\hat{x}_1, \dots, \hat{x}_k$ are independent and each $\alpha_i > 0$. Since an independent set in $X \times \boldsymbol{R}$ has at most $m+1$ members, $k \leqslant m+1$. But this reduces to the two equations $x = \sum_{i=1}^{k} \alpha_i \hat{x}_i$ and $1 = \sum_{i=1}^{k} \alpha_i$. In other words, x is a convex combination of $k \leqslant m+1$ vectors of A. ∎

Note that the theorem does not say that each point in $\operatorname{co} A$ is a convex combination of the same $m+1$ points, nor are these points unique. For example, in the plane a square is the convex hull of its four vertices, but any given point will lie in the convex hull of no more than three of the vertices

2.4.2 Remark We shall find the mapping that takes a vector x in X to the vector $\hat{x} = (x, 1)$ in $X \times \boldsymbol{R}$ quite useful. I wish I had a good name for it. We shall also occasionally map x to $\check{x} = (x, -1)$.

2.4.3 Remark Note that Lemma 2.3.3 already guarantees that if $x \in \operatorname{co} A \subset X$, then x can be written as a nonnegative linear combination of at most m points in A, but this linear combination is not guaranteed not be a convex combination. That is, the coefficients are not guaranteed not sum to one.

2.4.4 Corollary *The convex hull of a compact subset of $\boldsymbol{R}^{\mathrm{m}}$ is compact.*

Proof: Let K be compact and define the mapping from $K^{m+1} \times \Delta_m$ (where as you may recall, Δ_m is the unit simplex in $\boldsymbol{R}^{\mathrm{m+1}}$) into $\boldsymbol{R}^{\mathrm{m}}$ by

$$\Big(x_0, \dots, x_m, (\alpha_0, \dots, \alpha_m)\Big) \mapsto \alpha_0 x_0 + \cdots + \alpha_m x_m.$$

By Carathéodory's Theorem its image is the convex hull of K. The mapping is continuous and its domain is compact, so its image is compact (Lemma A.7.15). ∎

The convex hull of a compact set need not be compact in infinite dimensional spaces, see for instance Aliprantis and Border [1, Example 5.34, p. 185].

2.5 Shapley–Folkman Theorem I

A funny thing about the Shapley–Folkman Theorem is that neither Lloyd Shapley nor Jon Folkman published the theorem. Another curiosity is that the term Shapley–Folkman Theorem refers to two related, but on the surface quite different, results. The first mention of the result appears (with attribution) in the second appendix of a paper by Ross Starr [10], who was interested approximating nonconvex sets by convex sets.

Here is the version of the Shapley–Folkman Theorem that is simplest to state. We know (Exercise 2.1.4) that the convex hull of a sum of sets is the sum of their convex hulls. The first Shapley–Folkman Theorem strengthens this result for finite dimensional spaces. It asserts that we can replace the convex hull of a set in the sum by the set itself for all but at most m of the sets, where m is the dimension of the space. This dependence on the dimension of the space is reminiscent of Carathéodory's Theorem 2.4.1, and the proof is a simple variant, due to Lin Zhou [11].

2.5.1 Theorem (Shapley–Folkman I) *Let $A_1, \dots, A_n$ be nonempty subsets of $\boldsymbol{R}^{\mathrm{m}}$, and let*

$$x \in \operatorname{co}(A_1 + \cdots + A_n).$$

Then we may write x as a sum

$$x = x_1 + \cdots + x_n,$$

where for each i,

$$x_i \in \operatorname{co} A_i,$$

and

$$x_i \in A_i \quad \textit{for all but at most } m \textit{ of the sets } A_i \quad .$$

Proof: (Cf. Zhou [11]) Recall from Exercise 2.1.4 that

$$\operatorname{co}(A_1 + \cdots + A_n) = (\operatorname{co} A_1) + \cdots + (\operatorname{co} A_n),$$

so we can write x as a sum

$$x = y_1 + \cdots + y_n, \qquad y_i \in \operatorname{co} A_i,\ i = 1, \dots, n.$$

For each $i = 1, \dots, n$, since $y_i \in \operatorname{co} A_i$, there are vectors $y_{ij} \in A_i$ and coefficients $\lambda_{ij} > 0$, $j = 1, \dots, k_i$ such that for each i,

$$y_i = \lambda_{i1} y_{i1} + \cdots + \lambda_{ik_i} y_{ik_i}$$

where

$$\sum_{j=1}^{k_i} \lambda_{ij} = 1.$$

Thus we may write x as

$$x = (\underbrace{\lambda_{11} y_{11} + \cdots + \lambda_{1k_1} y_{1k_1}}_{y_1}) + \cdots + (\underbrace{\lambda_{n1} y_{n1} + \cdots + \lambda_{nk_n} y_{nk_n}}_{y_n}).$$

In the proof of Carathéodory's Theorem, we embedded each y_{ij} in $\boldsymbol{R}^{\mathrm{m}+1}$ by appending a one in the last coordinate. For this proof, we embed each y_{ij} in

$\boldsymbol{R}^{\mathrm{m+n}}$ by appending the i^{th} coordinate vector in $\boldsymbol{R}^{\mathrm{n}}$, and we embed x in $\boldsymbol{R}^{\mathrm{m+n}}$ by appending the unit vector $\mathbf{1} \in \boldsymbol{R}^{\mathrm{n}}$. Thus we can rewrite the above conditions as

$$\begin{bmatrix} x \\ \mathbf{1} \end{bmatrix} = \left(\lambda_{11} \begin{bmatrix} y_{11} \\ e^1 \end{bmatrix} + \cdots + \lambda_{1k_1} \begin{bmatrix} y_{1k_1} \\ e^1 \end{bmatrix} \right) + \cdots + \left(\lambda_{n1} \begin{bmatrix} y_{n1} \\ e^n \end{bmatrix} + \cdots + \lambda_{nk_n} \begin{bmatrix} y_{nk_n} \\ e^n \end{bmatrix} \right).$$

The first m coordinates simply reflect that

$$x = y_1 + \cdots + y_n.$$

The $m+i^{\text{th}}$ coordinate asserts that $\sum_{j=1}^{k_i} \lambda_{ij} = 1$, so we have $y_i \in \operatorname{co} A_i$. Recalling that k_i is the number of points taken from A_i, we have $k_i \geqslant 1$ for each $i = 1, \ldots, n$.

Using Lemma 2.3.3, we can write $(x, \mathbf{1})$, which is a nonzero vector in $\boldsymbol{R}^{\mathrm{m+n}}$, as a nonnegative linear combination depending on a linearly independent subset of at most $n+m$ of the vectors (y_{ij}, e^i) in $\boldsymbol{R}^{\mathrm{m+n}}$. In other words, we may without loss of generality assume that at most $n+m$ of the coefficients λ_{ij} are nonzero. Since k_i is the number of vectors from A_i used to represent y_i, we see that $\sum_{i=1}^{n} k_i \leqslant n+m$.

Since $\sum_{i=1}^{n} k_i \leqslant n+m$, and each k_i is at least unity, it must be that for no more than m values of $i \in \{1, \ldots, n\}$ can we have $k_i > 1$. For the remaining $n-m$ (or more) values of i, we must in fact have $k_i = 1$, so for these i we have $\lambda_{i1} = 1$, which means $y_i = y_{i1}$ belongs to A_i. This proves the theorem. ∎

2.6 Extreme points, faces, and extreme rays

Many different sets may have the same closed convex hull. In this section we partially characterize the minimal such set—the set of extreme points. In a sense, the extreme points of a convex set characterize all the members.

2.6.1 Definition *Let C be a convex set. An* ***extreme subset*** *of C is a subset E of C with the property that if x belongs to E it cannot be written as a proper convex combination of points of C outside E. That is,*

$$(x \in E \text{ and } x = (1-\lambda)y + \lambda z, \text{ where } 0 < \lambda < 1 \text{ and } y, z \in C) \implies y, z \in E.$$

A point x is an ***extreme point*** *of C if the singleton $\{x\}$ is an extreme set. The set of extreme points of C is denoted $\mathcal{E}(C)$.*

An extreme point of a polyhedron or polytope[1] *is called a* ***vertex****.*

A ***face*** *of C is an extreme subset F of C that is itself convex.*

An ***extreme ray*** *of a cone is a ray that is an extreme set.*

Note that the empty set is an extreme subset of any convex set. Also, x is an extreme point of C if x belongs to C and $x = (1-\lambda)y + \lambda z$ where $0 < \lambda < 1$ and $y, z \in C$ imply $y = z = x$.

Here are some examples.

[1] Later on (Corollaries 26.4.2 and 26.4.4) we shall see that every polytope is a polyhedron, and every polyhedron is the sum of a polytope and finitely generated cone.

- The set of extreme points of the empty set is the empty set. The empty set is a face of every convex set, including itself.
- The extreme points of a closed disk are all the points on its circumference.
- The set $\mathcal{E}(C)$ of extreme points of a convex set C is an extreme subset of C.
- The vertexes of a triangle C are its extreme points. The set of vertexes of C is an extreme subset, but not a face. Each edge of the triangle is a face.
- In $\boldsymbol{R}^{\mathrm{n}}$, the extreme points of a convex polyhedron are its vertexes. All its faces and edges are extreme sets.
- The nonnegative axes are the extreme rays of the usual positive cone in $\boldsymbol{R}^{\mathrm{n}}$.

The following useful property is easy to verify.

2.6.2 Exercise A point x in a convex set C is an extreme point of C if and only if $C \setminus \{x\}$ is a convex set. □

2.6.3 Remark In general, the set of extreme points of a convex set C may be empty. For example, any open convex set has no extreme points.

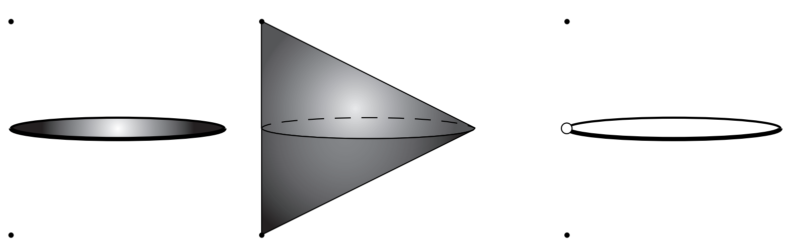
2.6.4 Example As an example of a compact convex set for which the set of extreme points is not closed, consider the subset of $\boldsymbol{R}^3$

$$A = \{(x, y, 0) \in \boldsymbol{R}^3 : x^2 + y^2 \leqslant 1\} \cup \{(0, -1, 1), (0, -1, -1)\}.$$

The convex hull of A is compact, but the set of extreme points of A is

$$\{(x, y, 0) \in \boldsymbol{R}^3 : x^2 + y^2 = 1\} \cup \{(0, -1, 1), (0, -1, -1)\} \setminus \{(0, -1, 0)\},$$

which is not closed. See Figure 2.6.1. □



A co A Extreme points of co A.

Figure 2.6.1. The set of extreme points of co A is not closed.

2.6.5 Exercise You should verify the following properties of extreme subsets.

1. If F is an extreme subset of E and E is an extreme subset of C, then F is an extreme subset of C.

2. The intersection of a collection of extreme subsets of a set C is an extreme subset of C.

□

Unfortunately, in general, the set of extreme points of a convex set need not even be a Borel set; see Bishop and DeLeeuw [4], and Jayne and Rogers [8].

The extreme points of a convex set are of interest primarily because of the Krein–Milman Theorem and its generalizations. The Krein–Milman Theorem asserts that a compact convex subset K of a locally convex Hausdorff space is the closed convex hull of its extreme points. That is, the convex hull of the set of extreme points is dense in K. This means that if every extreme point of K has some property P, and if P is preserved by taking limits and convex combinations, then every point in K also enjoys property P. For instance to show that a compact convex set K lies in the polar of a set A, it is enough to show that every extreme point lies in the polar of A. (This will make sense later. I hope.)

2.6.6 Lemma *The set of maximizers of a convex function is either an extreme set or is empty. Likewise, the set of minimizers of a concave function is either an extreme set or is empty.*

Proof: Let $f\colon C \to \boldsymbol{R}$ be a convex function. Suppose f achieves a maximum on C. Put $\mu = \max\{f(x) : x \in C\}$ and let $M = \{x \in C : f(x) = \mu\}$. Suppose that $x = (1-\alpha)z + \alpha y \in M$, $0 < \alpha < 1$, and $y, z \in C$. If $y \notin M$, then $f(y) < \mu$, so

$$\begin{aligned}\mu = f(x) = f\Big((1-\alpha)y + \alpha z\Big) &\leqslant (1-\alpha)f(y) + \alpha f(z)\\ &< (1-\alpha)\mu + \alpha\mu = \mu,\end{aligned}$$

a contradiction. Hence $y, z \in M$, so M is an extreme subset of C. ∎

We shall see in the Krein–Milman Theorem 19.1.4 that the set of extreme points of a compact convex set determine the set. Proposition 2.6.7 below is a special case of this. Recall that an extreme point of a polytope is called a vertex.

2.6.7 Proposition *Every polytope is the convex hull of the set of its vertexes.*

Proof: (Cf. Gale [7, Theorem 2.17, pp. 68–69]) The case of the empty polytope is trivial, so let F be a nonempty finite set and let $K = \operatorname{co} F$ be a polytope. Now some of the points in F may be redundant, so let

$$B = \{b_1, \ldots, b_n\}$$

be a minimal (in terms of number of elements) subset of F that generates K. That is, $\operatorname{co} B = K$ and no subset A of F with fewer than n elements satisfies $coA = K$. Note that minimality of B and Lemma 2.1.5 imply that

$$\text{no point in } B \text{ is a convex combination of the others.} \tag{M}$$

We want to show that the $\mathcal{E}(K) = B$. Let's start by showing that $B \subset \mathcal{E}(K)$. Suppose, say $b_n \in B$ satisfies

$$b_n = (1-\lambda)y + \lambda z,$$

where $0 < \lambda < 1$ and $y, z \in K$. But then we may write y and z as convex combinations of B:

$$y = \sum_{i=1}^{n} \alpha_i b_i$$
$$z = \sum_{i=1}^{n} \beta_i b_i$$

so

$$b_n = \sum_{i=1}^{n} \Big(\underbrace{(1-\lambda)\alpha_i + \lambda\beta_i}_{=\mu_i}\Big) b_i.$$

If $\mu_n < 1$, then $b_n = \sum_{i=1}^{n-1} \frac{\mu_i}{1-\mu_n} b_i$, violating (M), so $\mu_n = 1$. But then $\alpha_n = \beta_n = 1$, so $y = z = b_n$, proving b_n is extreme.

This also proves the $\mathcal{E}(K)$ is nonempty.

We now show that $\mathcal{E}(K) \subset B$, which is nearly obviously true. So let z be an extreme point of K. Then z must belong to $K = \operatorname{co} B$, so we can write z as a convex combination

$$z = \sum_{i=1}^{n} \lambda_i b_i$$

I claim that at most one $\lambda_i > 0$. For suppose that for more than one i, we have $\lambda_i > 0$. Say $\lambda_1 > 0$ and $\lambda_2 > 0$. Then $1 - \lambda_1 \geqslant \lambda_2 > 0$ and

$$z = \lambda_1 b_1 + (1-\lambda_1) \sum_{i=2}^{n} \tfrac{\lambda_i}{1-\lambda_1} b_i.$$

I claim that b_1 and $\sum_{i=2}^{n} \frac{\lambda_i}{1-\lambda_1} b_i$ are distinct points: For if $b_1 = \sum_{i=2}^{n} \frac{\lambda_i}{1-\lambda_1} b_i$, then we have expressed b_1 as a linear combination of $b_2, \dots, b_n$, contradicting (M).

So if z is an extreme point of K, then exactly one coefficient λ_i is nonzero, in which case, $\lambda_i = 1$ and $z = x_i \in B$. Thus the set of extreme points of K is a subset of B. (Did I say that was obvious?) ∎

References

[1] C. D. Aliprantis and K. C. Border. 2006. *Infinite dimensional analysis: A hitchhiker's guide*, 3d. ed. Berlin: Springer–Verlag.

[2] K. J. Arrow and F. H. Hahn. 1971. *General competitive analysis*. San Francisco: Holden–Day.

[3] Z. Artstein. 1980. Discrete and continuous bang-bang and facial spaces or: Look for the extreme points. *SIAM Review* 22(2):172–185. http://www.jstor.org/stable/2029960

[4] E. Bishop and K. DeLeeuw. 1959. The representation of linear functionals by measures on sets of extreme points. *Annales de l'Institut Fourier (Grenoble)* 9:305–331. http://www.numdam.org/item?id=AIF_1959__9__305_0

[5] C. Carathéodory. 1911. Über den Variabilitätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen. *Rendiconti del Circolo Matematico di Palermo* 32:193–217.

[6] J. W. S. Cassels. 1975. Measures of the non-convexity of sets and the Shapley–Folkman–Starr theorem. *Mathematical Proceedings of the Cambridge Philosophical Society* 78(3):433–436. DOI: 10.1017/S0305004100051884

[7] D. Gale. 1989. *Theory of linear economic models*. Chicago: University of Chicago Press. Reprint of the 1960 edition published by McGraw-Hill.

[8] J. E. Jayne and C. A. Rogers. 1977. The extremal structure of convex sets. *Journal of Functional Analysis* 26:251–288.

[9] H. W. Kuhn. 1956. Solvability and consistency for linear equations and inequalities. *American Mathematical Monthly* 63(4):217–232. http://www.jstor.org/stable/2310345

[10] R. M. Starr. 1969. Quasi-equilibria in markets with non-convex preferences. *Econometrica* 37:25–38. http://www.jstor.org/stable/1909201

[11] L. Zhou. 1993. A simple proof of the Shapley–Folkman theorem. *Economic Theory* 3(2):371–372. DOI: 10.1007/BF01212924