Section 2. Convex Sets

If $x, y \in \mathbb{E}^n$ then $\overline{xy}$ denotes the line segment joining x and y, thus

$$\overline{xy} = \{\lambda x + (1-\lambda)y : 0 \leq \lambda \leq 1\}.$$

A set S is *star shaped relative to the point* $x \in S$ if $\overline{xy} \subset S$ for each $y \in S$. A set S is *convex* if $\overline{xy} \subset S$ for each pair $x, y \in S$. The *kernel* K of a set S is defined by

$$K = \{x \in S : \overline{xy} \subset S \text{ for all } y \in S\}.$$

Theorem 2.6. *The kernel of any set S is a convex set.*

Proof. Let x, y be in the kernel of S. For each $0 \leq \lambda \leq 1$, we must show that $\lambda x + (1-\lambda)y$ is in the kernel of S. To this end, let $z \in S$, then we must show that for each $0 \leq \alpha \leq 1$

$$\alpha\lambda x + \alpha(1-\lambda)y + (1-\alpha)z \in S.$$

We will do this as follows. Find $w \in \overline{xz}$ so that

$$\alpha\lambda x + \alpha(1-\lambda)y + (1-\alpha)z \in \overline{wy}.$$

We know that $w \in S$ since x is in the kernel of S, then we know that $\overline{wy} \in S$ since y is in the kernel of S. We put $\theta = \dfrac{1-\alpha}{1-\alpha(1-\lambda)}$; clearly $0 \leq \theta \leq 1$ and

$$1 - \theta = \frac{\alpha\lambda}{1-\alpha(1-\lambda)}.$$

If we put $w = \theta z + (1-\theta)x$ then $w \in \overline{xz}$. Furthermore, $0 \leq \alpha(1-\lambda) \leq 1$ and

$$[1-\alpha(1-\lambda)]w + \alpha(1-\lambda)y = \alpha\lambda x + \alpha(1-\lambda)y + (1-\alpha)z$$

as required. $\square$

A set S is said to be *affine* if $x, y \in S$ implies that $\lambda x + (1-\lambda)y \in S$, for all real numbers λ. Notice that this means that the whole line containing x and y lies in S.

Theorem 2.13. *A set S in $\mathbb{E}^n$ is affine if and only if it is a translate of a subspace of $\mathbb{E}^n$.*

Proof. Assume S is affine and $x \in S$. We put $U = -x + S$, and aim to show that U is a subspace of $\mathbb{E}^n$. Clearly, the origin is in U. Now let $u_1, u_2 \in U$ and let α, β be any real numbers; it will suffice to prove that $\alpha u_1 + \beta u_2 \in U$. First note that if $u \in U$ then $-u \in U$. To see this, observe that there is an $s \in S$ such that $u = s - x$. Since S is affine, we know that $-s + 2x \in S$ and therefore $-u = -s + x = (-s + 2x) - x \in U$.

Consequently, we can assume, without loss of generality, that $\alpha + \beta \neq 0$. We choose $s_1, s_2 \in S$ such that

$$u_1 = s_1 - x \qquad \text{and} \qquad u_2 = s_2 - x.$$

Then

$$\frac{\alpha s_1}{\alpha+\beta} + \frac{\beta s_2}{\alpha+\beta} \in S$$

and so

$$\alpha u_1 + \beta u_2 = (\alpha+\beta)\left(\frac{\alpha s_1}{\alpha+\beta} + \frac{\beta s_2}{\alpha+\beta}\right) + (1-\alpha-\beta)x - x \in U.$$

Conversely, assume S is a translate of a subspace U of $\mathbb{E}^n$; thus $S = U + t$. If $x, y \in S$ then there are $u, v \in U$ such that

$$x = u + t \qquad y = v + t.$$

If $\lambda \in \mathbb{R}$, we have

$$\lambda x + (1-\lambda)y = \lambda u + (1-\lambda)v + t \in S$$

as required. □

In view of the above theorem, affine sets are often called *flats*, affine subspaces, or linear varieties. The dimension of a flat is the dimension of the parallel subspace. A flat of dimension 1 is called a *line* and a flat of dimension $n-1$ is called a *hyperplane*. Notice that points are zero dimensional flats. The following results are immediate consequences of the definitions of convex sets and affine sets:

a) every intersection of convex sets is convex;
b) every intersection of affine sets is affine.

For any set S, $\operatorname{conv} S$ is the intersection of all convex sets containing S and $\operatorname{aff} S$ is the intersection of all affine sets containing S. These are referred to as the *convex hull* and the *affine hull* of S. The interior of S relative to $\operatorname{aff} S$ is called the *relative interior* of S, and is denoted $\operatorname{relint} S$. For example, if S is a single point $S = \{x\}$ then $\operatorname{relint} S = \{x\}$. If S is the segment $\overline{xy}$ then

$$\operatorname{relint} S = \{\lambda x + (1-\lambda)y : 0 < \lambda < 1\}$$

Theorem 2.9. *Let C be a convex set in $\mathbb{E}^n$. If $x \in \operatorname{int} C$ and $y \in C$ then $\operatorname{relint} \overline{xy} \subset \operatorname{int} C$.*

Proof. First choose $\delta > 0$ so that $B(x,\delta) \subset C$, and let $w \in \operatorname{relint} \overline{xy}$. There is a $0 < \lambda < 1$ with

$$w = \lambda x + (1-\lambda)y.$$

We claim $B(w, \lambda\delta) \subset C$ and so $w \in \operatorname{int} C$, as required. To prove that $B(w,\lambda\delta) \subset C$, let $v \in B(w, \lambda\delta)$. Then $v = w + ru$ where $0 \le r < \lambda\delta$ and $\|u\| = 1$. Thus

$$v = \lambda x + (1-\lambda)y + ru = \lambda(x + \frac{r}{\lambda}u) + (1-\lambda)y.$$

Now $x + \frac{r}{\lambda}u \in B(x,\delta) \subset C$ and so $v \in C$, as required. □

Corollary 2.10. *If C is convex, then so is $\operatorname{int} C$.*

Theorem 2.11. *If C is convex, then so is* $\mathrm{cl}\, C$.

Proof. First note that, if $x \in \mathrm{cl}\, C$ and $\delta > 0$, there is an $x_0 \in B(x, \delta) \cap C$. If this were not the case then we would have $B(x, \delta) \subset \sim C$, in which case $x \in \mathrm{int}(\sim C)$. The latter is impossible because $x \in \mathrm{cl}\, C \subset \sim \mathrm{int}(\sim C)$. Therefore, for each $\delta > 0$, we can choose $x_0, y_0 \in C$ with $x_0 \in B(x, \delta)$ and $y_0 \in B(y, \delta)$. If $z = \lambda x + (1-\lambda) y$ then $z_0 = \lambda x_0 + (1-\lambda) y_0 \in B(z, \delta) \cap C$ since

$$\begin{aligned}
d(z, z_0) &= d(\lambda x_0 + (1-\lambda) y_0, \lambda x + (1-\lambda) y) \\
&\leq d(\lambda x_0 + (1-\lambda) y_0, \lambda x + (1-\lambda) y_0) \\
&\qquad + d(\lambda x + (1-\lambda) y_0, \lambda x + (1-\lambda) y) \\
&= d(\lambda x_0, \lambda x) + d((1-\lambda) y_0, (1-\lambda) y) \\
&= \lambda d(x_0, x) + (1-\lambda) d(y_0, y) \\
&\leq \lambda \delta + (1-\lambda) \delta = \delta.
\end{aligned}$$

Consequently, for every $\delta > 0$, $B(z, \delta) \cap C \neq \phi$. This implies that $z \in \mathrm{cl}\, C$ because, if not, there is a closed set $F \supset C$ with $z \notin F$. But then there is a $\delta > 0$ such that $B(z, \delta) \subset \sim F \subset \sim C$; since F is open. Thus $z \in \mathrm{cl}\, C$ and the proof is complete. □

If $\lambda_1, \ldots, \lambda_k$ are real numbers such that $\lambda_1 + \cdots + \lambda_k = 1$, then the point

$$y = \lambda_1 x_1 + \cdots + \lambda_k x_k$$

is called an *affine combination* of the points $x_1, \ldots, x_k$. If we also have $\lambda_i \geq 0$ for $i = 1, \ldots, k$ then y is called a *convex combination* of $x_1, \ldots, x_k$.

Theorem 2.15. *A set S is convex if and only if every convex combination of points of S lies in S.*

Theorem 2.16. *A set S is affine if and only if every affine combination of points of S lies in S.*

Proof of Theorems 2.15 and 2.16. If every convex(affine) combination of points of S lies in S then S is convex(affine). This is true since the definition of convex(affine) set just requires that every convex(affine) combination of two points of S lies in S.

For the converse, assume S is convex(affine) then, as above, every convex(affine) combination of two points lies in S. We now proceed by induction on k. We assume that the statement is true for all combinations of k points of S and let $x_1, \ldots, x_k \in S$. We assume

$$x = \lambda_1 x_1 + \cdots + \lambda_{k+1} x_{k+1}$$

is a convex(affine) combination. Clearly, there is a λ_i with $\lambda_i \neq 1$. If necessary relabel the points so that $\lambda_{k+1} \neq 1$, in which case $\lambda_1 + \cdots + \lambda_k \neq 0$. Then

$$x = (\lambda_1 + \cdots + \lambda_k) \left(\frac{\lambda_1}{\lambda_1 + \cdots + \lambda_k} x_1 + \cdots + \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_k} x_k \right) + \lambda_{k+1} x_{k+1}.$$

The inductive hypothesis implies that

$$y = \frac{\lambda_1}{\lambda_1 + \cdots + \lambda_k} x_1 + \cdots + \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_k} x_k \in S.$$

But then

$$x = (1 - \lambda_{k+1})y + \lambda_{k+1}x_{k+1} \in S$$

as required. □ □

The set $x_1, \ldots, x_k$ of points is said to be *affinely dependent* if there are real numbers $\lambda_1, \ldots, \lambda_k$ (not all zero) such that

$$\lambda_1 x_1 + \cdots + \lambda_k x_k = o \qquad \text{and} \qquad \lambda_1 + \cdots + \lambda_k = 0.$$

Otherwise the points are said to be *affinely independent.*

Theorem 2.18. *Any $n+1$ points in $\mathbb{E}^n$ are linearly dependent. Any $n+2$ points in $\mathbb{E}^n$ are affinely dependent.*

Proof. We recall from linear algebra that the dimension of a vector space is the maximum number of linearly independent vectors. So the first statement is immediate. For the second, let $x_1, \ldots, x_{n+2}$ be $n+2$ points in $\mathbb{E}^n$. Then $x_2 - x_1, x_3 - x_1, \ldots, x_{n+2} - x_1$ forms a set of $n+1$ vectors in $\mathbb{E}^n$ and these must be linearly dependent. So there are numbers $\lambda_2, \ldots, \lambda_{n+2}$ not all zero, such that

$$\lambda_2(x_2 - x_1) + \lambda_3(x_3 - x_1) + \cdots + \lambda_{n+2}(x_{n+2} - x_1) = O.$$

Hence if

$$\lambda_1 = -(\lambda_2 + \cdots + \lambda_{n+2})$$

then

$$\lambda_1 + \lambda_2 + \cdots + \lambda_{n+2} = 0$$

and

$$\lambda_1 x_1 + \cdots + \lambda_{n+2}x_{n+2} = o.$$

□

Theorem 2.22. *For any set S,* $\operatorname{conv} S(\operatorname{aff} S)$ *is the set of all convex(affine) combinations of elements of S.*

Proof. We denote by T the set of all convex(affine) combinations of points of S. We aim to show that $T = \operatorname{conv} S(\operatorname{aff} S)$. We know that $\operatorname{conv} S(\operatorname{aff} S)$ is convex(affine) and $S \subset \operatorname{conv} S(\operatorname{aff} S)$. It therefore follows from Theorem 2.15(2.16) that $T \subset \operatorname{conv} S(\operatorname{aff} S)$. Next we prove that T is convex(affine). Let $x, y \in T$ and put

$$z = \lambda x + (1 - \lambda)y \qquad \lambda \in \mathbb{R}.$$

There are points $x_1, \ldots, x_k$, $y_1, \ldots, y_m \in S$ and numbers $\alpha_1, \ldots, \alpha_k$, $\beta_1, \ldots, \beta_m$ with

$$\begin{aligned} x &= \alpha_1 x_1 + \cdots + \alpha_k x_k & y &= \beta_1 y_1 + \cdots + \beta_m y_m \\ \alpha_1 &+ \cdots + \alpha_k = 1 & \beta_1 &+ \cdots + \beta_m = 1. \end{aligned}$$

Hence

$$z = \lambda \sum_{i=1}^{k} \alpha_i x_i + (1-\lambda) \sum_{j=1}^{m} \beta_j y_j.$$

Notice

$$\sum_{i=1}^{k} \lambda \alpha_i + \sum_{j=1}^{m} (1-\lambda)\beta_j = \lambda + (1-\lambda) = 1,$$

and so z is an affine combination of points of S. It follows that T is affine. Clearly $T \supset S$ and so $T \supset \operatorname{aff} S$ which gives $T = \operatorname{aff} S$. For the convex case, we would have $\alpha_i \geq 0$, $\beta_j \geq 0$ for $i = 1, \ldots, k$, $j = 1, \ldots, m$. Then if $0 \leq \lambda \leq 1$ we have $\lambda \alpha_i \geq 0$ and $(1-\lambda)\beta_j \geq 0$. Hence T is convex and so $T = \operatorname{conv} S$, as above. □

Our first major convexity result is the following theorem of Carathéodory. It shows that the above result can be improved to the extent that we need only consider convex combinations of $n+1$ points. We will also see that this is the best possible.

Theorem 2.23 (Carathéodory). *If S is a non-empty subset of $\mathbb{E}^n$ then each point of* $\operatorname{conv} S$ *can be expressed as a convex combination of at most $n+1$ points of S.*

Proof. We assume $x \in \operatorname{conv} S$ and use Theorem 2.22 to choose points $x_1, \ldots, x_k \in S$ and numbers $\lambda_1, \ldots, \lambda_k$ such that $x = \lambda_1 x_1 + \cdots + \lambda_k x_k$ where $\lambda_1 + \cdots + \lambda_k = 1$, $\lambda_i \geq 0$ for all $i = 1, 2, \ldots, k$. Our objective is to show that we can make such a choice with $k \leq n+1$. We will achieve this by showing that if $k \geq n+2$ then we could reduce the number of points $x_1, \ldots, x_k$ chosen.

Note that, if $k \geq n+2$ then Theorem 2.18 shows that the points are affinely dependent. Consequently, we can choose numbers $\alpha_1, \ldots, \alpha_k$, not all zero, such that

$$\alpha_1 x_1 + \cdots + \alpha_k x_k = o \qquad \text{and} \qquad \alpha_1 + \cdots + \alpha_k = 0.$$

Now let t be any number and put

$$\mu_i = \lambda_i + t\alpha_i \qquad \text{for } i = 1, \ldots k.$$

We note that, no matter what the value of t,

$$\mu_1 x_1 + \cdots + \mu_k x_k = \lambda_1 x_1 + \cdots + \lambda_k x_k + t(\alpha_1 x_1 + \cdots + \alpha_k x_k) = x$$

and

$$\mu_1 + \cdots + \mu_k = \lambda_1 + \cdots + \lambda_k + t(\alpha_1 + \cdots + \alpha_k) = 1.$$

Now put

$$t = \max\{-\lambda_i/\alpha_i : \alpha_i > 0, i = 1, \ldots, k\}.$$

We note that there must be an i with $\alpha_i > 0$ and so $t \leq 0$. We also have the following:

a) if $\alpha_i = 0$, then $\mu_i = \lambda_i + t\alpha_i = \lambda_i \geq 0$
b) if $\alpha_i > 0$, then $\mu_i = \lambda_i + t\alpha_i \geq 0$
c) if $\alpha_i < 0$, then $\mu_i = \lambda_i + t\alpha_i \geq \lambda_i \geq 0$

Consequently $\mu_i \geq 0$ for $i = 1, \dots, k$. Furthermore, there is an i such that $\mu_i = 0$; namely the value of i which produces the maximum member of the set

$$\{-\lambda_i/\alpha_i : \alpha_i > 0, 1 \leq i \leq k\}.$$

Thus $x = \mu_1 x_1 + \cdots + \mu_k x_k$, $\mu_i \geq 0$, $\mu_1 + \cdots + \mu_k = 1$ and at least one of the μ_i's is zero. Hence x has been written as a convex combination of at most $k-1$ points of S. This completes the proof. □

The number $n+1$ above cannot be reduced. For example, if S comprises three points in $\mathbb{E}^2$ forming the vertices of a triangle, then the centroid of S cannot be expressed as a convex combination of fewer than three points of S.

The convex hull of a finite set of points is called a *polytope*. If $S = \{x_1, \dots, x_{k+1}\}$ and if $\dim \operatorname{aff} S = k$, then $\operatorname{conv} S$ is called a *k-dimensional simplex* or *k-simplex*. The points $x_1, \dots, x_{k+1}$ are its *vertices*.

Note that a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a triangle, and a 3-simplex is a tetrahedon.

If S is a k-simplex with vertices $x_1, \dots, x_{k+1}$ and if $x \in S$ we know from Theorem 2.22 that there are numbers $\alpha_1, \dots, \alpha_{k+1}$ such that

$$x = \alpha_1 x_1 + \cdots + \alpha_{k+1} x_{k+1} \qquad \alpha_1 + \cdots + \alpha_{k+1} = 1 \qquad \alpha_i \geq 0 \qquad 1 \leq i \leq k+1.$$

We will show that these numbers are uniquely determined by x. Assume that we have

$$x = \beta_1 x_1 + \cdots + \beta_{k+1} x_{k+1} \qquad \beta_1 + \cdots + \beta_{k+1} = 1 \qquad \beta_i \geq 0 \qquad 1 \leq i \leq k+1.$$

We put $\lambda_i = \beta_i - \alpha_i$ for $1 \leq i \leq k+1$ and note that

$$\lambda_1 + \cdots + \lambda_{k+1} = 0.$$

Furthermore,

$$\lambda_1 x_1 + \cdots + \lambda_{k+1} x_{k+1} = o.$$

However, the fact that $\dim \operatorname{aff} S = k$ shows that the $x_1, \dots, x_{k+1}$ are affinely independent. Consequently $\lambda_1 = \lambda_2 = \ldots = \lambda_{k+1} = 0$ and we are finished.

So we have proved

Theorem 2.25. *Let S be a k-simplex in $\mathbb{E}^n$ with vertices $x_1, \dots, x_{k+1}$, then each point of S has a unique representation as a convex combination of the vertices.*

We note that this result is not true in general. For example if $x_1, \dots, x_4$ are the vertices of a square. Then

$$\frac{1}{4}(x_1 + \cdots + x_4) = \frac{1}{2}(x_1 + x_3) = \frac{1}{2}(x_2 + x_4).$$

If $x_1, \dots, x_{k+1}$ are the vertices of a k-simplex and

$$x = \alpha_1 x_1 + \cdots + \alpha_{k+1} x_{k+1} \qquad \alpha_1 + \cdots + \alpha_{k+1} = 1 \qquad \alpha_i \geq 0$$

then the numbers $\alpha_1, \dots, \alpha_{k+1}$ are called the *barycentric coordinates* of x. The *centroid* of a k-simplex is the point with barycentric coordinates

$$\left(\frac{1}{k+1}, \dots, \frac{1}{k+1}\right).$$

Theorem 2.27. *Let S be a k-simplex, then* relint $S \neq \phi$.

Proof. Assume the vertices of S are $x_1, \ldots, x_{k+1}$ and let $V = \operatorname{aff} S - x_{k+1}$ be the subspace parallel to $\operatorname{aff} S$. Let x_0 be the centroid of S. Define $f : V \to \mathbb{E}^k$ by

$$f(\mu_1(x_1 - x_{k+1}) + \cdots + \mu_k(x_k - x_{k+1}) = (\mu_1, \ldots, \mu_k)$$

and note that f is continuous. Now choose $\varepsilon > 0$ so that, if the distance between $(\alpha_1, \ldots, \alpha_k)$ and $(\frac{1}{k+1}, \cdots, \frac{1}{k+1}) \in \mathbb{E}^k$ is less than $\varepsilon > 0$ then the α_i are all positive and $\alpha_1 + \cdots + \alpha_k < 1$. The continuity of f shows there is a $\delta > 0$ such that any point

$$(\alpha_1, \ldots, \alpha_k) \in f(B(x_0 - x_{k+1}, \delta))$$

satisfies the above condition. It follows that the $B(x_0, \delta) \subset S$. Hence $x_0 \in \operatorname{relint} S$. □

Corollary 2.28. *Let S be a k-dimensional convex subset of $\mathbb{E}^n$, then* relint $S \neq \phi$.

Proof. We can choose $k+1$ affinely independent points in S. Their convex hull is a k-simplex T. We have

$$\phi \neq \operatorname{relint} T \subset \operatorname{relint} S.$$

□

Theorem 2.29. *If S is open then so is* conv S.

Proof. Assume $x \in \operatorname{conv} S$. We can choose $x_1, \ldots, x_k \in S$ and numbers $\lambda_i \geq 0$ with $\lambda_1 + \cdots + \lambda_k = 1$ and

$$x = \lambda_1 x_1 + \cdots + \lambda_k x_k.$$

Also there is a number $\delta > 0$ such that

$$B(x, \delta) \subset S \qquad \text{for } i = 1, 2, \ldots, k.$$

If $y \in B(x, \delta)$ then $y = x + ru$ where $0 \leq r < \delta$ and $\|u\| = 1$. Put $y_i = x_i + ru$ for $i = 1, \ldots, k$ note that

$$\lambda_1 y_1 + \cdots + \lambda_k y_k = \lambda_1 x_1 + \cdots + \lambda_k x_k + (\lambda_1 + \cdots + \lambda_k) ru = x + ru = y.$$

Now each $y_i \in S$ and so $y \in \operatorname{conv} S$. Hence $\operatorname{conv} S$ is open. □

We note that the convex hull of a closed set is not necessarily closed. For example if

$$S = \{(x, y) \in \mathbb{E}^2 : x^2 y^2 = 1, x > 0\}$$

then S is closed, but

$$\operatorname{conv} S = \{(x, y) \in \mathbb{E}^2 : x > 0\}$$

which is not closed. However, we have the following theorem.

Theorem 2.30. *If S is compact then so is* $\operatorname{conv} S$ *(in $\mathbb{E}^n$).*

Proof. Let C be that compact subset of $\mathbb{E}^{n+1}$ defined by

$$C = \{(\alpha_1, \dots, \alpha_{n+1}) \in \mathbb{E}^{n+1} : \alpha_1 + \cdots + \alpha_{n+1} = 1 \text{ and } \alpha_i \geq 0, 1 \leq i \leq n+1\}.$$

Note that

$$\mathbb{E}^{n+1} \times \underbrace{\mathbb{E}^n \times \cdots \times \mathbb{E}^n}_{n+1} = \mathbb{E}^{(n+1)^2}.$$

Define $f : \mathbb{E}^{(n+1)^2} \to \mathbb{E}^n$ by

$$f(\underbrace{\alpha_1, \dots, \alpha_{n+1}}_{\in \mathbb{E}^{n+1}}, \underbrace{x_1}_{\in \mathbb{E}^n}, \underbrace{x_2}_{\in \mathbb{E}^n}, \dots, \underbrace{x_{n+1}}_{\in \mathbb{E}^n}) = \alpha_1 x_1 + \alpha_1 x_2 + \cdots + \alpha_{n+1} x_{n+1}.$$

It is clear that f is continuous. Furthermore, if S is compact in $\mathbb{E}^n$ then $C \times \underbrace{S \times \cdots \times S}_{n+1}$ is compact in $\mathbb{E}^{(n+1)^2}$ and so $f(C \times S \times \cdots \times S)$ is compact. But Caratheodory's theorem shows that

$$f(C \times S \times \cdots \times S) = \operatorname{conv} S.$$

□