Some Aspects of Convexity

V S Sunder

After spending over a decade at Indian Statistical Institute (Delhi and Bangalore), V S Sunder is currently at Institute of Mathematical Sciences, Chennai.

Part 2 of this article consisting of the applications of convexity, will appear in a subsequent issue.

Introduction

We aim to discuss one of the basic results in convexity – the so-called *Krein–Milman theorem* (for a brief discussion on convex sets, see *Box* 1). Actually, in view of the potential applications, this is best stated as a result about fairly general infinite-dimensional (topological vector) spaces; thus, this falls under the purview of the area generally referred to as 'functional analysis'. However, we shall content ourselves in these notes with discussing what might be called the 'Euclidean case' of this result, by which we mean the specialisation of this result to the case of finite-dimensional vector spaces.

To start with, recall that various notions of a geometric nature can be neatly re-phrased in purely algebraic terms; for instance, given two points $x = (x_1, \quad, x_n)$, $y = (y_1, \quad, y_n)$ in $\mathbb{R}^n$, a typical point on the line-segment joining x and y is of the form $z = (1-t)x + ty$ with $0 \leq t \leq 1$. (Thus, $z_j = (1-t)x_j + ty_j$, $j = 1, \quad, n$.) The following definition has two advantages: (a) it abstracts the property of a set that says that if it contains two points, then it contains the entire line-segment joining those two points; and (b) it makes sense (and is useful, at that level of generality) in a general vector space, possibly of infinite-dimension, possibly over any field, such as the field of complex numbers, which contains the real numbers.

DEFINITION 1. A subset C of an arbitrary vector space V is called **convex** if

$$x, y \in C,\ 0 \leq t \leq 1 \ \Rightarrow \ (1-t)x + ty \in C$$

EXAMPLE 2. (a) The unit interval $[0,1]$ is a convex subset of $\mathbb{R}$. The only convex sets of $\mathbb{R}$ are intervals

– an interval meaning a finite or infinite interval which may or may not contain either end-point.

(b) The set $\{(x,y) \in I\!R^2 : y \geq x^2\}$ is a convex set in $I\!R^2$ The crucial property that a function – such as $f(x) = x^2$ in this case – needs to satisfy, in order that its 'supergraph' $\{(x,y) : y \geq f(x)\}$ is convex, is that (its domain is an interval and)

$$f((1-t)x+ty) \leq (1-t)f(x) + tf(y)$$

for all $t \in [0,1]$ and all x, y (in the domain of definition of f).

(c) The smallest convex set containing three non-collinear points is the triangle with vertices at those points.

The following proposition is prompted by the last example, and is a source of all examples of convex sets.

PROPOSITION 3. (a) The intersection of any family of convex subsets of a vector space is convex.

(b) Let S be a subset of a vector space V; then there exists a smallest convex set containing S; such a set is necessarily unique, and is denoted by conv(S) and called the **convex hull** of the set S.

(c) conv(S) = $\{\sum_{i=0}^{n} \theta_i x_i : \theta_i \in [0,1]\ \forall 0 \leq i \leq n,\ \sum_{i=0}^{n} \theta_i = 1,\ x_i \in S,\ n = 1, 2, \quad \}$.

Proof: (a) is a consequence of the definition.

(b) Define conv$(S) = \cap\{C : C \in \mathcal{C}\}$ where $\mathcal{C}$ denotes the collection of all convex sets C satisfying $S \subset C \subset V$; note that $V \in \mathcal{C}$, and that hence $\mathcal{C} \neq \emptyset$ and that conv(S) is indeed a convex superset of S which is a subset of any other such convex superset.

(c) It is not hard to verify that the set displayed in (c) is indeed a convex set which contains S, and is necessarily contained in any convex superset of S.

□

An expression, such as $\sum_i \theta_i x_i$, as in Proposition 3(c) is called a **convex combination** (of the vectors $x_0, \cdots, x_n$). The following definition is prompted by an examination of the special role that the vertices of a triangle play in the convex set that 'they span'.

DEFINITION 4. A vector v is said to be an **extreme point** of a convex set C if $v \in C$ and the conditions

$$v = (1-t)x + ty, \ x, y \in C, \ 0 < t < 1$$

admit only the trivial solution $v = x = y$. The set of all extreme points of C is denoted by $\partial_e(C)$.

More generally, a set F is said to be a **face** of the convex set C if F is convex and if the conditions

$$v = (1-t)x + ty, \ v \in F, \ x, y \in C, \ 0 < t < 1$$

can be satisfied only if $x, y \in F$

EXERCISE 5. (1) Verify that the following sets are convex, and determine their extreme points:

$\mathbf{B}^n = \{\mathbf{x} = (x_1, \cdots x_n) \in I\!R^n : ||\mathbf{x}||^2 = \sum_{i=1}^n x_i^2 < 1\}$;

$\overline{\mathbf{B}^n} = \{\mathbf{x} = (x_1, \cdots, x_n) \in I\!R^n : ||\mathbf{x}||^2 = \sum_{i=1}^n x_i^2 \leq 1\}$;

$\mathcal{H}^n = \{\mathbf{x} = (x_1, \cdots, x_{n+1}) \in I\!R^{n+1} : x_{n+1} > 0\}$.

(2) Show that if C is a convex subset of $I\!R^n$, then its closure $\overline{C}$ is also convex. (Recall that the closure of a set $C \subset I\!R^n$ is the set of those points $x \in I\!R^n$ for which it is possible to find a sequence $\{x_k : k = 1, 2 \cdots\}$ in C which converges to x (meaning that $\lim_{k\to\infty} ||x_k - x|| = 0$).)

(3) Prove that if $S \subset I\!R^n$ is any set, then $\overline{\text{conv}}(S)$ is the smallest closed convex set containing S.

(4) Show that the only faces of the set $[0, 1] \subset I\!R$ are $\emptyset$, $\{0\}$, $\{1\}$, $[0, 1]$.

(5) What are the faces of the cube? Generalise.

(6) The **standard** n**-simplex** is the set

$$\Delta_n = \{(x_1, x_2, \quad , x_{n+1}) \in I\!R^{n+1} : x_i \geq 0, \ \sum_{i=1}^{n+1} x_i = 1\}$$

(a) Show that Δ_n is the convex hull of the so-called **standard basis** $\{e_i : 1 \leq i \leq n+1\}$, where e_i is the vector with j-th co-ordinate being given by δ_{ij} (i.e., 1 if $i = j$, and 0 otherwise).

(b) What are the extreme points and the faces of Δ_n?

(7) Verify the following easy consequences of the definition of a face:

(a) An intersection of an arbitrary collection of faces of a convex set C is again a face of C.

(b) If F is a face of a convex set C, and if F_0 is a face of (the convex set) F then F_0 is a face of C.

2. The Krein–Milman Theorem (Euclidean case)

Exercise 5 shows that although the notions of convexity and extreme point are purely 'vector space notions', the behaviour of the topological properties of a set seem to influence the behaviour of the set of its extreme points. A result which could well claim to be the ultimate theorem in this direction is the celebrated Krein–Milman theorem, whose finite-dimensional specialisation we state below:

THEOREM 6. A compact convex subset of $I\!R^n$ is the closure of the convex hull of the set of its extreme points.

The only word in the above statement that has not been explained so far is 'compact', and it will do us well to briefly dwell on this notion and some of the pleasant consequences that this property has.

Of the three sets occurring in Exercise 5 (1), the only compact one is $\overline{\mathbb{B}^n}$ We list some consequences of the assumption of compactness in a 'proposition' for convenience of reference. (We do not prove the proposition here; the interested reader can find a proof in any standard text; see [1], [2] or [3], for instance.)

PROPOSITION 7 Suppose K is a compact set. Then,

(i) if $\{F_i : i \in I\}$ is a collection of non-empty closed subsets of K such that any finite subcollection $\{F_s : s \in S\}$ (where S is some finite subset of I) has a point in common, then the entire family has a point in common; in symbols,

$$\cap_{j=1}^{n} F_{i_j} \neq \emptyset \,\forall\, i_1, \quad , i_n \in I,\ n = 1, 2, \qquad \Rightarrow \cap_{i \in I} F_i \neq \emptyset\,;$$

(ii) any continuous real-valued function f on K is bounded, and attains its bounds (i.e., there exist $a, b \in K$ such that $f(a) \leq f(x) \leq f(b)\ \forall\ x \in K$).

EXERCISE 8. If K is a compact set in $\mathbb{R}^n$, deduce the following consequences of Proposition 7(ii):

(a) if $\{x_k\}$ is a sequence in K and if $\lim_k x_k = x$, then $x \in K$; (hint: if $x \notin K$ the function $f(y) = ||y - x||^{-1}$ would be a continuous function on K which does not attain its minimum;)

(b) there exists $0 < N < \infty$ such that $||x|| \leq N\ \forall\, x \in K$

The last exercise may be paraphrased thus: a compact set in $\mathbb{R}^n$ is necessarily **closed** (i.e., satisfies the property asserted in (a) of the exercise) and **bounded** (i.e., satisfies the property asserted in (b)). It is true, conversely, that every closed, bounded set in $\mathbb{R}^n$ is compact.

Notice that we have so far not said what a compact set is; we shall adopt the prescription that a subset of $\mathbb{R}^n$ is said to be compact precisely when it is closed and bounded, and assume the validity (for a proof, see any standard text such as [2] or [3]) of Proposition 7 for closed and bounded subsets of $\mathbb{R}^n$ and go on with the proof of our (Euclidean) formulation of the Krein–Milman theorem – for which, we need a definition and a simple fact.

DEFINITION 9. (1) The **inner product** of two vectors $x = (x_1, \quad, x_n), y = (y_1, \quad, y_n) \in I\!R^n$ is denoted by the symbol (x, y) and defined by:

$$(x, y) = \sum_{i=1}^{n} x_i y_i$$

The expression (x, x) is normally denoted by $||x||^2$ – see Exercise 5 (1); and the expression $||x||$ is called the **norm** of x, and is seen to be the 'distance' from x to the origin 0.

(2) A **linear function** on $I\!R^n$ is a function $f : I\!R^n \rightarrow I\!R$ of the form $f(x) = \alpha_1 x_1 + \quad + \alpha_n x_n$, where $\alpha_1, \quad, \alpha_n \in I\!R$. (Thus every linear function on $I\!R^n$ has the form $f(x) = f_y(x) = (x, y)$ for a uniquely determined $y \in I\!R^n$.)

It must be noticed that linear functions have (and are characterised by) the property that

$$f(\sum_{i=1}^{m} c_i x_i) = \sum_{i=1}^{m} c_i f(x_i),$$

for arbitrary

$$m = 0, 1, 2, \qquad \{c_1, \quad, c_m\} \subset I\!R, \ \{x_1, \quad, x_m\} \subset I\!R^n$$

PROPOSITION 10. Suppose K is a compact convex set in $I\!R^n$ and suppose $z \notin K$ Then there exists a linear function f on $I\!R^n$ and a real number a such that $f(z) < a < f(x)$ for all $x \in K$

Proof: We first prove the proposition under the (inessential) assumption that $z = 0$. Apply Proposition 7 to the continuous function $g(x) = ||x||$ and the compact set K to find that there exists $x_0 \in K$ such that $||x_0|| \leq ||x||$

for all $x \in K$ Since K is convex, conclude that if $x \in K$ and $t \in [0,1]$, then

$$0 \leq ||tx + (1-t)x_0||^2 - ||x_0||^2$$
$$= ||x_0 + t(x - x_0)||^2 - ||x_0||^2$$
$$= t^2||x - x_0||^2 + 2t(x - x_0, x_0)$$

It follows that

$$(x - x_0, x_0) \geq 0 \ \forall \ x \in K$$

Thus, the linear function $f = f_{x_0}$ satisfies

$$f(x_0) \ \leq \ f(x) \ \forall \ x \in K$$

Thus the proposition is valid with $z = 0$, $f = f_{x_0}$ and $a = \frac{1}{2}||x_0||^2$.

If $z \neq 0$, set $C = K - z = \{c = x - z : x \in K\}$, and if f satisfies $0 = f(0) < a < f(c) \ \forall c \in C$, we then see that $f(z) < a + f(z) < f(x) \ \forall \ x \in K$, thus completing the proof of the lemma.

□

The one final ingredient that is needed for a proof of the Krein–Milman theorem is the celebrated Zorn's lemma. This result's usefulness lies in its allowing us to infer the existence of 'maximal' (or 'minimal') elements. We briefly digress to describe some of the terms occurring in its statement.

Recall that a *partial order* on a set $\mathcal{F}$ is a relation $\leq$ that holds for some pairs of elements of $\mathcal{F}$ (i.e., there exists a subset $R_{\leq} \subset \mathcal{F} \times \mathcal{F}$, defined by $R_{\leq} = \{(x,y) : x \leq y\}$), and such that this relation satisfies the requirements of reflexivity ($x \leq x \ \forall x \in \mathcal{F}$), anti-symmetry ($x, y \in \mathcal{F}$, $x \leq y$ and $y \leq x$ imply $x = y$), and transitivity ($x, y, z \in \mathcal{F}$, $x \leq y$ and $y \leq z$ imply $x \leq z$). The set $\mathcal{F}$, equipped with a specific partial order, is called a **partially ordered set**. An element x_1 (resp., x_0) is said to be a maximal (resp., minimal) element of a partially ordered set $\mathcal{F}$ if $x \in \mathcal{F}$, and $x_1 \leq x$ (resp., $x \leq x_0$) necessarily imply that $x = x_1$ (resp., $x = x_0$).

For instance, the relation of 'inclusion' – i.e., $S \leq T \Leftrightarrow S \subset T$ – defines a partial order on any class of subsets of a given set. (What are the maximal (resp., minimal) elements in this example? It should be noted that being a maximal element does not mean being a 'greatest element' since a maximal element need not be comparable with all other elements.)

We state **Zorn's Lemma** in the form that it will be convenient for our use. (This is equivalent to the usual one that you will find in other texts, in which 'upper bounds' and 'maximal elements' feature, rather than 'lower bounds' and 'minimal elements' as in our formulation.)

Suppose $\mathcal{F}$ is a non-empty partially ordered set with the following property: every totally ordered set in $\mathcal{F}$ (i.e., a subset $\mathcal{C} \subset \mathcal{F}$ such that $x, y \in \mathcal{C} \Rightarrow$ either $x \leq y$ or $y \leq x$) admits a lower bound (i.e., there exists an element $z \in \mathcal{F}$ such that $z \leq x \forall\ x \in \mathcal{C}$). Then $\mathcal{F}$ contains at least one minimal element.

This 'lemma' is actually equivalent to one of the axioms of set theory, and so there is no question of our 'proving' this lemma. (What this means is: one of the 'unproved axioms' of our system of logic is the so-called **axiom of choice** which is the (seemingly self-evident) statement that 'the Cartesian product of an arbitrary family of non-empty sets is non-empty'; and it is true that it is possible – see [1], for instance – to derive the validity of either of the statements 'Zorn's lemma', 'axiom of choice', from the other.)

We shall see a more or less typical application of Zorn's lemma in the following proof.

Proof of Theorem 6: There is nothing to prove if $K = \emptyset$, so we assume that $K \neq \emptyset$. Let $\mathcal{F}$ denote the collection of non-empty closed faces of the compact convex set $K \subset \mathbb{R}^n$ Then $\mathcal{F}$ is partially ordered by inclusion, and it is non-empty since $K \in \mathcal{F}$. We assert that this $\mathcal{F}$ satisfies the conditions of Zorn's lemma. Indeed, suppose $\mathcal{C} = \{\mathcal{F}_\rangle : \rangle \in \mathcal{I}\}$ is a collection of closed non-empty faces of K such that for any $i, j \in I$ it is true that either $F_i \subset F_j$ or $F_j \subset F_i$. Then it is seen that if S is any finite subset of I, we can order its elements as $S = \{i_1, \imath_2, \quad, i_m\}$, in such a way that $F_{i_1} \subset F_{i_2} \subset \quad \subset F_{i_m}$, and in particular, this means that

$$\emptyset \neq F_{i_1} = \cap_{s \in S} F_s$$

Since S was arbitrary, it follows from Proposition 7 that $\cap_{i \in I} F_i \neq \emptyset$. If we set $F_0 = \cap_{i \in I} F_i$, we thus see that F_0 is a non-empty closed subset of K On the other hand, since the intersection of any collection of faces is a face (see Exercise 5(7)(a)), we see that F_0 is actually a lower bound for $\mathcal{C}$, as desired.

Hence, by Zorn's lemma, we may conclude the existence of a minimal element P of $\mathcal{F}$. We wish, next, to be able to conclude that P is actually a singleton. So, suppose, if possible, that P contains two distinct points x, y. Appeal to Proposition 10 to find a linear function f and a real number a such that $f(x) < a < f(y)$. Set $a_0 = \min\{f(z) : z \in P\}$; note that the set $P_0 = \{z \in P : f(z) = a_0\}$ is a non-empty closed face in P; since a face of a face of K is necessarily again a face of K (see Exercise 5(7)(b)) it is seen that P_0 is a non-empty closed face of K, that $P_0 \subset P$ and that $y \notin P_0$. This contradicts the minimality of P. Thus, indeed P is a singleton, say $\{v\}$; but, to say that $\{v\}$ is a face of K is the same as saying that v is an extreme point of K

We have thus proved that any compact convex set in $\mathbb{R}^n$ admits at least one extreme point. Thus, $\partial_e(K) \neq \emptyset$. If we let K_0 denote the closure of the convex hull of

Box 1.*

The notion of convexity (actually that of a convex function) was described precisely first by Archimedes (287-212 BC). Implicit contributions, though, were made earlier by various people including Eudoxus (400-347 BC) and Euclid (around 300 BC).

Convex sets have many interesting properties. Similar properties are often derived for other sets by imposing some analytic assumptions like smoothness. Convexity is, in many cases, an interesting substitute for these analytic requirements. An illustrative example is the fact that the boundary of a compact, convex set in $\mathbf{R}^n$ has a finite ($n-1$ dimensional) area. (See discussion following Theorem 6 for a description of compactness.) Various isoperimetric inequalities, including the classical one are also true for convex sets without any further smoothness assumptions. (See the article by A Sitaram, *Resonance*, September 1997 for a discussion of the classical isoperimetric inequality.)

A closed convex set C in $\mathbf{R}^n$ can be one of three types:

(i) Bounded

(ii) Unbounded but not containing a line in $\mathbf{R}^n$ (though it may contain a 'half-line').

(iii)Unbounded and containing a line in $\mathbf{R}^n$.

In case (i) C is homeomorphic to a closed ball, in case (ii) C is homeomorphic to a closed half space and in case (iii) C is homeomorphic to a cylinder with a convex (possibly unbounded) cross-section. (A *homeomorphism* is a continuous, one-to-one and onto map with a continuous inverse. A *half space* in $\mathbf{R}^n$ is a region in $\mathbf{R}^n$ defined by a linear inequality in the co-ordinates i.e. it is the region lying on one side of a hyperplane.)

Every point on the boundary of a convex set C in $\mathbf{R}^n$ has at least one hyperplane containing it such that C lies entirely in one of the two closed half spaces determined by the hyperplane. Such a hyperplane is called a *supporting hyperplane* for the set C.

This yields the following characterisation: A closed set in $\mathbf{R}^n$ is convex if and only if each point on the boundary lies on a supporting hyperplane for the set. Moreover, a face of the convex set (see Definition 4) is just the intersection of the set with some supporting hyperplane.

The standard n-simplex (see Exercise 5) is a special case of what are called convex

Box 1 continued...

polyhedra. A *convex polyhedron* is the closed convex hull of a finite number of points in $\mathbb{R}^n$. Such a polyhedron is always the intersection of a finite collection of closed half spaces and is thus described by a finite set of linear inequalities. The subject of linear programming which deals with extrema of linear forms subject to linear constraints can, in many cases, be thought of as a study of extrema of linear forms on convex polyhedra.

One approach to the study of convex sets is to approximate them by convex polyhedra. Since a convex polyhedron is defined by a finite set of data, this method is especially effective.

* Thanks are due to Dr Cherian Varughese for preparing this box.

$\partial_e(K)$, then K_0 is easily seen to be a compact convex subset of K We need to show that $K_0 = K$. Suppose, if possible, that there exists a point $z \in K$ such that $z \notin K_0$. Then, by another application of Proposition 10, we can choose a linear function g and a real number a_0 such that $g(x) < a_0 < g(z)\ \forall\ x \in K_0$. (To be totally correct, we should first note that, in the notation of Proposition 10, if we set $g = -f$ and $a_0 = -a$, then indeed $g(x) < a_0 < g(z)\ \forall\ x \in K$.) Argue next that the set $K_1 = \{y \in K : g(y) = \max\{g(w) : w \in K\}\ \}$ is a non-empty compact convex set and in fact, a face of K, which is disjoint from K_0; the first sentence of this paragraph, when applied to K_1, says that there exists an extreme point, say v_1 of K_1; it follows now from Exercise 5(7)(b) that an extreme point of the face K_1 of K is necessarily an extreme point of K – i.e., $v_1 \in K_0$ by the definition of K_0; the desired contradiction, and hence the end of the proof of the theorem, has been reached.

Suggested Reading

[1] Paul R Halmos, *Naive Set Theory*, Van Nostrand, Princeton, 1960.

[2] Walter Rudin, *Principles of Mathematical Analysis*, 3rd Ed., McGraw Hill, New York, 1991.

[3] George F Simmons, *Introduction to Topology and Modern Analysis*, McGraw Hill, New York, 1963.

Address for Correspondence
V S Sunder
The Institute of Mathematical Sciences
Taramani P.O.
Chennai 600 113, India.