Chapter 5

Topological vector spaces

One way to think of functional analysis is as the branch of mathematics that studies the extent to which the properties possessed by finite dimensional spaces generalize to infinite dimensional spaces. In the finite dimensional case there is only one natural linear topology. In that topology every linear functional is continuous, convex functions are continuous (at least on the interior of their domains), the convex hull of a compact set is compact, and nonempty disjoint closed convex sets can always be separated by hyperplanes. On an infinite dimensional vector space, there is generally more than one interesting topology, and the topological dual, the set of continuous linear functionals, depends on the topology. In infinite dimensional spaces convex functions are not always continuous, the convex hull of a compact set need not be compact, and nonempty disjoint closed convex sets cannot generally be separated by a hyperplane. However, with the right topology and perhaps some additional assumptions, each of these results has an appropriate infinite dimensional version.

Continuous linear functionals are important in economics because they can often be interpreted as prices. Separating hyperplane theorems are existence theorems asserting the existence of a continuous linear functional separating disjoint convex sets. These theorems are the basic tools for proving the existence of efficiency prices, state-contingent prices, and Lagrange multipliers in Kuhn–Tucker type theorems. They are also the cornerstone of the theory of linear inequalities, which has applications in the areas of mechanism design and decision theory. Since there is more than one topology of interest on an infinite dimensional space, the choice of topology is a key modeling decision that can have economic as well as technical consequences.

The proper context for separating hyperplane theorems is that of linear topologies, especially locally convex topologies. The classic works of N. Dunford and J. T. Schwartz [110, Chapter V], and J. L. Kelley and I. Namioka, *et al.* [199], as well as the more modern treatments by R. B. Holmes [166], H. Jarchow [181], J. Horváth [168], A. P. Robertson and W. J. Robertson [287], H. H. Schaefer [293], A. E. Taylor and D. C. Lay [330], and A. Wilansky [341] are good references on the general theory of linear topologies. R. R. Phelps [278] gives an excellent treatment of convex functions on infinite dimensional spaces. For applications to problems of optimization, we recommend J.-P. Aubin and I. Ekeland [23], I. Ekeland and R. Temam [115], I. Ekeland and T. Turnbull [116], and R. R. Phelps [278].

Here is the road map for this chapter. We start by defining a *topological vector space* (tvs) as a vector space with a topology that makes the vector operations continuous. Such a topology is *translation invariant* and can therefore be characterized by the neighborhood base at zero. While the topology may not be metrizable, there is a base of neighborhoods that behaves in some ways like the family of balls of positive radius (Theorem 5.6). In particular, if V is a neighborhood of zero, it includes another neighborhood W such that $W + W \subset V$. So if we think of V as an ε-ball, then W is like the $\varepsilon/2$-ball.

There is a topological characterization of finite dimensional topological vector spaces. (Finite dimensionality is an algebraic, not topological property.) A Hausdorff tvs is finite dimensional if and only if it is locally compact (Theorem 5.26). There is a unique Hausdorff linear topology on any finite dimensional space, namely the Euclidean topology (Theorem 5.21). Any finite dimensional subspace of a Hausdorff tvs is closed (Corollary 5.22) and *complemented* (Theorem 5.89) in locally convex spaces.

There is also a simple characterization of metrizable topological vector spaces. A Hausdorff tvs is metrizable if and only if there is a countable neighborhood base at zero (Theorem 5.10).

Without additional structure, these spaces can be quite dull. In fact, it is possible to have an infinite dimensional metrizable tvs where zero is the only continuous linear functional (Theorem 13.31). The additional structure comes from convexity. A set is *convex* if it includes the line segments joining any two of its points. A real function f is convex if its epigraph, $\{(x, \alpha) : \alpha \geqslant f(x)\}$, is convex. All linear functionals are convex. A convex function on an open convex set is continuous if it is bounded above on a neighborhood of a point (Theorem 5.43). Thus linear functions are continuous if and only if they are bounded on a neighborhood of zero. When zero has a base of convex neighborhoods, the space is *locally convex*. These are the spaces we really want. A convex neighborhood of zero gives rise to a convex homogeneous function known as its *gauge*. The gauge function of a set tells for each point how much the set must be enlarged to include it. In a normed space, the *norm* is the gauge of the unit ball. Not all locally convex spaces are normable, but the family of gauges of symmetric convex neighborhoods of zero, called *seminorms*, are a good substitute. The best thing about locally convex spaces is that they have lots of continuous linear functionals. This is a consequence of the seemingly innocuous Hahn–Banach Extension Theorem 5.53. The most important consequence of the Hahn–Banach Theorem is that in a locally convex space, there are *hyperplanes* that strictly separate points from closed convex sets that don't contain them (Corollary 5.80). As a result, every closed convex set is the intersection of all closed *half spaces* including it.

Another of the consequences of the Hahn–Banach Theorem is that the set of continuous linear functionals on a locally convex space separates points. The collection of continuous linear functionals on X is known as the (topological) *dual space*, denoted X'. Now each $x \in X$ defines a linear functional on X' by $x(x') = x'(x)$. Thus we are led to the study of *dual pairs* $\langle X, X' \rangle$ of spaces and their associated *weak topologies*. These weak topologies are locally convex. The weak topology on X' induced by X is called the *weak*∗ topology on X'. The most familiar example of a dual pair is probably the pairing of functions and measures—each defines a linear functional via the integral $\int f \, d\mu$, which is linear in f for a fixed μ, and linear in μ for a fixed f. (The weak topology induced on probability measures by this duality with continuous functions is the topology of convergence in distribution that is used in Central Limit Theorems.) Remarkably, in a dual pair $\langle X, X' \rangle$, any subspace of X' that separates the points of X is weak∗ dense in X' (Corollary 5.108).

G. Debreu [84] introduced dual pairs in economics in order to describe the duality between commodities and prices. According to this interpretation, a dual pair $\langle X, X' \rangle$ represents the commodity-price duality, where X is the commodity space, X' is the price space, and $\langle x, x' \rangle$ is the value of the bundle x at prices x'. This is the basic ingredient of the Arrow–Debreu–McKenzie model of general economic equilibrium; see [9].

If we put the weak topology on X generated by X', then X' is the set of all continuous linear functionals on X (Theorem 5.93). Given a weak neighborhood V of zero in X, we look at all the linear functionals that are bounded on this neighborhood. Since they are bounded, they are continuous and so lie in X'. We further normalize them so that they are bounded by unity on V. The resulting set is called the *polar* of V, denoted V°. The remarkable Alaoglu Theorem 5.105 asserts that V° is compact in the weak topology X generates on X'. Its proof relies on the Tychonoff Product Theorem 2.61. The useful Bipolar Theorem 5.103 states the polar of the polar of a set A is the closed convex circled hull of A.

We might ask what other topologies besides the weak topology on X give X' as the dual. The Mackey–Arens Theorem 5.112 answers this question. The answer is that for a topology on X to have X' as its dual, there must be a base at zero consisting of the duals of a family of weak∗ compact convex circled subsets of X'. Thus the topology generated by the polars of all the weak∗ compact convex circled sets in X' is the strongest topology on X for which X' is the dual. This topology is called the *Mackey topology* on X, and it has proven to be extremely useful in the study of infinite dimensional economies. It was introduced to economics by T. F. Bewley [40]. The usefulness stems from the fact that once the dual space of continuous linear functionals has been fixed, the Mackey topology allows the greatest number of continuous real (nonlinear) functions.

There are entire volumes devoted to the theory of topological vector spaces, so we cannot cover everything in one chapter. Chapter 6 describes the additional properties of spaces where the topology is derived from a norm. Chapter 7 goes into more depth on the properties of convex sets and functions. Convexity involves a strange synergy between the topological structure of the space and its algebraic structure. A number of results there are special to the finite dimensional case. Another important aspect of the theory is the interaction of the topology and the order structure of the space. Chapter 8 covers *Riesz spaces*, which are partially ordered topological vector spaces where the partial order has topological and algebraic restrictions modeled after the usual order on $\mathbb{R}^n$. Chapter 9 deals with normed partially ordered spaces.

5.1 Linear topologies

Recall that a (real) **vector space** or (real) **linear space** is a set X (whose elements are called vectors) with two operations: addition, which assigns to each pair of vectors x, y the vector $x + y$, and scalar multiplication, which assigns to vector x and each scalar (real number) α the vector αx. There is a special vector 0. These operations satisfy the following properties: $x + y = y + x$, $(x + y) + z = x + (y + z)$, $x + 0 = x$, $x + (-1)x = 0$, $1x = x$, $\alpha(\beta x) = (\alpha\beta)x$, $\alpha(x + y) = \alpha x + \alpha y$, and $(\alpha + \beta)x = \alpha x + \beta x$. (There are also complex vector spaces, where the scalars are complex numbers, but we won't have occasion to refer to them.)

A subset of a vector space is called a **vector subspace** or (**linear subspace**) if it is a vector space in its own right under the induced operations. The **(linear) span** of a subset is the smallest vector subspace including it. A function $f: X \to Y$ between two vector spaces is **linear** if it satisfies

$$f(\alpha x + \beta z) = \alpha f(x) + \beta f(z)$$

for every $x, z \in X$ and $\alpha, \beta \in \mathbb{R}$. Linear functions between vector spaces are usually called **linear operators**. A linear operator from a vector space to the real line is called a **linear functional**.

A topology τ on a vector space X is called a **linear topology** if the operations addition and scalar multiplication are τ-continuous. That is, if $(x, y) \mapsto x + y$ from $X \times X$ to X and $(\alpha, x) \mapsto \alpha x$ from $\mathbb{R} \times X$ to X are continuous. Then (X, τ) is called a **topological vector space** or **tvs** for short. (A topological vector space may also be called a **linear topological space**, especially in older texts.) A tvs need not be a Hausdorff space.

A mapping $\varphi: L \to M$ between two topological vector spaces is a **linear homeomorphism** if φ is one-to-one, linear, and $\varphi: L \to \varphi(L)$ is a homeomorphism. The linear homeomorphism φ is also called an **embedding** and $\varphi(L)$ is referred to as a **copy** of L in M. Two topological vector spaces are **linearly homeomorphic** if there exists a linear homeomorphism from one onto the other.

5.1 Lemma *Every vector subspace of a tvs with the induced topology is a topological vector space in its own right.*

Products of topological vector spaces are topological vector spaces.

5.2 Theorem *The product of a family of topological vector spaces is a tvs under the pointwise algebraic operations and the product topology.*

Proof: Let $\{(X_i, \tau_i)\}_{i \in I}$ be a family of topological vector spaces and let $X = \prod_{i \in I} X_i$ and $\tau = \prod_{i \in I} \tau_i$. We show only that addition on X is continuous and leave the case of scalar multiplication as an exercise.

Let $(x_i^\alpha) \xrightarrow[\alpha]{\tau} (x_i)$ and $(y_i^\lambda) \xrightarrow[\lambda]{\tau} (y_i)$ in X. Then $x_i^\alpha \xrightarrow[\alpha]{\tau_i} x_i$ and $y_i^\lambda \xrightarrow[\lambda]{\tau_i} y_i$ in X_i for each i, so also $x_i^\alpha + y_i^\lambda \xrightarrow[\alpha,\lambda]{\tau_i} x_i + y_i$ in X_i for each i. Since the product topology on X is the topology of pointwise convergence, we see that

$$(x_i^\alpha) + (y_i^\lambda) = (x_i^\alpha + y_i^\lambda) \xrightarrow[\alpha,\lambda]{\tau} (x_i + y_i) = (x_i) + (y_i),$$

and the proof is finished. ∎

Linear topologies are **translation invariant**. That is, a set V is open in a tvs X if and only if the translation $a + V$ is open for all a. Indeed, the continuity of addition implies that for each $a \in X$, the function $x \mapsto a + x$ is a linear homeomorphism. In particular, every neighborhood of a is of the form $a + V$, where V is a neighborhood of zero. In other words, the neighborhood system at zero determines the neighborhood system at every point of X by translation. Also note that the mapping $x \mapsto \alpha x$ is linear homeomorphism for any $\alpha \neq 0$. In particular, if V is a neighborhood of zero, then so is αV for all $\alpha \neq 0$.

The most familiar linear topologies are derived from norms. A **norm** on a vector space is a real function $\|\cdot\|$ satisfying

1. $\|x\| \geqslant 0$ for all vectors x, and $\|x\| = 0$ implies $x = 0$.
2. $\|\alpha x\| = |\alpha| \, \|x\|$ for all vectors x and all scalars α.
3. $\|x + y\| \leqslant \|x\| + \|y\|$ for all vectors x and y.

A neighborhood base at zero consists of all sets of the form $\{x : \|x\| < \varepsilon\}$ where ε is a positive number. The **norm topology** for a norm $\|\cdot\|$ is the metrizable topology generated by the metric $d(x, y) = \|x - y\|$.

The next lemma presents some basic facts about subsets of topological vector spaces. Most of the proofs are straightforward.

5.3 Lemma *In a topological vector space:*

1. *The algebraic sum of an open set and an arbitrary set is open.*
2. *Nonzero multiples of open sets are open.*
3. *If B is open, then for any set A we have $\overline{A} + B = A + B$.*
4. *The algebraic sum of a compact set and a closed set is closed. (However, the algebraic sum of two closed sets need not be closed.)*

5. *The algebraic sum of two compact sets is compact.*

6. *Scalar multiples of closed sets are closed.*

7. *Scalar multiples of compact sets are compact.*

8. *A linear functional is continuous if and only if it is continuous at* 0.

Proof: We shall prove only parts (3) and (4).

(3) Clearly $\overline{A+B} \subset \overline{A} + B$. For the reverse inclusion, let $y \in \overline{A} + B$ and write $y = x + b$ where $x \in \overline{A}$ and $b \in B$. Then there is an open neighborhood V of zero such that $b + V \subset B$. Since $x \in \overline{A}$, there exists some $a \in A \cap (x - V)$. Then $y = x + b = a + b + (x - a) \in a + b + V \subset A + B$.

(4) Let A be compact and B be closed, and let a net $\{x_\alpha + y_\alpha\}$ in $A + B$ satisfy $x_\alpha + y_\alpha \to z$. Since A is compact, we can assume (by passing to a subnet) that $x_\alpha \to x \in A$. The continuity of the algebraic operations yields

$$y_\alpha = (x_\alpha + y_\alpha) - x_\alpha \to z - x = y.$$

Since B is closed, $y \in B$, so $z = x + y \in A + B$, proving that $A + B$ is closed. ∎

5.4 Example (Sum of closed sets) To see that the sum of two closed sets need not be closed, consider the closed sets $A = \{(x, y) : x > 0, y \geqslant \frac{1}{x}\}$ and $B = \{(x, y) : x < 0, y \geqslant -\frac{1}{x}\}$ in $\mathbb{R}^2$. While A and B are closed, neither is compact, and $A + B = \{(x, y) : y > 0\}$ is not closed. ∎

5.2 Absorbing and circled sets

We now describe some special algebraic properties of subsets of vector spaces. The **line segment** joining vectors x and y is the set $\{\lambda x + (1 - \lambda)y : 0 \leqslant \lambda \leqslant 1\}$.

5.5 Definition *A subset A of a vector space is:*

- ***convex*** *if it includes the line segment joining any pair of its points.*

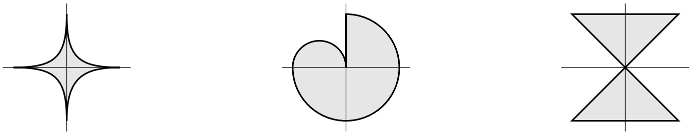
- ***absorbing*** *(or* ***radial****) if for any x some multiple of A includes the line segment joining x and zero. That is, if there exists some* $\alpha_0 > 0$ *satisfying* $\alpha x \in A$ *for every* $0 \leqslant \alpha \leqslant \alpha_0$.

  Equivalently, A is absorbing if for each vector x there exists some $\alpha_0 > 0$ *such that* $\alpha x \in A$ *whenever* $-\alpha_0 \leqslant \alpha \leqslant \alpha_0$.

- ***circled*** *(or* ***balanced****) if for each* $x \in A$ *the line segment joining x and* $-x$ *lies in A. That is, if for any* $x \in A$ *and any* $|\alpha| \leqslant 1$ *we have* $\alpha x \in A$.

- ***symmetric*** *if* $x \in A$ *implies* $-x \in A$.

- ***star-shaped about zero*** *if it includes the line segment joining each of its points with zero. That is, if for any* $x \in A$ *and any* $0 \leqslant \alpha \leqslant 1$ *we have* $\alpha x \in A$.



Circled and absorbing, but not convex.

Star-shaped, but neither symmetric nor convex.

Circled, but neither absorbing nor convex.

Figure 5.1. Shapes of sets in $\mathbb{R}^2$.

Note that an absorbing set must contain zero, and any set including an absorbing set is itself absorbing. For any absorbing set A, the set $A \cap (-A)$ is nonempty, absorbing, and symmetric. Every circled set is symmetric. Every circled set is star-shaped about zero, as is every convex set containing zero. See Figure 5.1 for some examples.

Let X be a topological vector space. For each fixed scalar $\alpha \neq 0$ the mapping $x \mapsto \alpha x$ is a linear homeomorphism, so αV is a neighborhood of zero whenever V is and $\alpha \neq 0$. Now if V is a neighborhood of zero, then the continuity of the function $(\alpha, x) \mapsto \alpha x$ at $(0, 0)$ guarantees the existence of a neighborhood W at zero and some $\alpha_0 > 0$ such that $x \in W$ and $|\alpha| \leqslant \alpha_0$ imply $\alpha x \in V$. Thus, if $U = \bigcup_{|\alpha| \leqslant \alpha_0} \alpha W$, then U is a neighborhood of zero, $U \subset V$, and U is circled. Moreover, from the continuity of the addition map $(x, y) \mapsto x + y$ at $(0, 0)$, we see that there is a neighborhood W of zero such that $x, y \in W$ implies $x+y \in V$, that is, $W+W \subset V$. Also note that since $W+W \subset V$, it follows that $\overline{W} \subset V$. (For if $x \in \overline{W}$, then $x - W$ is a neighborhood of x, so $(x - W) \cap W \neq \varnothing$ implies $x \in W + W \subset V$.)

Since the closure of an absorbing circled set remains absorbing and circled (why?), we have just shown that zero has a neighborhood base consisting of closed, absorbing, and circled sets. We cannot conclude that zero has a neighborhood base consisting of convex sets. If the tvs does have a neighborhood base at zero of convex sets, it is called a **locally convex space**.

The following theorem establishes the converse of the results above and characterizes the structure of linear topologies.

5.6 Structure Theorem *If (X, τ) is a tvs, then there is a neighborhood base $\mathcal{B}$ at zero such that:*

1. *Each $V \in \mathcal{B}$ is absorbing.*

2. *Each $V \in \mathcal{B}$ is circled.*

3. *For each $V \in \mathcal{B}$ there exists some $W \in \mathcal{B}$ with $W + W \subset V$.*

4. *Each $V \in \mathcal{B}$ is closed.*

Conversely, if a filter base $\mathcal{B}$ on a vector space X satisfies properties (1), (2), and (3) above, then there exists a unique linear topology τ on X having $\mathcal{B}$ as a neighborhood base at zero.

Proof: If τ is a linear topology, then by the discussion preceding the theorem, the collection of all τ-closed circled neighborhoods of zero satisfies the desired properties. For the converse, assume that a filter base $\mathcal{B}$ of a vector space X satisfies properties (1), (2), and (3).

We have already mentioned that a linear topology is translation invariant and so uniquely determined by the neighborhoods of zero. So define τ to be the collection of all subsets A of X satisfying

$$A = \{x \in A : \exists\, V \in \mathcal{B} \text{ such that } x + V \subset A\}. \qquad (\star)$$

Then clearly $\varnothing, X \in \tau$ and the collection τ is closed under arbitrary unions. If $A_1, \dots, A_k \in \tau$ and $x \in A_1 \cap \cdots \cap A_k$, then for each $i = 1, \dots, k$ there exists some $V_i \in \mathcal{B}$ such that $x + V_i \subset A_i$. Since $\mathcal{B}$ is a filter base, there exists some $V \in \mathcal{B}$ with $V \subset V_1 \cap \cdots \cap V_k$. Now note that $x + V \subset A_1 \cap \cdots \cap A_k$ and this proves that $A_1 \cap \cdots \cap A_k \in \tau$. Therefore, we have established that τ is a topology on X.

The next thing we need to observe is that if for each subset A of X we let

$$A^\sharp = \{x \in A : \exists\, V \in \mathcal{B} \text{ such that } x + V \subset A\},$$

then $A^\sharp$ coincides with the τ-interior of A, that is, $A^\circ = A^\sharp$. If $x \in A^\circ$, then by $(\star)$ and the fact that A° is τ-open, there exists some $V \in \mathcal{B}$ such that $x + V \subset A^\circ \subset A$, so $x \in A^\sharp$. Therefore, $A^\circ \subset A^\sharp$. To see that equality holds, it suffices to show that $A^\sharp$ is τ-open. To this end, $y \in A^\sharp$. Pick some $V \in \mathcal{B}$ such that $y + V \subset A$. By (3) there exists some $W \in \mathcal{B}$ such that $W + W \subset V$. Now if $w \in W$, then we have $y + w + W \subset y + W + W \subset y + V \subset A$, so that $y + w \in A^\sharp$ for each $w \in W$, that is, $y + W \subset A^\sharp$. This proves that $A^\sharp$ is τ-open, so $A^\circ = A^\sharp$.

Now it easily follows that for each $x \in X$ the collection $\{x + V : V \in \mathcal{B}\}$ is a τ-neighborhood base at x. Next we shall show that the addition map $(x, y) \mapsto x + y$ is a continuous function. To see this, fix $x_0, y_0 \in X$ and a set $V \in \mathcal{B}$. Choose some $U \in \mathcal{B}$ with $U + U \subset V$ and note that $x \in x_0 + U$ and $y \in y_0 + U$ imply $x + y \in x_0 + y_0 + V$. Consequently, the addition map is continuous at (x_0, y_0) and therefore is a continuous function.

Finally, let us prove the continuity of scalar multiplication. Fix $\lambda_0 \in \mathbb{R}$ and $x_0 \in X$ and let $V \in \mathcal{B}$. Pick some $W \in \mathcal{B}$ such that $W + W \subset V$. Since W is an absorbing set there exists some $\varepsilon > 0$ such that for each $-\varepsilon < \delta < \varepsilon$ we have $\delta x_0 \in W$. Next, select a natural number $n \in \mathbb{N}$ with $|\lambda_0| + \varepsilon < n$ and note that if $\lambda \in \mathbb{R}$ satisfies $|\lambda - \lambda_0| < \varepsilon$, then $\left|\frac{\lambda}{n}\right| \leqslant \frac{|\lambda_0|+\varepsilon}{n} < 1$. Now since W is (by (2)) a circled set, for each $\lambda \in \mathbb{R}$ with $|\lambda - \lambda_0| < \varepsilon$ and all $x \in x_0 + \frac{1}{n}W$ we have

$$\lambda x = \lambda_0 x_0 + (\lambda - \lambda_0)x_0 + \lambda(x - x_0) \in \lambda_0 x_0 + W + \tfrac{\lambda}{n}W \subset \lambda_0 x_0 + W + W \subset \lambda_0 x_0 + V.$$

This shows that multiplication is a continuous function at (λ_0, x_0). ∎

In a topological vector space the interior of a circled set need not be a circled set; see, for instance, the third set in Figure 5.1. However, the interior of a circled neighborhood V of zero is automatically an open circled set. To see this, note first that 0 is an interior point of V. Now let $x \in V^\circ$ and fix some nonzero $\lambda \in \mathbb{R}$ with $|\lambda| \leqslant 1$. Pick some neighborhood W of zero with $x + W \subset V$ and note that the neighborhood λW of zero satisfies $\lambda x + \lambda W = \lambda(x + W) \subset \lambda V \subset V$. Therefore $\lambda x \in V^\circ$ for each $|\lambda| \leqslant 1$, so V° is a circled set. This conclusion yields the following.

5.7 Lemma *In a topological vector space the collection of all open and circled neighborhoods of zero is a base for the neighborhood system at zero.*

If τ is a linear topology on a vector space and $\mathcal{N}$ denotes the τ-neighborhood system at zero, then the set $K_\tau = \bigcap_{V \in \mathcal{N}} V$ is called the **kernel of the topology** τ. From Theorem 5.6 it is not difficult to see that K_τ is a closed vector subspace. The vector subspace K_τ is the trivial subspace $\{0\}$ if and only if τ is a Hausdorff topology. The proof of the next result is straightforward and is left for the reader.

5.8 Lemma *A linear topology τ on a vector space is Hausdorff if and only if its kernel K_τ is trivial (and also if and only if $\{0\}$ is a τ-closed set).*

Property (3) of the Theorem 5.6 allows to use "$\varepsilon/2$ arguments" even when we don't have a metric. As an application of this result, we offer another instance of the informal principle that compact sets behave like points.

5.9 Theorem *Let K be a compact subset of a topological vector space X, and suppose $K \subset U$, where U is open. Then there exist an open neighborhood W of zero and a finite subset Φ of K such that*

$$K \subset \Phi + W \subset K + W \subset U.$$

Proof: Since $K \subset U$, for each $x \in K$, there is open neighborhood W_x of zero such that $x + W_x + W_x \subset U$. Since K is compact, there is a finite set $\{x_1, \ldots, x_n\}$ with $K \subset \bigcup_{i=1}^n (x_i + W_{x_i})$. Let $W = \bigcap_{i=1}^n W_{x_i}$ and note that W is an open neighborhood of zero. Since the open sets $x_i + W_{x_i}$, $i = 1, \ldots, n$, cover K, given $y \in K$, there is an x_i satisfying $y \in x_i + W_{x_i}$. For this x_i we have $y + W \subset x_i + W_{x_i} + W_{x_i} \subset U$, and from this we see that $K + W \subset U$.

Now from $K \subset K + W = \bigcup_{y \in K} (y + W) \subset U$ and the compactness of K, it follows that there exists a finite subset $\Phi = \{y_1, \ldots, y_m\}$ of K such that $K \subset \bigcup_{j=1}^m (y_j + W)$. Now note that $K \subset \Phi + W \subset K + W \subset U$. ∎

5.3 Metrizable topological vector spaces

A metric d on a vector space is said to be **translation invariant** if it satisfies $d(x + a, y + a) = d(x, y)$ for all x, y, and a. Every metric induced by a norm is translation invariant, but the converse is not true (see Example 5.78 below). For Hausdorff topological vector spaces, the existence of a compatible translation invariant metric is equivalent to first countability.

5.10 Theorem *A Hausdorff topological vector space is metrizable if and only if zero has a countable neighborhood base. In this case, the topology is generated by a translation invariant metric.*

Proof: Let (X, τ) be a Hausdorff tvs. If τ is metrizable, then τ has clearly a countable neighborhood base at zero. For the converse, assume that τ has a countable neighborhood base at zero. Choose a countable base $\{V_n\}$ of circled neighborhoods of zero such that $V_{n+1} + V_{n+1} + V_{n+1} \subset V_n$ holds for each n. Now define the function $\rho\colon X \to [0, \infty)$ by

$$\rho(x) = \begin{cases} 1 & \text{if } x \notin V_1, \\ 2^{-k} & \text{if } x \in V_k \setminus V_{k+1}, \\ 0 & \text{if } x = 0. \end{cases}$$

Then it is easy to check that for each $x \in X$ we have the following:

1. $\rho(x) \geqslant 0$ and $\rho(x) = 0$ if and only if $x = 0$.
2. $x \in V_k$ for some k if and only if $\rho(x) \leqslant 2^{-k}$.
3. $\rho(x) = \rho(-x)$ and $\rho(\lambda x) \leqslant \rho(x)$ for all $|\lambda| \leqslant 1$.
4. $\lim_{\lambda\to 0} \rho(\lambda x) = 0$.

We also note the following property.

- $x_n \xrightarrow{\tau} 0$ *if and only if* $\rho(x_n) \to 0$.

Now by means of the function ρ we define the function $\pi\colon X \to [0, \infty)$ via the formula:

$$\pi(x) = \inf\Big\{\sum_{i=1}^{n} \rho(x_i) : x_1, \ldots, x_n \in X \text{ and } \sum_{i=1}^{n} x_i = x\Big\}.$$

The function π satisfies the following properties.

a. $\pi(x) \geqslant 0$ *for each* $x \in X$.

b. $\pi(x + y) \leqslant \pi(x) + \pi(y)$ *for all* $x, y \in X$.

c. $\frac{1}{2}\rho(x) \leqslant \pi(x) \leqslant \rho(x)$ *for each* $x \in X$ (*so* $\pi(x) = 0$ *if and only if* $x = 0$).

Property (a) follows immediately from the definition of π. Property (b) is straightforward. The proof of (c) will be based upon the following property:

$$\textit{If } \sum_{i=1}^{n} \rho(x_i) < \tfrac{1}{2^m}, \textit{ then } \sum_{i=1}^{n} x_i \in V_m. \tag{$\star$}$$

To verify ($\star$), we use induction on n. For $n = 1$ we have $\rho(x_1) < 2^{-m}$, and consequently $x_1 \in V_{m+1} \subset V_m$ is trivially true. For the induction step, assume that if $\{x_i : i \in I\}$ is any collection of at most n vectors satisfying $\sum_{i \in I} \rho(x_i) < 2^{-m}$ for some $m \in \mathbb{N}$, then $\sum_{i \in I} x_i \in V_m$.

Suppose that $\sum_{i=1}^{n+1} \rho(x_i) < \frac{1}{2^m}$ for some $m \in \mathbb{N}$. Clearly, we have $\rho(x_i) \leqslant \frac{1}{2^{m+1}}$, so $x_i \in V_{m+1}$ for each $1 \leqslant i \leqslant n+1$. We now distinguish the following two cases.

Case 1: $\sum_{i=1}^{n+1} \rho(x_i) < \frac{1}{2^{m+1}}$

Clearly $\sum_{i=1}^{n} \rho(x_i) < \frac{1}{2^{m+1}}$, so by the induction hypothesis $\sum_{i=1}^{n} x_i \in V_{m+1}$. Thus

$$\sum_{i=1}^{n+1} x_i = \sum_{i=1}^{n} x_i + x_{n+1} \in V_{m+1} + V_{m+1} \subset V_m.$$

Case 2: $\sum_{i=1}^{n+1} \rho(x_i) \geqslant \frac{1}{2^{m+1}}$

Let $1 \leqslant k \leqslant n+1$ be the largest k such that $\sum_{i=k}^{n+1} \rho(x_i) \geqslant \frac{1}{2^{m+1}}$. If $k = n+1$, then $\rho(x_{n+1}) = \frac{1}{2^{m+1}}$, so from $\sum_{i=1}^{n+1} \rho(x_i) < \frac{1}{2^m}$ we have $\sum_{i=1}^{n} \rho(x_i) < \frac{1}{2^{m+1}}$. But then, as in Case 1, we get $\sum_{i=1}^{n+1} x_i \in V_m$.

Thus, we can assume that $k < n+1$. Assume first that $k > 1$. From the inequalities $\sum_{i=1}^{n+1} \rho(x_i) < \frac{1}{2^m}$ and $\sum_{i=k}^{n+1} \rho(x_i) \geqslant \frac{1}{2^{m+1}}$, we obtain $\sum_{i=1}^{k-1} \rho(x_i) < \frac{1}{2^{m+1}}$. So our induction hypothesis yields $\sum_{i=1}^{k-1} x_i \in V_{m+1}$. Also, by the choice of k we have $\sum_{i=k+1}^{n+1} \rho(x_i) < \frac{1}{2^{m+1}}$, and thus by our induction hypothesis also we have $\sum_{i=k+1}^{n+1} x_i \in V_{m+1}$. Therefore, in this case we obtain

$$\sum_{i=1}^{n+1} x_i = \sum_{i=1}^{k-1} x_i + x_k + \sum_{i=k+1}^{n+1} x_i \in V_{m+1} + V_{m+1} + V_{m+1} \subset V_m.$$

If $k = 1$, then we have $\sum_{i=2}^{n+1} \rho(x_i) < \frac{1}{2^{m+1}}$, so $\sum_{i=2}^{n+1} x_i \in V_{m+1}$. This implies $\sum_{i=1}^{n+1} x_i = x_1 + \sum_{i=2}^{n+1} x_i \in V_{m+1} + V_{m+1} \subset V_m$. This completes the induction and the proof of ($\star$).

Next, we verify (c). To this end, let $x \in X$ satisfy $\rho(x) = 2^{-m}$ for some $m \geqslant 0$. Also, assume by way of contradiction that the vectors $x_1, \ldots, x_k$ satisfy $\sum_{i=1}^{k} x_i = x$ and $\sum_{i=1}^{k} \rho(x_i) < \frac{1}{2}\rho(x) = 2^{-m-1}$. But then, from ($\star$) we get $x = \sum_{i=1}^{k} x_i \in V_{m+1}$, so $\rho(x) \leqslant 2^{-m-1} < 2^{-m} = \rho(x)$, which is impossible. This contradiction, establishes the validity of (c).

Finally, for each $x, y \in X$ define $d(x, y) = \pi(x-y)$ and note that d is a translation invariant metric that generates τ. ∎

Even if a tvs is not metrizable, it is nonetheless uniformizable by a translation invariant uniformity. For a proof of this result, stated below, see, for example, H. H. Schaefer [293, §1.4, pp. 16–17].

5.11 Theorem *A topological vector space is uniformizable by a unique translation invariant uniformity. A base for the uniformity is the collection of sets of the form $\{(x, y) : x - y \in V\}$ where V ranges over a neighborhood base $\mathcal{B}$ at zero.*

A **Cauchy net** in a topological vector space is a net $\{x_\alpha\}$ such that for each neighborhood V of zero there is some α_0 such that $x_\alpha - x_\beta \in V$ for all $\alpha, \beta \geqslant \alpha_0$. Every convergent net is Cauchy. (Why?) Similarly, a filter $\mathcal{F}$ on a topological vector space is called a **Cauchy filter** if for each neighborhood V of zero there exists some $A \in \mathcal{F}$ such that $A - A \subset V$. Convergent filters are clearly Cauchy. From the discussion in Section 2.6, it is easy to see that a filter is Cauchy if and only if the net it generates is a Cauchy net (and that a net is Cauchy if and only if the filter it generates is Cauchy).

A topological vector space (X, τ) is **topologically complete**, or simply **complete** (and τ is called a **complete topology**), if every Cauchy net is convergent, or equivalently, if every Cauchy filter is convergent.

The proof of the next lemma is straightforward and is omitted.

5.12 Lemma *Let $\{(X_i, \tau_i)\}_{i \in I}$ be a family of topological vector spaces, and let $X = \prod_{i \in I} X_i$ endowed with the product topology $\tau = \prod_{i \in I} \tau_i$. Then (X, τ) is τ-complete if and only if each factor (X_i, τ_i) is τ_i-complete.*

If a linear topology τ on a vector space X is generated by a translation invariant metric d, then (X, d) is a complete metric space if and only if (X, τ) is topologically complete as defined above, that is, (X, d) is a complete metric space if and only if every τ-Cauchy sequence in X is τ-convergent. Not every consistent metric of a metrizable topological vector space is translation invariant. For instance, consider the three metrics d_1, d_2, and d_3 on $\mathbb{R}$ defined by:

$$\begin{aligned} d_1(x, y) &= |x - y|, \\ d_2(x, y) &= |x - y| + \left| \tfrac{1}{1+|x|} - \tfrac{1}{1+|y|} \right|, \text{ and} \\ d_3(x, y) &= \left| e^{-x} - e^{-y} \right|. \end{aligned}$$

Then d_1, d_2, and d_3 are equivalent metrics, d_1 is complete and translation invariant, d_2 is complete but not translation invariant, and d_3 is neither complete nor translation invariant.

5.13 Definition *A **completely metrizable topological vector space** is a topologically complete metrizable topological vector space. In other words, a completely metrizable tvs is a topologically complete tvs having a countable neighborhood base at zero.*

Note that (according to Theorem 5.10) every completely metrizable topological vector space admits a compatible translation invariant complete metric. Clearly, the class of completely metrizable topological vector spaces includes the class of Banach spaces.

A complete Hausdorff topological vector space Y is called a **topological completion** or simply a **completion** of another Hausdorff topological vector space X if there is a linear homeomorphism $T: X \to Y$ such that $T(X)$ is dense in Y; identifying X with $T(X)$, we can think of X as a subspace of Y. This leads to the next result, which appears in many places; see, for instance, J. Horváth [168, Theorem 1, p. 131].

5.14 Theorem *Every Hausdorff topological vector space has a unique (up to linear homeomorphism) topological completion.*

The concept of uniform continuity makes sense for functions defined on subsets of topological vector spaces. A function $f: A \to Y$, where A is a subset of a tvs X and Y is another tvs, is **uniformly continuous** if for each neighborhood V of zero in Y there exists a neighborhood W of zero in X such that $x, y \in A$ and $x - y \in W$ imply $f(x) - f(y) \in V$. You should notice that if X is a tvs, then both addition $(x, y) \mapsto x + y$, from $X \times X$ to X, and scalar multiplication $(\alpha, x) \mapsto \alpha x$, from $\mathbb{R} \times X$ to X, are uniformly continuous.

The analogue of Lemma 3.11 can now be stated as follows—the proof is left as an exercise.

5.15 Theorem *Let A be a subset of a tvs, let Y be a complete Hausdorff topological vector space, and let $f: A \to Y$ be uniformly continuous. Then f has a unique uniformly continuous extension to the closure $\overline{A}$ of A.*

5.4 The Open Mapping and Closed Graph Theorems

In this section we prove two basic theorems of functional analysis, the Open Mapping Theorem and the Closed Graph Theorem. We do this in the setting of completely metrizable topological vector spaces. For more on these theorems and extensions to general topological vector spaces we recommend T. Husain [174]. We start by recalling the definition of an operator.

5.16 Definition *A function $T: X \to Y$ between two vector spaces is a **linear operator** (or simply an **operator**) if*

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$$

*for all $x, y \in X$ and all scalars $\alpha, \beta \in \mathbb{R}$. When Y is the real line $\mathbb{R}$, we call T a **linear functional**.*

It is common to denote the vector $T(x)$ by Tx, and we do it quite often. If $T\colon X \to Y$ is not a linear operator, then T is referred to as a **nonlinear operator**. The following lemma characterizes continuity of linear operators.

5.17 Lemma (Continuity at zero) *An operator $T\colon X \to Y$ between topological vector spaces is continuous if and only if it is continuous at zero (in which case it is uniformly continuous).*

Proof: Everything follows from the identity $T(x) - T(y) = T(x - y)$. ∎

Recall that a function between topological spaces is called an **open mapping** if it carries open sets to open sets.

5.18 The Open Mapping Theorem *A surjective continuous operator between completely metrizable topological vector spaces is an open mapping.*

Proof: Let $T\colon (X_1, \tau_1) \to (X_2, \tau_2)$ be a surjective continuous operator between completely metrizable topological vector spaces and let U be a circled τ_1-closed neighborhood of zero. It suffices to show that the set $T(U)$ is a τ_2-neighborhood of zero. We first establish the following claim.

- *For any τ_1-neighborhood W of zero in X_1 there exists a τ_2-neighborhood V of zero in X_2 satisfying $V \subset \overline{T(W)}$.*

To see this, let W and W_0 be circled τ_1-neighborhoods of zero that satisfy $W_0 + W_0 \subset W$. From $X_1 = \bigcup_{n=1}^{\infty} nW_0$ and the fact that T is surjective, it follows that $X_2 = T(X_1) = \bigcup_{n=1}^{\infty} nT(W_0)$. Therefore, by the Baire Category Theorem 3.47, for some n the set $\overline{nT(W_0)} = n\overline{T(W_0)}$ must have an interior point. This implies that there exists some $y \in \overline{T(W_0)}$ and some circled τ_2-neighborhood V of zero with $y + V \subset \overline{T(W_0)}$. Since $\overline{T(W_0)}$ is symmetric, we see that $v - y \in \overline{T(W_0)}$ for each $v \in V$. Thus, if $v \in V$, then it follows from $v = (v-y)+y \in \overline{T(W_0)} + \overline{T(W_0)} \subset \overline{T(W)}$ that $v \in \overline{T(W)}$, so $V \subset \overline{T(W)}$.

Now pick a countable base $\{W_n\}$ at zero for τ_1 consisting of τ_1-closed circled sets satisfying $W_{n+1} + W_{n+1} \subset W_n$ for all $n = 1, 2, \dots$ and $W_1 + W_1 \subset U$. The claim established above and an easy inductive argument guarantee the existence of a countable base $\{V_n\}$ at zero for τ_2 consisting of circled and τ_2-closed sets satisfying $V_{n+1} + V_{n+1} \subset V_n$ and $V_n \subset \overline{T(W_n)}$ for all $n = 1, 2, \dots$. We finish the proof by showing that $V_1 \subset T(U)$.

To this end, let $y \in V_1$. From $V_1 \subset \overline{T(W_1)}$ and the fact that $y + V_2$ is a τ_2-neighborhood of y, it follows that there exists some $w_1 \in W_1$ with $y - T(w_1) \in V_2$, so $y - T(w_1) \in \overline{T(W_2)}$. Now by an inductive argument, we can construct a sequence $\{w_n\}$ in X_1 such that for each $n = 1, 2, \dots$ we have $w_n \in W_n$ and

$$y - \sum_{i=1}^{n} T(w_i) = y - T\Big(\sum_{i=1}^{n} w_i\Big) \in V_{n+1}. \tag{$\star$}$$

Next, let $x_n = \sum_{i=1}^{n} w_i$ and note that from

$$x_{n+p} - x_n = \sum_{i=n+1}^{n+p} w_i \in W_{n+1} + W_{n+2} + \cdots + W_{n+p} \subset W_n,$$

we see that $\{x_n\}$ is a τ_1-Cauchy sequence. Since (X_1, τ_1) is τ_1-complete, there is some $x \in X_1$ such that $x_n \xrightarrow{\tau_1} x$. Rewriting ($\star$) as $y - T(x_n) \in V_{n+1}$ for each n, we see that $y - T(x_n) \xrightarrow{\tau_2} 0$ in X_2. On the other hand, the continuity of T yields $T(x_n) \xrightarrow{\tau_2} T(x)$, and from this we get $y = T(x)$.

Finally, from $x_n = \sum_{i=1}^{n} w_i \in W_1 + W_2 + \cdots + W_n \subset W_1 + W_1 \subset U$ and the τ_1-closedness of U, we easily infer that $x \in U$, so $y = T(x) \in T(U)$. In other words $V_1 \subset T(U)$, and the proof is finished. ∎

5.19 Corollary *A surjective continuous one-to-one operator between completely metrizable topological vector spaces is a homeomorphism.*

Recall that the **graph** of a function $f : A \to B$ is simply the subset of the Cartesian product $A \times B$ defined by

$$\operatorname{Gr} f = \{(a, f(a)) : a \in A\}.$$

Notice that if $T : X \to Y$ is an operator between vector spaces, then the graph $\operatorname{Gr} T$ of T is a vector subspace of $X \times Y$.

5.20 The Closed Graph Theorem *An operator between completely metrizable topological vector spaces is continuous if and only if it has closed graph.*

Proof: Assume that $T : (X_1, \tau_1) \to (X_2, \tau_2)$ is an operator between completely metrizable topological vector spaces such that its graph $\operatorname{Gr} T = \{(x, T(x)) : x \in X_1\}$ is a closed subspace of $X_1 \times X_2$. It follows that $\operatorname{Gr} T$ (with the induced product topology from $X_1 \times X_2$) is also a completely metrizable topological vector space. Since the mapping $S : \operatorname{Gr} T \to X_1$ defined by $S(x, T(x)) = x$ is a surjective continuous one-to-one operator, it follows from Corollary 5.19 that S is a homeomorphism. In particular, the operator $x \mapsto (x, T(x)) = S^{-1}(x)$, from X_1 to $\operatorname{Gr} T$, is continuous. Since the projection $P_2 : X_1 \times X_2 \to X_2$, defined by $P_2(x_1, x_2) = x_2$, is continuous it follows that the operator $T = P_2 S^{-1}$ is likewise continuous. ∎

5.5 Finite dimensional topological vector spaces

This section presents some distinguishing properties of finite dimensional vector spaces. Recall that the **Euclidean norm** $\|\cdot\|_2$ on $\mathbb{R}^n$ is defined by $\|x\|_2 = (\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}}$. It generates the **Euclidean topology**. Remarkably, this is the only Hausdorff linear topology on $\mathbb{R}^n$. In particular, any two norms on a finite dimensional vector space are equivalent: Two norms $\|\cdot\|$ and $|||\cdot|||$ on a vector space X are **equivalent** if they generate the same topology. In view of Theorem 6.17, this occurs if and only if there exist two positive constants K and M satisfying $K\|x\| \leqslant |||x||| \leqslant M\|x\|$ for each $x \in X$.

5.21 Theorem *Every finite dimensional vector space admits a unique Hausdorff linear topology, namely the complete Euclidean topology.*

Proof: Let $X = \mathbb{R}^n$, let τ_1 be a Hausdorff linear topology on X, and let τ denote the linear topology generated by the Euclidean norm $\|\cdot\|_2$. Clearly, (X, τ) is topologically complete.

We know that a net $\{x_\alpha = (x_1^\alpha, \dots, x_n^\alpha)\}$ in $\mathbb{R}^n$, satisfies $x_\alpha \xrightarrow{\|\cdot\|_2} 0$ if and only if $x_i^\alpha \xrightarrow[\alpha]{} 0$ in $\mathbb{R}$ for each i. Thus, if $x_\alpha \xrightarrow{\|\cdot\|_2} 0$, then since addition and scalar multiplication are τ_1-continuous,

$$x_\alpha = \sum_{i=1}^{n} x_i^\alpha e_i \xrightarrow[\alpha]{\tau_1} \sum_{i=1}^{n} 0 e_i = 0,$$

where as usual, e_i denotes the i coordinate unit vector of $\mathbb{R}^n$. Thus, the identity $I\colon (X, \tau) \to (X, \tau_1)$ is continuous and so $\tau_1 \subset \tau$.

Now let $B = \{x \in X : \|x\|_2 < 1\}$. Since $S = \{x \in X : \|x\|_2 = 1\}$ is τ-compact, it follows from $\tau_1 \subset \tau$ that S is also τ_1-compact. Therefore (since τ_1 is Hausdorff) S is τ_1-closed. Since $0 \notin S$, we see that there exists a circled τ_1-neighborhood V of zero such that $V \cap S = \varnothing$. Since V is circled, we have $V \subset B$: For if there exists some $x \in V$ such that $x \notin B$ (that is, $\|x\|_2 \geqslant 1$), then $\frac{x}{\|x\|_2} \in V \cap S$, a contradiction.

Thus, B is a τ_1-neighborhood of zero. Since scalar multiples of B form a τ-neighborhood base at zero, we see that $\tau \subset \tau_1$. Therefore $\tau_1 = \tau$. ∎

When we deal with finite dimensional vector spaces, we shall assume tacitly (and without any specific mention) that they are equipped with their Euclidean topologies and all topological notions will be understood in terms of Euclidean topologies.

The remaining results in this section are consequences of Theorem 5.21.

5.22 Corollary *A finite dimensional vector subspace of a Hausdorff topological vector space is closed.*

Proof: Let Y be a finite dimensional subspace of a Hausdorff topological vector space (X, τ), and let $\{y_\alpha\}$ be a net in Y satisfying $y_\alpha \xrightarrow{\tau} x$ in X. Therefore it is a Cauchy net in X, and hence also in Y. By Theorem 5.21, τ induces the Euclidean topology on Y. Since Y (with its Euclidean metric) is a complete metric space, it follows that $y_\alpha \xrightarrow{\tau} y$ in Y. Since τ is Hausdorff, we see that $x = y \in Y$, so Y is a closed subspace of X. ∎

5.23 Corollary *Every Hamel basis of an infinite dimensional completely metrizable topological vector space is uncountable.*

Proof: Let $\{e_1, e_2, \ldots\}$ be a countable Hamel basis of an infinite dimensional completely metrizable tvs X. For each n let X_n be the finite dimensional vector subspace generated by $\{e_1, \ldots, e_n\}$. By Theorem 5.21 each X_n is closed. Now note that $X = \bigcup_{n=1}^{\infty} X_n$ and then use the Baire Category Theorem 3.47 to conclude that some X_n has a nonempty interior. This implies $X = X_n$ for some n, which is impossible. ∎

5.24 Corollary *Let $v_1, v_2, \ldots, v_m$ be linearly independent vectors in a Hausdorff topological vector space (X, τ). For each n let $x_n = \sum_{i=1}^m \lambda_i^n v_i$. If $x_n \xrightarrow{\tau} x$ in X, then there exist $\lambda_1, \ldots, \lambda_m$ such that $x = \sum_{i=1}^m \lambda_i v_i$ (that is, x is in the linear span of $\{v_1, \ldots, v_m\}$) and $\lambda_i^n \xrightarrow[n\to\infty]{} \lambda_i$ for each i.*

Proof: Let Y be the linear span of $\{v_1, \ldots, v_m\}$. By Corollary 5.22, Y is a closed vector subspace of X, so $x \in Y$. That is, there exist scalars $\lambda_1, \ldots, \lambda_m$ such that $x = \sum_{i=1}^m \lambda_i v_i$.

Now for each $y = \sum_{i=1}^m \alpha_i v_i \in Y$, let $\|y\| = \sum_{i=1}^m |\alpha_i|$. Then $\|\cdot\|$ is a norm on Y, and thus (by Theorem 5.21) the topology induced by τ on Y coincides with the topology generated by the norm $\|\cdot\|$ on Y. Now note that

$$\|x_n - x\| = \Big\| \sum_{i=1}^m \lambda_i^n v_i - \sum_{i=1}^m \lambda_i v_i \Big\| = \sum_{i=1}^m |\lambda_i^n - \lambda_i| \xrightarrow[n\to\infty]{} 0$$

if and only if $\lambda_i^n \xrightarrow[n\to\infty]{} \lambda_i$ for each i. ∎

A **ray** in a vector space X is the set of nonnegative multiples of some vector, that is, a set of the form $\{\alpha v : \alpha \geqslant 0\}$, where $v \in X$. It is **trivial** if it contains only zero. We may also refer to a translate of such a set as a ray or a **half line**. A **cone** is a set of rays, or in other words a set that contains every nonnegative multiple of each of its members. That is, C is a cone if $x \in C$ implies $\alpha x \in C$ for every $\alpha \geqslant 0$.[1] In particular, we consider linear subspaces to be cones. A cone is **pointed** if it includes no lines. (A **line** is a translate of a one-dimensional subspace, that is, a set of the form $\{x + \alpha v : \alpha \in \mathbb{R}\}$, where $x, v \in X$ and $v \neq 0$.)

Let S be a nonempty subset of a vector space. The **cone generated by** S is the smallest cone that includes S and is thus $\{\alpha x : \alpha \geqslant 0 \text{ and } x \in S\}$. The **convex cone generated by** S is the smallest convex cone generated by S. You should verify that it consists of all nonnegative linear combinations from S.

5.25 Corollary *In a Hausdorff topological vector space, the convex cone generated by a finite set is closed.*

[1] Some authors, notably R. T. Rockafellar [288] and G. Choquet [76], define a cone to be a set closed under multiplication by strictly positive scalars. The point zero may or may not belong to such a cone. Other authorities, e.g., W. Fenchel [123] and D. Gale [133] use our definition.

Proof: Let $S = \{x_1, x_2, \dots, x_k\}$ be a nonempty finite subset of a Hausdorff topological vector space X. Then the convex cone K generated by S is given by

$$K = \Big\{\sum_{i=1}^{k} \lambda_i x_i : \lambda_i \geqslant 0 \text{ for each } i\Big\}.$$

Now fix a nonzero $x = \sum_{i=1}^{k} \lambda_i x_i \in K$. We claim that there is a linearly independent subset T of S and nonnegative scalars $\{\beta_t : t \in T\}$ such that $x = \sum_{t\in T} \beta_t t$.

To see this, start by noticing that we can assume that $\lambda_i > 0$ for each i; otherwise drop the terms with $\lambda_i = 0$. Now if the set S is linearly independent, then there is nothing to prove. So assume that S is linearly dependent. This means that there exist scalars $\alpha_1, \dots, \alpha_k$, not all zero, such that $\sum_{i=1}^{k} \alpha_i x_i = 0$. We can assume that $\alpha_i > 0$ for some i; otherwise multiply them by -1. Now let $\mu = \max\{\frac{\alpha_i}{\lambda_i} : i = 1, \dots, k\}$, and notice that $\mu > 0$. In particular, we have $\lambda_i \geqslant \frac{1}{\mu}\alpha_i$ for each i and $\lambda_i = \frac{1}{\mu}\alpha_i$ for some i. This implies that

$$x = \sum_{i=1}^{k} \lambda_i x_i = \sum_{i=1}^{k} \lambda_i x_i - \frac{1}{\mu}\sum_{i=1}^{k} \alpha_i x_i = \sum_{i=1}^{k} (\lambda_i - \frac{1}{\mu}\alpha_i) x_i$$

is a linear combination of the x_i with nonnegative coefficients, and one of them is zero. In other words, we have shown that if the set S is not a linearly independent set, then we can write x as a linear combination with positive coefficients of at most $k-1$ vectors of S. Our claim can now be completed by repeating this process.

Now assume that a sequence $\{y_n\}$ in K satisfies $y_n \to y$ in X. Since the collection of all linearly independent subsets of S is a finite set, by the above discussion, there exist a linearly independent subset of S, say $\{z_1, \dots, z_m\}$, and a subsequence of $\{y_n\}$, which we shall denote by $\{y_n\}$ again, such that

$$y_n = \sum_{i=1}^{m} \mu_i^n z_i$$

with all coefficients μ_i^n nonnegative. It follows from Corollary 5.24 that y belongs to K, so K is closed. ∎

There are no infinite dimensional locally compact Hausdorff topological vector spaces. This is essentially due to F. Riesz.

5.26 Theorem (F. Riesz) *A Hausdorff topological vector space is locally compact if and only if is finite dimensional.*

Proof: Let (X, τ) be a Hausdorff topological vector space. If X is finite dimensional, then τ coincides with the Euclidean topology and since the closed balls are compact sets, it follows that (X, τ) is locally compact.

For the converse assume that (X, τ) is locally compact and let V be a τ-compact neighborhood of zero. From $V \subset \bigcup_{x \in V}(x + \frac{1}{2}V)$, we see that there exists a finite subset $\{x_1, \ldots, x_k\}$ of V such that

$$V \subset \bigcup_{i=1}^{k}(x_i + \tfrac{1}{2}V) = \{x_1, \ldots, x_k\} + \tfrac{1}{2}V. \qquad (\star)$$

Let Y be the linear span of $x_1, \ldots, x_k$. From $(\star)$, we get $V \subset Y + \frac{1}{2}V$. This implies $\frac{1}{2}V \subset \frac{1}{2}(Y + \frac{1}{2}V) = Y + \frac{1}{2^2}V$, so $V \subset Y + (Y + \frac{1}{2^2}V) = Y + \frac{1}{2^2}V$. By induction we see that

$$V \subset Y + \frac{1}{2^n}V \qquad (\star\star)$$

for each n. Next, fix $x \in V$. From $(\star\star)$, it follows that for each n there exist $y_n \in Y$ and $v_n \in V$ such that $x = y_n + \frac{1}{2^n}v_n$. Since V is τ-compact, there exists a subnet $\{v_{n_\alpha}\}$ of the sequence $\{v_n\}$ such that $v_{n_\alpha} \xrightarrow{\tau} v$ in X (and clearly $\frac{1}{2^{n_\alpha}} \to 0$ in $\mathbb{R}$). So

$$y_{n_\alpha} = x - \tfrac{1}{2^{n_\alpha}}v_{n_\alpha} \xrightarrow{\tau} x - 0v = x.$$

Since (by Corollary 5.22) Y is a closed subspace, $x \in Y$. That is, $V \subset Y$. Since V is also an absorbing set, it follows that $X = Y$, so that X is finite dimensional. ∎

5.6 Convex sets

Recall that a subset of a vector space is **convex** if it includes the line segment joining any two of its points. Or in other words, a set C is convex if whenever $x, y \in C$, the line segment $\{\alpha x + (1 - \alpha)y : \alpha \in [0, 1]\}$ is included in C. By induction, a set C is convex if and only if for every finite subset $\{x_1, \ldots, x_n\}$ of C and nonnegative scalars $\{\alpha_1, \ldots, \alpha_n\}$ with $\sum_{i=1}^n \alpha_i = 1$, the linear combination $\sum_{i=1}^n \alpha_i x_i$ lies in C. Such a linear combination is called a **convex combination**, and the coefficients may be called **weights**.

The next lemma presents some elementary properties of convex sets.

5.27 Lemma *In any vector space:*

1. *The sum of two convex sets is convex.*
2. *Scalar multiples of convex sets are convex.*
3. *A set C is convex if and only if $\alpha C + \beta C = (\alpha + \beta)C$ for all nonnegative scalars α and β.*
4. *The intersection of an arbitrary family of convex sets is convex.*
5. *A convex set containing zero is circled if and only if it is symmetric.*
6. *In a topological vector space, both the interior and the closure of a convex set are convex.*

Proof: We prove only the first part of the last claim and leave the proofs of everything else as an exercise.

Let C be a convex subset of a tvs and let $0 \leqslant \alpha \leqslant 1$. Since C° is an open set, the set $\alpha C^\circ + (1-\alpha)C^\circ$ is likewise open. (Why?) The convexity of C implies $\alpha C^\circ + (1-\alpha)C^\circ \subset C$. Since C° is the largest open set included in C, we see that $\alpha C^\circ + (1-\alpha)C^\circ \subset C^\circ$. This shows that C° is convex. ∎

In topological vector spaces we can say a little bit more about the interior and closure of a convex set.

5.28 Lemma *If C is a convex subset of a tvs, then:*

$$0 < \alpha \leqslant 1 \quad \Longrightarrow \quad \alpha C^\circ + (1-\alpha)\overline{C} \subset C^\circ. \tag{$\star$}$$

In particular, if $C^\circ \neq \varnothing$, then:

1. *The interior of C is dense in $\overline{C}$, that is, $\overline{C^\circ} = \overline{C}$.*
2. *The interior of $\overline{C}$ coincides with the interior of C, that is, $\overline{C}^\circ = C^\circ$.*

Proof: The case $\alpha = 1$ in ($\star$) is immediate. So let C be convex, $x \in C^\circ$, $y \in \overline{C}$, and let $0 < \alpha < 1$. Choose an open neighborhood U of zero such that $x + U \subset C$. Since $y - \frac{\alpha}{1-\alpha}U$ is a neighborhood of y, there is some $z \in C \cap (y - \frac{\alpha}{1-\alpha}U)$, so that $(1-\alpha)(y-z)$ belongs to αU. Since C is convex, the (nonempty) open set $V = \alpha(x+U) + (1-\alpha)z = \alpha x + \alpha U + (1-\alpha)z$ lies entirely in C. Moreover, from

$$\alpha x + (1-\alpha)y = \alpha x + (1-\alpha)(y-z) + (1-\alpha)z \in \alpha x + \alpha U + (1-\alpha)z = V \subset C,$$

we see that $\alpha x + (1-\alpha)y \in C^\circ$. This proves ($\star$), and letting $\alpha \to 0$ proves (1).

For (2), fix $x_0 \in C^\circ$ and $x \in \overline{C}^\circ$. Pick a neighborhood W of zero satisfying $x + W \subset \overline{C}$. Since W is absorbing, there is some $0 < \varepsilon < 1$ such that $\varepsilon(x - x_0) \in W$, so $x + \varepsilon(x - x_0) \in \overline{C}$. By ($\star$), we have $x - \varepsilon(x - x_0) = \varepsilon x_0 + (1-\varepsilon)x \in C^\circ$. But then, using ($\star$) once more, we obtain $x = \frac{1}{2}[x - \varepsilon(x-x_0)] + \frac{1}{2}[x + \varepsilon(x - x_0)] \in C^\circ$. Therefore, $\overline{C}^\circ \subset C^\circ \subset \overline{C}^\circ$ so that $\overline{C}^\circ = C^\circ$. ∎

Note that a convex set with an empty interior may have a closure with a nonempty interior. For instance, any dense (proper) vector subspace has this property.

The **convex hull** of a set A, denoted $\operatorname{co} A$, consists precisely of all convex combinations from A. That is,

$$\operatorname{co} A = \Big\{x : \exists\, x_i \in A,\ \alpha_i \geqslant 0\,(1 \leqslant i \leqslant n), \sum_{i=1}^n \alpha_i = 1, \text{ and } x = \sum_{i=1}^n \alpha_i x_i\Big\}.$$

The convex hull $\operatorname{co} A$ is the smallest convex set including A and by Lemma 5.27 (4) is the intersection of all convex sets that include A. In a topological vector space, the **closed convex hull** of a set A, denoted $\overline{\mathrm{co}}\, A$, is the smallest closed convex set including A. By Lemma 5.27 (6) it is the closure of $\mathrm{co}\, A$, that is, $\overline{\mathrm{co}}\, A = \overline{\mathrm{co}\, A}$.

The next lemma presents further results on the relationship between topological and convexity properties. The **convex circled hull** of a subset A of a vector space is the smallest convex and circled set that includes A. It is the intersection of all convex and circled sets that include A. The **closed convex circled hull** of A is the smallest closed convex circled set including A. It is the closure of the convex circled hull of A.

5.29 Lemma *For nonempty convex sets $A_1, \dots, A_n$ in a tvs we have:*

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

$$\mathrm{co}\Big(\bigcup_{i=1}^n A_i\Big) = \Big\{\sum_{i=1}^n \lambda_i x_i : \lambda_i \geqslant 0,\ x_i \in A_i,\ \textit{and} \sum_{i=1}^n \lambda_i = 1\Big\}.$$

In particular, if each A_i is also compact, then $\mathrm{co}(\bigcup_{i=1}^n A_i)$ is compact.

2. *If, in addition, each A_i is circled, then the convex circled hull of the union $\bigcup_{i=1}^n A_i$ is the set*

$$\Big\{\sum_{i=1}^n \lambda_i x_i : \lambda_i \in \mathbb{R},\ x_i \in A_i,\ \textit{and} \sum_{i=1}^n |\lambda_i| \leqslant 1\Big\}.$$

Furthermore, if each A_i is also compact, then the convex circled hull of $\bigcup_{i=1}^n A_i$ is compact.

Proof: Let X be a vector space and let $A_1, \dots, A_n$ be convex subsets of X. You can easily verify that the indicated sets coincide with the convex and convex circled hull of the union $\bigcup_{i=1}^n A_i$, respectively.

Now let X be equipped with a linear topology. Consider the compact sets

$$C = \Big\{\lambda \in \mathbb{R}^n_+ : \sum_{i=1}^n \lambda_i = 1\Big\} \quad \text{and} \quad K = \Big\{\lambda \in \mathbb{R}^n : \sum_{i=1}^n |\lambda_i| \leqslant 1\Big\}.$$

Define the continuous function $f\colon \mathbb{R}^n \times A_1 \times \cdots \times A_n \to X$ by

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

The compactness assertions follow from the fact that the continuous image of a compact set is compact, and the observations that

$$\mathrm{co}\Big(\bigcup_{i=1}^n A_i\Big) = f(C \times A_1 \times \cdots \times A_n)$$

and that $f(K \times A_1 \times \cdots \times A_n)$ is the convex circled hull of $\bigcup_{i=1}^n A_i$. ∎

The convexity of the sets A_i is crucial for the results in Lemma 5.29; see Example 5.34 below. Here are some straightforward corollaries. The convex hull of a finite set is called a **polytope**.

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

5.31 Corollary *The convex circled hull of a compact convex subset of a tvs is compact.*

Proof: Note that if C is a compact convex set, then its convex circled hull coincides with the convex circled hull of $C \cup (-C)$. ∎

In finite dimensional vector spaces, the convex hull of a set is characterized by the celebrated Carathéodory convexity theorem.

5.32 Carathéodory's Convexity Theorem *In an n-dimensional vector space, every vector in the convex hull of a nonempty set can be written as a convex combination using no more than $n+1$ vectors from the set.*

Proof: Let A be a nonempty subset of some n-dimensional vector space, and let $x \in \operatorname{co} A$. Consider the nonempty set of natural numbers

$$S = \{\ell \in \mathbb{N} : x \text{ is a convex combination of some } \ell \text{ vectors from } A\},$$

and let $k = \min S$. We must show that $k \leqslant n+1$.

Assume by way of contradiction that $k > n+1$. Pick $x_1, \dots, x_k \in A$ and positive constants $\alpha_1, \dots, \alpha_k$ with $\sum_{i=1}^k \alpha_i = 1$ and $x = \sum_{i=1}^k \alpha_i x_i$. Since $k-1 > n$, the $k-1$ vectors $x_2 - x_1, x_3 - x_1, \dots, x_k - x_1$ of the n-dimensional vector space X must be linearly dependent. Consequently, there exist scalars $\lambda_2, \lambda_3, \dots, \lambda_k$, not all zero, such that $\sum_{i=2}^k \lambda_i (x_i - x_1) = 0$. Letting $c_1 = -\sum_{i=2}^k \lambda_i$ and $c_i = \lambda_i$ $(i = 2, 3, \dots, k)$, we see that not all the c_i are zero and satisfy

$$\sum_{i=1}^k c_i x_i = 0 \quad \text{and} \quad \sum_{i=1}^k c_i = 0.$$

Without loss of generality we can assume that $c_j > 0$ for some j. Next, put $c = \min\{\alpha_i / c_i : c_i > 0\}$, and pick some m with $\alpha_m / c_m = c > 0$. Note that

1. $\alpha_i - c c_i \geqslant 0$ for each i and $\alpha_m - c c_m = 0$; and
2. $\sum_{i=1}^k (\alpha_i - c c_i) = 1$ and $x = \sum_{i=1}^k (\alpha_i - c c_i) x_i$.

The above shows that x can be written as a convex combination of fewer than k vectors of A, contrary to the definition of k. ∎

Since continuous images of compact sets are compact, Carathéodory's theorem immediately implies the following. (Cf. proof of Lemma 5.29.)

5.33 Corollary *The convex hull and the convex circled hull of a compact subset of a finite dimensional vector space are compact sets.*

The convex hull of a compact subset of an infinite dimensional topological vector space need not be a compact set.

5.34 Example (Noncompact convex hull) Consider ℓ_2, the space of all square summable sequences. For each n let $u_n = (\underbrace{0,\ldots,0}_{n-1},\frac{1}{n},0,0,\ldots)$. Observe that $\|u_n\|_2 = \frac{1}{n}$, so $u_n \to 0$. Consequently,

$$A = \{u_1, u_2, u_3, \ldots\} \cup \{0\}$$

is a norm compact subset of ℓ_2. Since $0 \in A$, it is easy to see that

$$\operatorname{co} A = \Big\{\sum_{i=1}^{k} \alpha_i u_i : \alpha_i \geqslant 0 \text{ for each } i \text{ and } \sum_{i=1}^{k} \alpha_i \leqslant 1\Big\}.$$

In particular, each vector of $\operatorname{co} A$ has only finitely many nonzero components. We claim that $\operatorname{co} A$ is not norm compact. To see this, set

$$x_n = \left(\tfrac{1}{2}, \tfrac{1}{2}\cdot\tfrac{1}{2^2}, \tfrac{1}{3}\cdot\tfrac{1}{2^3}, \ldots, \tfrac{1}{n}\cdot\tfrac{1}{2^n}, 0, 0, \ldots\right) = \sum_{i=1}^{n} \tfrac{1}{2^i} u_i,$$

so $x_n \in \operatorname{co} A$. Now $x_n \xrightarrow{\|\cdot\|_2} x = \left(\frac{1}{2}, \frac{1}{2}\frac{1}{2^2}, \frac{1}{3}\cdot\frac{1}{2^3}, \ldots, \frac{1}{n}\cdot\frac{1}{2^n}, \frac{1}{n+1}\cdot\frac{1}{2^{n+1}}, \ldots\right)$ in ℓ_2. But $x \notin \operatorname{co} A$, so $\operatorname{co} A$ is not even closed, let alone compact.

In this example, the convex hull of a compact set failed to be closed. The question remains as to whether the closure of the convex hull is compact. In general, the answer is no. To see this, let X be the space of sequences that are eventually zero, equipped with the ℓ_2-norm. Let A be as above, and note that $\overline{\operatorname{co}}\, A$ (where the closure is taken in X, not ℓ_2) is not compact either. To see this, observe that the sequence $\{x_n\}$ defined above has no convergent subsequence (in X). ∎

However there are three important cases when the closed convex hull of a compact set is compact. The first is when the compact set is a finite union of compact convex sets. This is just Lemma 5.29. The second is when the space is completely metrizable and locally convex. This includes the case of all Banach spaces with their norm topologies. Failure of completeness is where the last part of Example 5.34 goes awry. The third case is a compact set in the weak topology on a Banach space; this is the Krein–Šmulian Theorem 6.35 ahead. Here is the proof for the completely metrizable locally convex case.

5.35 Theorem (Closed convex hull of a compact set) *In a completely metrizable locally convex space, the closed convex hull of a compact set is compact.*

Proof: Let K be compact subset of a completely metrizable locally convex space X. By Theorem 5.10 the topology is generated by some compatible complete metric d. By Theorem 3.28, it suffices to prove that $\overline{\text{co}}\, K$ is d-totally bounded. So let $\varepsilon > 0$ be given. By local convexity there is a convex neighborhood V of zero satisfying $V + V \subset B_\varepsilon$, the d-open ball of radius ε at zero. Since K is compact, there is a finite set Φ with $K \subset \Phi + V$. Clearly, $\text{co}\, K \subset \text{co}\, \Phi + V$. (Why?) By Corollary 5.30, $\text{co}\, \Phi$ is compact, so there is a finite set F satisfying $\text{co}\, \Phi \subset F + V$. Therefore

$$\text{co}\, K \subset \text{co}\, \Phi + V \subset F + V + V \subset F + B_\varepsilon.$$

Thus $\text{co}\, K$, and hence $\overline{\text{co}}\, K$, is d-totally bounded. ∎

Note that the proof above does not require the entire space to be completely metrizable. The same argument works provided $\overline{\text{co}}\, K$ lies in a subset of a locally convex space that is completely metrizable.

Finally, we shall present a case where the convex hull of the union of two closed convex sets is closed. But first, we need a definition.

5.36 Definition *A subset A of a topological vector space (X, τ) is **(topologically) bounded**, or more specifically **τ-bounded**, if for each neighborhood V of zero there exists some $\lambda > 0$ such that $A \subset \lambda V$.*

Observe that for a normed space, the topologically bounded sets coincide with the norm bounded sets. Also, notice that if $\{x_\alpha\}$ is a topologically bounded net in a tvs and $\lambda_\alpha \to 0$ in $\mathbb{R}$, then $\lambda_\alpha x_\alpha \to 0$.

5.37 Lemma *If A and B are two nonempty convex subsets of a Hausdorff topological vector space such that A is compact and B is closed and bounded, then $\text{co}(A \cup B)$ is closed.*

Proof: Let $z_\alpha = (1 - \lambda_\alpha)x_\alpha + \lambda_\alpha y_\alpha \to z$, where $0 \leqslant \lambda_\alpha \leqslant 1$, $x_\alpha \in A$, and $y_\alpha \in B$ for each α. By passing to a subnet, we can assume that $x_\alpha \to x \in A$ and $\lambda_\alpha \to \lambda \in [0, 1]$. If $\lambda > 0$, then $y_\alpha \to \frac{z-(1-\lambda)x}{\lambda} = y \in B$, and consequently $z = (1 - \lambda)x + \lambda y \in \text{co}(A \cup B)$.

Now consider the case $\lambda = 0$. The boundedness of B implies $\lambda_\alpha y_\alpha \to 0$, so $z_\alpha = (1 - \lambda_\alpha)x_\alpha + \lambda_\alpha y_\alpha \to x$. Since the space is Hausdorff, $z = x \in \text{co}(A \cup B)$. ∎

5.7 Convex and concave functions

The interaction of the algebraic and topological structure of a topological vector space is manifested in the properties of the important class of convex functions. The definition is purely algebraic.

5.38 Definition *A function $f\colon C \to \mathbb{R}$ on a convex set C in a vector space is:*

- ***convex** if $f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y)$ for all $x, y \in C$ and all $0 \leqslant \alpha \leqslant 1$.*

- ***strictly convex** if $f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$ for all $x, y \in C$ with $x \neq y$ and all $0 < \alpha < 1$.*

- ***concave** if $-f$ is a convex function.*

- ***strictly concave** if $-f$ is strictly convex.*

Note that a real function f on a convex set is convex if and only if

$$f\Big(\sum_{i=1}^{n} \alpha_i x_i\Big) \leqslant \sum_{i=1}^{n} \alpha_i f(x_i)$$

for every convex combination $\sum_{i=1}^{n} \alpha_i x_i$.

You may verify the following lemma.

5.39 Lemma *A function $f\colon C \to \mathbb{R}$ on a convex subset of a vector space is convex if and only if its **epigraph**, $\{(x,\alpha) \in C \times \mathbb{R} : \alpha \geqslant f(x)\}$, is convex. Similarly, f is concave if and only if its **hypograph**, $\{(x,\alpha) \in C \times \mathbb{R} : \alpha \leqslant f(x)\}$, is convex.*

Some important properties of convex functions are immediate consequences of the definition. There is of course a corresponding lemma for concave functions. We omit it.

5.40 Lemma *The collection of convex functions on a fixed convex set has the following properties.*

1. *Sums and nonnegative scalar multiples of convex functions are convex.*
2. *The (finite) pointwise limit of a net of convex functions is convex.*
3. *The (finite) pointwise supremum of a family of convex functions is convex.*

The next simple inequality is useful enough that it warrants its own lemma. It requires no topology.

5.41 Lemma *Let $f\colon C \to \mathbb{R}$ be a convex function, where C is a convex subset of a vector space. Let x belong to C and suppose z satisfies $x + z \in C$ and $x - z \in C$. Let $\delta \in [0,1]$. Then*

$$\big|f(x+\delta z) - f(x)\big| \leqslant \delta \max\{f(x+z) - f(x), f(x-z) - f(x)\}$$

Proof: Now $x + \delta z = (1-\delta)x + \delta(x+z)$, so $f(x+\delta z) \leqslant (1-\delta)f(x) + \delta f(x+z)$. Rearranging terms yields

$$f(x+\delta z) - f(x) \leqslant \delta\,[f(x+z) - f(x)], \tag{1}$$

and replacing z by $-z$ gives

$$f(x-\delta z) - f(x) \leqslant \delta\,[f(x-z) - f(x)]. \tag{2}$$

Also, since $x = \frac{1}{2}(x+\delta z) + \frac{1}{2}(x-\delta z)$, we have $f(x) \leqslant \frac{1}{2}f(x+\delta z) + \frac{1}{2}f(x-\delta z)$. Multiplying by two and rearranging terms we obtain

$$f(x) - f(x+\delta z) \leqslant f(x-\delta z) - f(x). \tag{3}$$

Combining (2) and (3) yields

$$f(x) - f(x+\delta z) \leqslant f(x-\delta z) - f(x) \leqslant \delta\,[f(x-z) - f(x)].$$

This in conjunction with (1) yields the conclusion of the lemma. ∎

5.42 Theorem (Local continuity of convex functions) *Let $f\colon C \to \mathbb{R}$ be a convex function, where C is a convex subset of a topological vector space. If f is bounded above on a neighborhood of an interior point of C, then f is continuous at that point.*

Proof: Assume that for some $x \in C$ there exist a circled neighborhood V of zero and some $M > 0$ satisfying $x + V \subset C$ and $f(y) < f(x) + M$ for each $y \in x + V$. Fix $\varepsilon > 0$ and choose some $0 < \delta \leqslant 1$ so that $\delta M < \varepsilon$. But then if $y \in x + \delta V$, then from Lemma 5.41 it follows that for each $y \in x + \delta V$ we have $|f(y) - f(x)| < \varepsilon$. This shows that f is continuous at x. ∎

Amazingly, continuity at a single point implies global continuity for convex functions on open sets.

5.43 Theorem (Global continuity of convex functions) *For a convex function $f\colon C \to \mathbb{R}$ on an open convex subset of a topological vector space, the following statements are equivalent.*

1. *f is continuous on C.*
2. *f is upper semicontinuous on C.*
3. *f is bounded above on a neighborhood of each point in C.*
4. *f is bounded above on a neighborhood of some point in C.*
5. *f is continuous at some point in C.*

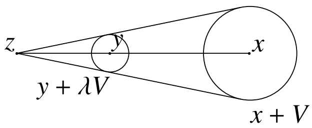
Proof: (1) $\implies$ (2) Obvious.

(2) $\implies$ (3) Assume that f is upper semicontinuous and $x \in C$. Then the set $\{y \in C : f(y) < f(x) + 1\}$ is an open neighborhood of x on which f is bounded.

(3) $\implies$ (4) This is trivial.

(4) $\implies$ (5) This is Theorem 5.42.

(5) $\implies$ (1) Suppose f is continuous at the point x, and let y be any other point in C. Since scalar multiplication is continuous, $\{\beta \in \mathbb{R} : x + \beta(y - x) \in C\}$ includes an open neighborhood of 1. This implies that there exist $z \in C$ and $0 < \lambda < 1$ such that $y = \lambda x + (1 - \lambda)z$.



Also, since f is continuous at x, there is a circled neighborhood V of zero such that $x + V \subset C$ and f is bounded above on $x + V$, say by μ. We claim that f is bounded above on $y + \lambda V$. To see this, let $v \in V$. Then $y+\lambda v = \lambda(x+v)+(1-\lambda)z \in C$. The convexity of f thus implies

$$f(y + \lambda v) \leqslant \lambda f(x + v) + (1 - \lambda)f(z) \leqslant \lambda\mu + (1 - \lambda)f(z).$$

That is, f is bounded above by $\lambda\mu + (1 - \lambda)f(z)$ on $y + \lambda V$. So by Theorem 5.42, f is continuous at y. ∎

If the topology of a tvs is generated by a norm, continuity of a convex function at an interior point implies local Lipschitz continuity. The proof of the next result is adapted from A. W. Roberts and D. E. Varberg [285].

5.44 Theorem *Let $f : C \to \mathbb{R}$ be convex, where C is a convex subset of a normed tvs. If f is continuous at the interior point x of C, then f is Lipschitz continuous on a neighborhood of x. That is, there exists $\delta > 0$ and $\mu > 0$, such that $B_\delta(x) \subset C$ and for $y, z \in B_\delta(x)$, we have*

$$|f(y) - f(z)| \leqslant \mu \|y - z\|.$$

Proof: Since f is continuous at x, there exists $\delta > 0$ such that $B_{2\delta}(x) \subset C$ and $w, z \in B_{2\delta}(x)$ implies $|f(w) - f(z)| < 1$. Given distinct y and z in $B_\delta(x)$, let $\alpha = \|y - z\|$ and let $w = y + \frac{\delta}{\alpha}(y - z)$, so $\|w - y\| = \frac{\delta}{\alpha}\|y - z\| = \delta$. Then w belongs to $B_{2\delta}(x)$ and we may write y as the convex combination $y = \frac{\alpha}{\alpha+\delta}w + \frac{\delta}{\alpha+\delta}z$. Therefore

$$f(y) \leqslant \frac{\alpha}{\alpha + \delta}f(w) + \frac{\delta}{\alpha + \delta}f(z).$$

Subtracting $f(z)$ from each side gives

$$f(y) - f(z) \leqslant \frac{\alpha}{\alpha + \delta}[f(w) - f(z)] < \frac{\alpha}{\alpha + \delta}.$$

Switching the roles of y and z allows us to conclude

$$|f(y) - f(z)| < \frac{\alpha}{\alpha+\delta} < \frac{\alpha}{\delta} = \frac{1}{\delta}\|y - z\|,$$

so $\mu = 1/\delta$ is the desired Lipschitz constant. ∎

We also point out that strictly convex functions on infinite dimensional spaces are quite special. In order for a continuous function to be strictly convex on a compact convex set, the relative topology of the set must be metrizable. This result relies on facts about metrizability of uniform spaces that we do not wish to explore, but if you are interested, see G. Choquet [76, p. II-139].

5.8 Sublinear functions and gauges

A real function f defined on a vector space is **subadditive** if

$$f(x + y) \leqslant f(x) + f(y)$$

for all x and y. Recall that a nonempty subset C of a vector space is a cone if $x \in C$ implies $\alpha x \in C$ for every $\alpha \geqslant 0$. A real function f defined on a cone C is **positively homogeneous** if

$$f(\alpha x) = \alpha f(x)$$

for every $\alpha \geqslant 0$. Clearly, if f is positively homogeneous, then $f(0) = 0$ and f is completely determined by its values on any absorbing set. In other words, two positively homogeneous functions are equal if and only if they agree on an absorbing set.

5.45 Definition *A real function on a vector space is* ***sublinear*** *if it is both positively homogeneous and subadditive, or equivalently, if it is both positively homogeneous and convex.*

To see the equivalence in the definition above, observe that for a subadditive positively homogeneous function f we have

$$f(\lambda x + (1-\lambda)y) \leqslant f(\lambda x) + f((1-\lambda x)) = \lambda f(x) + (1-\lambda)f(x),$$

so f is convex. Conversely, to see that a positively homogeneous convex function is subadditive, note that

$$f(x) + f(y) = \tfrac{1}{2}f(2x) + \tfrac{1}{2}f(2y) \leqslant f\Big(\tfrac{1}{2}2x + \tfrac{1}{2}2y\Big) = f(x + y).$$

Clearly every linear functional is sublinear, and so too is every norm. An important subclass of sublinear functions consists of functions called seminorms, which satisfy most of the properties norms, and which turn out to be crucial to the study of locally convex spaces.

5.46 Definition *A* ***seminorm*** *is a subadditive function $p\colon X \to \mathbb{R}$ on a vector space satisfying*

$$p(\alpha x) = |\alpha| p(x)$$

for all $\alpha \in \mathbb{R}$ and all $x \in X$.[2]

A seminorm p that satisfies $p(x) = 0$ if and only if $x = 0$ is called a ***norm****.*

Note that every seminorm is indeed sublinear, and every sublinear function satisfying $p(-x) = p(x)$ for all x is a seminorm. In particular, if f is a linear functional, then $p(x) = |f(x)|$ defines a seminorm. A seminorm p defines a semimetric d via $d(x, y) = p(x - y)$. If p is a norm, then the semimetric is actually a metric.

We now state some simple properties of sublinear functions. The proofs are left as exercises.

5.47 Lemma (Sublinearity) *If $p\colon X \to \mathbb{R}$ is sublinear, then:*

1. $p(0) = 0$.
2. *For all x we have $-p(x) \leqslant p(-x)$. Consequently p is linear if and only if $p(-x) = -p(x)$ for all $x \in X$.*
3. *The function q defined by $q(x) = \max\{p(x), p(-x)\}$ is a seminorm.*
4. *If p is a seminorm, then $p(x) \geqslant 0$ for all x.*
5. *If p is a seminorm, then the set $\{x : p(x) = 0\}$ is a linear subspace.*

We now come to the important class of Minkowski functionals, or gauges.

5.48 Definition *The* ***gauge****,*[3] *or the* ***Minkowski functional****, p_A, of a subset A of a vector space is defined by*

$$p_A(x) = \inf\{\alpha > 0 : x \in \alpha A\},$$

where, by convention, $\inf \varnothing = \infty$. In other words, $p_A(x)$ is the smallest factor by which the set A must be enlarged to contain the point x.

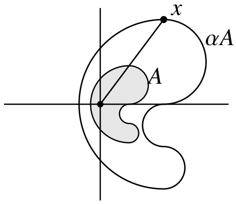


Figure 5.2. The gauge of A.

The next lemma collects a few elementary properties of gauges. The proof is left as an exercise.

[2] Be assured at once that, as we shall see in the following result, every seminorm $p\colon X \to \mathbb{R}$ satisfies $p(x) \geqslant 0$ for each $x \in X$.

[3] Dunford and Schwartz [110, p. 411] use the term support functional instead of gauge. We however have another, more standard, use in mind for the term support functional.

5.49 Lemma *For nonempty subsets B and C of a vector space X:*

1. $p_{-C}(x) = p_C(-x)$ *for all* $x \in X$.
2. *If C is symmetric, then* $p_C(x) = p_C(-x)$ *for all* $x \in X$.
3. $B \subset C$ *implies* $p_C \leqslant p_B$.
4. *If C includes a subspace M, then* $p_C(x) = 0$ *for all* $x \in M$.
5. *If C is star-shaped about zero, then*
$$\{x \in X : p_C(x) < 1\} \subset C \subset \{x \in X : p_C(x) \leqslant 1\}.$$
6. *If X is a tvs and C is closed and star-shaped about zero, then*
$$C = \{x \in X : p_C(x) \leqslant 1\}.$$
7. *If B and C are star-shaped about zero, then* $p_{B \cap C} = p_B \vee p_C$, *where as usual* $[p_B \vee p_C](x) = \max\{p_B(x), p_C(x)\}$.

Absorbing sets are of interest in part because any positively homogeneous function is completely determined by its values on any absorbing set.

5.50 Lemma *For a nonnegative function $p\colon X \to \mathbb{R}$ on a vector space we have the following.*

1. *p is positively homogeneous if and only if it is the gauge of an absorbing set—in which case for every subset A of X satisfying*
$$\{x \in X : p(x) < 1\} \subset A \subset \{x \in X : p(x) \leqslant 1\}$$
we have $p_A = p$.
2. *p is sublinear if and only if it is the gauge of a convex absorbing set C, in which case we may take* $C = \{x \in X : p(x) \leqslant 1\}$.
3. *p is a seminorm if and only if it is the gauge of a circled convex absorbing set C, in which case we may take* $C = \{x \in X : p(x) \leqslant 1\}$.
4. *When X is a tvs, p is a continuous seminorm if and only if it is the gauge of a unique closed, circled and convex neighborhood V of zero, namely* $V = \{x \in X : p(x) \leqslant 1\}$.
5. *When X is finite dimensional, p is a norm if and only if it is the gauge of a unique circled, convex and compact neighborhood V of zero, namely* $V = \{x \in X : p(x) \leqslant 1\}$.

Proof: (1) If $p = p_A$ for some absorbing subset A of X, then it is easy to see that p is positively homogeneous. For the converse, assume that p is positively homogeneous, and let A be any subset of X satisfying

$$\{x \in X : p(x) < 1\} \subset A \subset \{x \in X : p(x) \leqslant 1\}.$$

Clearly, A is an absorbing set, so $p_A \colon X \to \mathbb{R}$ is a nonnegative real-valued positively homogeneous function.

Now fix $x \in X$. If some $\alpha > 0$ satisfies $x \in \alpha A$, then pick some $u \in A$ such that $x = \alpha u$ and note that $p(x) = p(\alpha u) = \alpha p(u) \leqslant \alpha$. From this, we easily infer that $p(x) \leqslant p_A(x)$. On the other hand, the positive homogeneity of p implies that for each $\beta > p(x)$ we have $\frac{x}{\beta} \in A$ or $x \in \beta A$, so $p_A(x) \leqslant \beta$ for all $\beta > p(x)$. Hence $p_A(x) \leqslant p(x)$ is also true. Therefore $p_A(x) = p(x)$ for all $x \in X$.

(2) Let $p = p_C$, the gauge of the absorbing convex set C. Clearly p_C is nonnegative and positively homogeneous. For the subadditivity of p_C, let $\alpha, \beta > 0$ satisfy $x \in \alpha C$ and $y \in \beta C$. Then $x+y \in \alpha C + \beta C = (\alpha+\beta)C$, so $p_C(x+y) \leqslant \alpha+\beta$. Taking infima yields $p_C(x + y) \leqslant p_C(x) + p_C(y)$, so p_C is subadditive. For the converse, assume that p is a sublinear function. Let $C = \{x \in X : p(x) \leqslant 1\}$ and note that C is convex and absorbing. Now a glance at part (1) shows that $p = p_C$.

(3) Repeat the arguments of the preceding part.

(4) If p is a continuous seminorm, then the set $V = \{x \in X : p(x) \leqslant 1\}$ is a closed, circled and convex neighborhood of zero such that $p = p_V$. Conversely, if V is a closed, circled and convex neighborhood of zero and $p = p_V$, then p_V is (by part (3)) a seminorm. But then $p_V \leqslant 1$ on V and Theorem 5.43 guarantee that p is continuous.

For the uniqueness of the set V, assume that W is any other closed, circled and convex neighborhood of zero satisfying $p = p_V = p_W$. If $x \in W$, then $p(x) = p_W(x) \leqslant 1$, so $x \in V$. Therefore, $W \subset V$. For the reverse inclusion, let $x \in V$. This implies $p_W(x) = p_V(x) \leqslant 1$. If $p_W(x) < 1$, then pick $0 \leqslant \alpha < 1$ and $w \in W$ such $x = \alpha w$. Since W is circled, $x \in W$. On the other hand, if $p_W(x) = 1$, then pick a sequence $\{\alpha_n\}$ of real numbers and a sequence $\{w_n\} \subset W$ satisfying $\alpha_n \downarrow 1$ and $x = \alpha_n w_n$ for each n. But then $w_n = \frac{x}{\alpha_n} \to x$ and the closedness of W yield $x \in W$. Thus, $V \subset W$ is also true, so $W = V$.

(5) If p is a norm, then p generates the Euclidean topology on X, so the set $V = \{x \in X : p(x) \leqslant 1\}$ is circled, convex and compact neighborhood of zero and satisfies $p = p_V$. Its uniqueness should be obvious. On the other hand, if $p = p_V$, where $V = \{x \in X : p(x) \leqslant 1\}$ is a circled, convex and compact neighborhood of zero, then it is not difficult to see that the seminorm p is indeed a norm. ∎

The continuity of a sublinear functional is determined by its behavior near zero. Recall that a real function $f \colon D \to \mathbb{R}$ on a subset of a tvs is uniformly continuous on D if for every $\varepsilon > 0$, there is a neighborhood V of zero such that $|f(x) - f(y)| < \varepsilon$ whenever $x, y \in D$ satisfy $x - y \in V$.

5.51 Lemma *A sublinear function on a tvs is (uniformly) continuous if and only if it is bounded on some neighborhood of zero.*[4]

Proof: Let $h\colon X \to \mathbb{R}$ be a sublinear function on a tvs. Note that h is bounded on $h^{-1}((-1,1))$, which is a neighborhood of zero if h is continuous.

For the converse, continuity follows from Theorem 5.43, but uniform continuity is easy to prove directly. Assume that $|h(x)| < M$ for each x in some circled neighborhood V of zero. Note that for any x and y we have

$$h(x) = h(x - y + y) \leqslant h(x - y) + h(y),$$

so $h(x) - h(y) \leqslant h(x - y)$. In a similar fashion, $h(y) - h(x) \leqslant h(y - x)$. Thus, $|h(x) - h(y)| \leqslant \max\{h(x - y), h(y - x)\}$. So if $x - y \in \frac{\varepsilon}{M}V$, then $|h(x) - h(y)| < \varepsilon$, which shows that h is uniformly continuous. ∎

The next result elaborates on Lemma 5.50.

5.52 Theorem (Semicontinuity of gauges) *A nonnegative sublinear function on a topological vector space is:*

1. *Lower semicontinuous if and only if it is the gauge of an absorbing closed convex set.*

2. *Continuous if and only if it is the gauge of a convex neighborhood of zero.*

Proof: Let $p\colon X \to \mathbb{R}$ be a nonnegative sublinear function on a tvs.

(1) Suppose first that the function p is lower semicontinuous on X. Then $C = \{x \in X : p(x) \leqslant 1\}$ is absorbing, closed and convex. By Lemma 5.50, $p = p_C$, the gauge of C.

Let C be an arbitrary absorbing, closed and convex subset of X. Then for $0 < \alpha < \infty$ the lower contour set $\{x \in X : p_C(x) \leqslant \alpha\} = \alpha C$ (why?), which is closed. The set $\{x \in X : p_C(x) \leqslant 0\} = \bigcap_{\alpha>0} \alpha C$, which is closed, being the intersection of closed sets. Finally, $\{x \in X : p_C(x) \leqslant \alpha\}$ for $\alpha < 0$ is empty. Thus, p_C is lower semicontinuous.

(2) If p is continuous, then the set $C = \{x \in X : p(x) \leqslant 1\}$ includes the set $\{x \in X : p(x) < 1\}$, which is open. Thus C is a (closed) convex neighborhood of zero, and $p = p_C$. On the other hand, if C is a neighborhood of zero and $p = p_C$, then $p_C \leqslant 1$ on C, so by Lemma 5.51 it is continuous. ∎

[4] By Theorem 7.24, every sublinear function on a finite dimensional vector space is continuous, since it is convex.

5.9 The Hahn–Banach Extension Theorem

Let X^* denote the vector space of all linear functionals on the linear space X. The space X^* is called the **algebraic dual** of X to distinguish it from the **topological dual** X', the vector space of all continuous linear functionals on a tvs X.[5]

The algebraic dual X^* is in general very large. To get a feeling for its size, fix a Hamel basis $\mathcal{H}$ for X. Every $x \in X$ has a unique representation $x = \sum_{h \in \mathcal{H}} \lambda_h h$, where only a finite number of the λ_h are nonzero; see Theorem 1.8. If $f^* \in X^*$, then $f^*(x) = \sum_{h \in \mathcal{H}} \lambda_h f^*(h)$, so the action of f^* on X is completely determined by its action on $\mathcal{H}$. This implies that every $f \in \mathbb{R}^{\mathcal{H}}$ gives rise to a (unique) linear functional f^* on X via the formula $f^*(x) = \sum_{h \in \mathcal{H}} \lambda_h f(h)$. The mapping $f \mapsto f^*$ is a linear isomorphism from $\mathbb{R}^{\mathcal{H}}$ onto X^*, so X^* can be identified with $\mathbb{R}^{\mathcal{H}}$.[6] In general, when we use the term dual space, we mean the topological dual.

One of the most important and far-reaching results in analysis is the following seemingly mild theorem. It is usually stated for the case where p is sublinear, but this more general statement is as easy to prove. Recall that a real-valued function f **dominates** a real-valued function g on A if $f(x) \geqslant g(x)$ for all $x \in A$.

5.53 Hahn–Banach Extension Theorem *Let X be a vector space and let $p\colon X \to \mathbb{R}$ be any convex function. Let M be a vector subspace of X and let $f\colon M \to \mathbb{R}$ be a linear functional dominated by p on M. Then there is a (not generally unique) linear extension $\hat{f}$ of f to X that is dominated by p on X.*

---

[5] Be warned! Some authors use X' for the algebraic dual and X^* for the topological dual.

[6] This depends on the fact that any two Hamel bases $\mathcal{H}$ and $\mathcal{H}'$ of X have the same cardinality. From elementary linear algebra, we know that this is true if $\mathcal{H}$ is finite. We briefly sketch the proof of this claim when $\mathcal{H}$ and $\mathcal{H}'$ are infinite. The proof is based upon the fact that $\mathcal{H} \times \mathbb{N}$ has the same cardinality as $\mathcal{H}$. To see this, let $\mathcal{X}$ be the set of all pairs (S, f), where S is a nonempty subset of $\mathcal{H}$ and the function $f\colon S \times \mathbb{N} \to S$ is one-to-one and surjective. Since $\mathcal{X}$ contains the countable subsets of $\mathcal{H}$, the set $\mathcal{X}$ is nonempty. On $\mathcal{X}$ we define a partial order $\geqslant$ by letting $(S, f) \geqslant (T, g)$ whenever $S \supset T$ and $f = g$ on T. It is not difficult to see that $\geqslant$ is indeed a partial order on $\mathcal{X}$ and that every chain in $\mathcal{X}$ has an upper bound. By Zorn's Lemma 1.7, $\mathcal{X}$ has a maximal element, say (R, φ). We claim that $\mathcal{H} \setminus R$ is a finite set. Otherwise, if $\mathcal{H} \setminus R$ is an infinite set, then $\mathcal{H} \setminus R$ must include a countable subset A. Let $R' = R \cup A$ and fix any one-to-one and surjective function $g\colon A \times \mathbb{N} \to A$. Now define $\psi\colon R' \times \mathbb{N} \to R'$ by $\psi(r, n) = \varphi(r, n)$ if $(r, n) \in R \times \mathbb{N}$ and $\psi(a, n) = g(a, n)$ if $(a, n) \in A \times \mathbb{N}$. But then we have $(R', \psi) \in \mathcal{X}$ and $(R', \psi) > (R, \varphi)$, contrary to the maximality property of (R, φ). Therefore, $\mathcal{H} \setminus R$ is a finite set. Next, pick a countable set Y of R and fix a one-to-one and surjective function $h\colon [(\mathcal{H} \setminus R) \cup Y] \times \mathbb{N} \to [\varphi(Y \times \mathbb{N}) \cup (\mathcal{H} \setminus R)]$ and then define the function $\theta\colon \mathcal{H} \times \mathbb{N} \to \mathcal{H}$ by $\theta(x, n) = \varphi(x, n)$ if $(x, n) \in (R \setminus Y) \times \mathbb{N}$ and $\theta(x, n) = h(x, n)$ if $(x, n) \in [(\mathcal{H} \setminus R) \cup Y] \times \mathbb{N}$. Clearly, $\theta\colon \mathcal{H} \times \mathbb{N} \to \mathcal{H}$ is one-to-one and surjective.

For each $x \in \mathcal{H}$ there exists a unique nonempty finite subset $\mathcal{H}'(x) = \{y_1^x, \ldots, y_{k_x}^x\}$ of $\mathcal{H}'$ and nonzero scalars $\lambda_1^x, \ldots, \lambda_{k_x}^x$ such that $x = \sum_{i=1}^{k_x} \lambda_i^x y_i^x$. Since $\mathcal{H}$ and $\mathcal{H}'$ are Hamel bases, it follows that $\mathcal{H}' = \bigcup_{x \in \mathcal{H}} \mathcal{H}'(x)$. Now define the function $\alpha\colon \mathcal{H} \times \mathbb{N} \to \mathcal{H}'$ by $\alpha(x, n) = y_1^x$ if $n > k_x$ and $\alpha(x, n) = y_n^x$ if $1 \leqslant n \leqslant k_x$. Clearly, α is surjective and from this we infer that there exists a one-to-one function $\beta\colon \mathcal{H}' \to \mathcal{H} \times \mathbb{N}$. But then the scheme $\mathcal{H}' \xrightarrow{\beta} \mathcal{H} \times \mathbb{N} \xrightarrow{\theta} \mathcal{H}$ shows that $\mathcal{H}$ has cardinality at least as large as $\mathcal{H}'$. By symmetry, $\mathcal{H}'$ has cardinality at least as large as $\mathcal{H}$ and a glance at the classical Cantor–Schröder–Bernstein Theorem 1.2 shows that $\mathcal{H}$ and $\mathcal{H}'$ have the same cardinality.

Proof: The proof is an excellent example of what is known as transfinite induction. It has two parts. One part says that an extension of f whose domain is not all of X can be extended to a larger subspace and still satisfy $\hat{f} \leqslant p$. The second part says that this is enough to conclude that we can extend f all the way to X and still satisfy $\hat{f} \leqslant p$.

Let $f \leqslant p$ on the subspace M. If $M = X$, then we are done. So suppose there exists $v \in X \setminus M$. Let N be the linear span of $M \cup \{v\}$. For each $x \in N$ there is a unique decomposition of x of the form $x = z + \lambda v$ where $z \in M$. (To see the uniqueness, suppose $x = z_1 + \lambda_1 v = z_2 + \lambda_2 v$. Then $z_1 - z_2 = (\lambda_2 - \lambda_1)v$. Since $z_1 - z_2 \in M$ and $v \notin M$, it must be the case that $\lambda_2 - \lambda_1 = 0$. But then $\lambda_1 = \lambda_2$ and $z_1 = z_2$.)

Any linear extension $\hat{f}$ of f to N must satisfy $\hat{f}(z + \lambda v) = f(z) + \lambda \hat{f}(v)$. Thus what we need to show is that we can choose $c = \hat{f}(v) \in \mathbb{R}$ so that $\hat{f} \leqslant p$ on N. That is, we must demonstrate the existence of a real number c satisfying

$$f(z) + \lambda c \leqslant p(z + \lambda v) \tag{1}$$

for all $z \in M$ and all $\lambda \in \mathbb{R}$. It is a routine matter to verify that (1) is true if and only if there exists some real number c satisfying

$$\tfrac{1}{\lambda}[f(x) - p(x - \lambda v)] \leqslant c \leqslant \tfrac{1}{\mu}[p(y + \mu v) - f(y)] \tag{2}$$

for all $x, y \in M$ and all $\lambda, \mu > 0$. Now notice that (2) is true for some $c \in \mathbb{R}$ if and only if

$$\sup_{x \in M, \lambda > 0} \tfrac{1}{\lambda}[f(x) - p(x - \lambda v)] \leqslant \inf_{y \in M, \mu > 0} \tfrac{1}{\mu}[p(y + \mu v) - f(y)], \tag{3}$$

which is equivalent to

$$\tfrac{1}{\lambda}[f(x) - p(x - \lambda v)] \leqslant \tfrac{1}{\mu}[p(y + \mu v) - f(y)] \tag{4}$$

for all $x, y \in M$ and $\lambda, \mu > 0$. Rearranging terms, we see that (4) is equivalent to

$$f(\mu x + \lambda y) \leqslant \mu p(x - \lambda v) + \lambda p(y + \mu v) \tag{5}$$

for all $x, y \in M$ and all $\lambda, \mu > 0$. Thus, an extension of f to all of N exists if and only if (5) is valid. For the validity of (5) note that if $x, y \in M$ and $\lambda, \mu > 0$, then

$$\begin{aligned}
f(\mu x + \lambda y) &= (\lambda + \mu) f\Big(\frac{\mu}{\lambda + \mu} x + \frac{\lambda}{\lambda + \mu} y\Big) \\
&\leqslant (\lambda + \mu) p\Big(\frac{\mu}{\lambda + \mu} x + \frac{\lambda}{\lambda + \mu} y\Big) \\
&= (\lambda + \mu) p\Big(\frac{\mu}{\lambda + \mu}[x - \lambda v] + \frac{\lambda}{\lambda + \mu}[y + \mu v]\Big) \\
&\leqslant (\lambda + \mu)\Big[\frac{\mu}{\lambda + \mu} p(x - \lambda v) + \frac{\lambda}{\lambda + \mu} p(y + \mu v)\Big] \\
&= \mu p(x - \lambda v) + \lambda p(y + \mu v).
\end{aligned}$$

This shows that as long as there is some $v \notin M$, there is an extension of f to a larger subspace containing v that satisfies $\hat{f} \leqslant p$.

To conclude the proof, consider the set of all pairs (g, N) of partial extensions of f such that: N is a linear subspace of X with $M \subset N$, $g\colon N \to \mathbb{R}$ is a linear functional, $g|_M = f$, and $g(x) \leqslant p(x)$ for all $x \in N$. On this set, we introduce the partial order $(h, L) \geqslant (g, N)$ whenever $L \supset N$ and $h|_N = g$; note that this relation is indeed a partial order.

It is easy to verify that if $\{(g_\alpha, N_\alpha)\}$ is a chain, then the function g defined on the linear subspace $N = \bigcup_\alpha N_\alpha$ by $g(x) = g_\alpha(x)$ for $x \in N_\alpha$ is well defined and linear, $g(x) \leqslant p(x)$ for all $x \in N$, and $(g, N) \geqslant (g_\alpha, N_\alpha)$ for each α. By Zorn's Lemma 1.7, there is a maximal extension $\hat{f}$ satisfying $\hat{f} \leqslant p$. By the first part of the argument, $\hat{f}$ must be defined on all of X. ∎

The next result tells us when a sublinear functional is actually linear.

5.54 Theorem *A sublinear function $p\colon X \to \mathbb{R}$ on a vector space is linear if and only if it dominates exactly one linear functional on X.*

Proof: First let $p\colon X \to \mathbb{R}$ be a sublinear functional on a vector space. If p is linear and $f(x) \leqslant p(x)$ for all $x \in X$ and some linear functional $f\colon X \to \mathbb{R}$, then $-f(x) = f(-x) \leqslant p(-x) = -p(x)$, so $p(x) \leqslant f(x)$ for all $x \in X$, that is, $f = p$.

Now assume that p dominates exactly one linear functional on X. Note that p is linear if and only if $p(-x) = -p(x)$ for each $x \in X$. So if we assume by way of contradiction that p is not linear, then there exists some $x_0 \neq 0$ such that $-p(-x_0) < p(x_0)$. Let $M = \{\lambda x_0 : \lambda \in \mathbb{R}\}$, the vector subspace generated by x_0, and define the linear functionals $f, g\colon M \to \mathbb{R}$ by $f(\lambda x_0) = \lambda p(x_0)$ and $g(\lambda x_0) = -\lambda p(-x_0)$. From $f(x_0) = p(x_0)$ and $g(x_0) = -p(-x_0)$, we see that $f \neq g$. Next, notice that $f(z) \leqslant p(z)$ and $g(z) \leqslant p(z)$ for each $z \in M$, that is, p dominates both f and g on the subspace M. Now by the Hahn–Banach Theorem 5.53, the two distinct linear functionals f and g have linear extensions to all of X that are dominated by p, a contradiction. ∎

5.10 Separating hyperplane theorems

There is a geometric interpretation of the Hahn–Banach Theorem that is more useful. Assume that X is a vector space. Taking a page from the statisticians' notational handbook, let $[f = \alpha]$ denote the level set $\{x : f(x) = \alpha\}$, and $[f > \alpha]$ denote $\{x : f(x) > \alpha\}$, etc. A **hyperplane** is a set of the form $[f = \alpha]$, where f is a nonzero linear functional on X and α is a real number. (Note well that it is a crucial part of the definition that f be nonzero.) A hyperplane defines two **strict half spaces**, $[f > \alpha]$ and $[f < \alpha]$, and two **weak half spaces**, $[f \geqslant \alpha]$ and $[f \leqslant \alpha]$. A set in a vector spaces is a **polyhedron** if it is the intersection of finitely many weak half spaces.

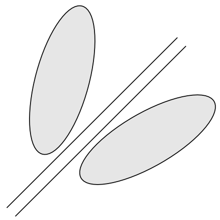


Figure 5.3. Strong separation.

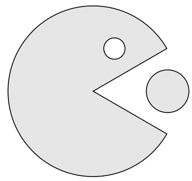


Figure 5.4. These sets cannot be separated by a hyperplane.

The hyperplane $[f = \alpha]$ **separates** two sets A and B if either $A \subset [f \leqslant \alpha]$ and $B \subset [f \geqslant \alpha]$ or if $B \subset [f \leqslant \alpha]$ and $A \subset [f \geqslant \alpha]$. We say that the hyperplane $[f = \alpha]$ **properly separates** A and B if it separates them and $A \cup B$ is not included in H. A hyperplane $[f = \alpha]$ **strictly separates** A and B if it separates them and in addition, $A \subset [f > \alpha]$ and $B \subset [f < \alpha]$ or vice-versa. We say that $[f = \alpha]$ **strongly separates** A and B if there is some $\varepsilon > 0$ with $A \subset [f \leqslant \alpha]$ and $B \subset [f \geqslant \alpha + \varepsilon]$ or vice-versa. We may also say that the linear functional f itself separates the sets when some hyperplane $[f = \alpha]$ separates them, etc. (Note that this terminology is inconsistent with the terminology of Chapter 2 regarding separation by continuous functions. Nevertheless, it should not lead to any confusion.)

It is obvious—but we shall spell it out anyhow, because it is such a useful trick—that if $[f = \alpha]$ separates two sets, then so does $[-f = -\alpha]$, but the sets are in the opposite half spaces. This means we can take our choice of putting A in $[f \geqslant \alpha]$ or in $[f \leqslant \alpha]$.

5.55 Lemma *A hyperplane $H = [f = \alpha]$ in a topological vector space is either closed or dense, but not both; it is closed if and only if f is continuous, and dense if and only if f is discontinuous.*

Proof: If e satisfies $f(e) = \alpha$ and $H_0 = [f = 0]$, then $H = e + H_0$. This shows that we can assume that $\alpha = 0$. If f is continuous, then clearly H_0 is closed. Also, if H_0 is dense, then f cannot be continuous (otherwise f is the zero functional).

Now assume that H_0 is closed and let $x_\lambda \to 0$. Also, fix some u with $f(u) = 1$. If $f(x_\lambda) \not\to 0$, then (by passing to a subnet if necessary) we can assume that $|f(x_\lambda)| \geqslant \varepsilon$ for each λ and some $\varepsilon > 0$. Put $y_\lambda = u - \frac{f(u)}{f(x_\lambda)} x_\lambda$ and note that $y_\lambda \in H_0$ for each λ and $y_\lambda \to u$. So $u \in H_0$, which is impossible. Thus $f(x_\lambda) \to 0$, so f is continuous.

Next, suppose that f is discontinuous. Then there exist a net $\{x_\lambda\}$ and some $\varepsilon > 0$ satisfying $x_\lambda \to 0$ and $|f(x_\lambda)| \geqslant \varepsilon$ for each λ. If x is arbitrary, then put $z_\lambda = x - \frac{f(x)}{f(x_\lambda)} x_\lambda \in H_0$ and note that $z_\lambda \to x$. So H_0 (and hence H) is dense, and the proof is finished. ∎

Ordinary separation is a weak notion because it does not rule out that both sets might actually lie in the hyperplane. The following example illustrates some of the possibilities.

5.56 Example (Kinds of separation) Consider the plane $\mathbb{R}^2$ and set $f(x,y) = y$. Put $A_1 = \{(x,y) : y > 0 \text{ or } (y = 0 \text{ and } x > 0)\}$ and $B_1 = -A_1$. Also define $A_2 = \{(x,y) : x > 0 \text{ and } y \geqslant \frac{1}{x}\}$ and $B_2 = \{(x,y) : x > 0 \text{ and } y \leqslant -\frac{1}{x}\}$. Then the hyperplane $[f = 0]$ separates A_1 and B_1 and strictly separates A_2 and B_2. But the sets A_1 and B_1 cannot be strictly separated, while the sets A_2 and B_2 cannot be strongly separated. ∎

The following simple facts are worth pointing out, and we may use these facts without warning.

5.57 Lemma *If a linear functional f separates the sets A and B, then f is bounded above or below on each set. Consequently, if say A is a linear subspace, then f is identically zero on A.*

Likewise, if B is a cone, then f can take on values of only one sign on B and the opposite sign on A.

Proof: Suppose $f(x) \neq 0$ for some x in the subspace A. For any real number λ define $x_\lambda = \frac{\lambda}{f(x)}x$. Then x_λ also belongs to A and $f(x_\lambda) = \lambda$, which contradicts the fact f is bounded on A.

For the case where B is a cone, observe that either $\lambda f(b) = f(\lambda b) \leqslant f(a)$ holds for all $b \in B$, $a \in A$ and $\lambda \geqslant 0$ or $\lambda f(b) \geqslant f(a)$ for all $b \in B$, $a \in A$ and $\lambda \geqslant 0$. This implies either $f(b) \leqslant 0 \leqslant f(a)$ for all $b \in B$ and $a \in A$ or $f(b) \geqslant 0 \geqslant f(a)$ for all $b \in B$ and $a \in A$. ∎

We may say that a linear functional **annihilates** a subspace when it is bounded, and hence zero, on the subspace.

Another cheap trick stems from the following observation. In a vector space, for nonempty sets A and B we have:

$$A \cap B = \varnothing \quad \Longleftrightarrow \quad 0 \notin A - B.$$

We use this fact repeatedly.

The first important separation theorem is a plain vanilla separating hyperplane theorem—it holds in arbitrary linear spaces and requires no topological assumptions. Instead, a purely algebraic property is assumed.

5.58 Definition *A point x in a vector space is an **internal point** of a set B if there is an absorbing set A such that $x + A \subset B$, or equivalently if the set $B - x$ is absorbing.*

In other words, a point x is an internal point of a set B if and only if for each vector u there exists some $\alpha_0 > 0$ depending on u such that $x + \alpha u \in B$ whenever $|\alpha| \leqslant \alpha_0$.

5.59 Example (Internal point vs. interior point) It should be clear that interior points are internal points. We shall show later (see Lemma 5.60) that a vector in a convex subset of a finite dimensional vector space is an internal point if and only if it is an interior point. However, in infinite dimensional topological vector spaces an internal point of a convex set need not be an interior point. For an example, let $X = C[0, 1]$, the vector space of all continuous real-valued functions defined on $[0, 1]$. On X we consider the two norms $\|f\|_\infty = \max_{x\in[0,1]} |f(x)|$ and $\|f\| = \int_0^1 |f(x)|\,dx$, and let τ_∞ and τ be the Hausdorff linear topologies generated by $\|\cdot\|_\infty$ and $\|\cdot\|$, respectively. If $C = \{f \in C[0, 1] : \|f\|_\infty < 1\}$, then C is a convex set and has 0 as a τ_∞-interior point. In particular, 0 is an internal point of C. Now notice that 0 is not a τ-interior point of C. ∎

As mentioned in the preceding example, in finite dimensional vector spaces the internal points of a convex set are precisely the interior points of the set.

5.60 Lemma *Let C be a nonempty convex subset of a finite dimensional vector space X. Then a vector of C is an internal point of C if and only if it is an interior point of C (for the Euclidean topology on X).*

Proof: Let x_0 be an internal point of C. Replacing C by $C - x_0$, we can assume that $x_0 = 0$. It is easy to see that there exists a basis $\{e_1, \dots, e_k\}$ of X such that $\pm e_i \in C$ for all $i = 1, \dots, k$. Now note that the norm $\left\|\sum_{i=1}^k \alpha_i e_i\right\| = \sum_{i=1}^k |\alpha_i|$ must be equivalent to the Euclidean norm; see Theorem 5.21. If $x = \sum_{i=1}^k \alpha_i e_i \in B_1(0)$, then x can be written as a convex combination of the collection of vectors $\{0, \pm e_1, \dots, \pm e_k\}$ of C, so since C is convex we have $x \in C$. Thus $B_1(0) \subset C$ so that 0 is an interior point of C. (For more details see also the proof of Theorem 7.24.) ∎

We are now ready for the fundamental separating hyperplane theorem.

5.61 Basic Separating Hyperplane Theorem *Two nonempty disjoint convex subsets of a vector space can be properly separated by a nonzero linear functional, provided one of them has an internal point.*

Proof: Let A and B be disjoint nonempty convex sets in a vector space X, and suppose A has an internal point. Then the nonempty convex set $A - B$ has an internal point. Let z be an internal point of $A - B$. Clearly, $z \neq 0$ and the set $C = A - B - z$ is nonempty, convex, absorbing, and satisfies $-z \notin C$. (Why?) By part (2) of Lemma 5.50, the gauge p_C of C is a sublinear function.

We claim that $p_C(-z) \geqslant 1$. Indeed, if $p_C(-z) < 1$, then there exist $0 \leqslant \alpha < 1$ and $c \in C$ such that $-z = \alpha c$. Since $0 \in C$, it follows that $-z = \alpha c + (1 - \alpha)0 \in C$, a contradiction. Hence $p_C(-z) \geqslant 1$.

Let $M = \{\alpha(-z) : \alpha \in \mathbb{R}\}$, the one-dimensional subspace generated by $-z$, and define $f\colon M \to \mathbb{R}$ by $f(\alpha(-z)) = \alpha$. Clearly, f is linear and moreover $f \leqslant p_C$ on M, since for each $\alpha \geqslant 0$ we have $p_C(\alpha(-z)) = \alpha p_C(-z) \geqslant \alpha = f(\alpha(-z))$, and $\alpha < 0$ yields $f(\alpha(-z)) < 0 \leqslant p_C(\alpha(-z))$. By the Hahn–Banach Extension Theorem 5.53, f extends to $\hat{f}$ defined on all of X satisfying $\hat{f}(x) \leqslant p_C(x)$ for all $x \in X$. Note that $\hat{f}(z) = -1$, so $\hat{f}$ is nonzero.

To see that $\hat{f}$ separates A and B let $a \in A$ and $b \in B$. Then we have

$$\begin{aligned}\hat{f}(a) &= \hat{f}(a-b-z) + \hat{f}(z) + \hat{f}(b) \leqslant p_C(a-b-z) + \hat{f}(z) + \hat{f}(b)\\ &= p_C(a-b-z) - 1 + \hat{f}(b) \leqslant 1 - 1 + \hat{f}(b) = \hat{f}(b).\end{aligned}$$

This shows that the nonzero linear functional $\hat{f}$ separates the convex sets A and B.

To see that the separation is proper, let $z = a-b$, where $a \in A$ and $b \in B$. Since $\hat{f}(z) = -1$, we have $\hat{f}(a) \neq \hat{f}(b)$, so A and B cannot lie in the same hyperplane. ∎

5.62 Corollary *Let A and B be two nonempty disjoint convex subsets of a vector space X. If there exists a vector subspace Y including A and B such that either A or B has an internal point in Y, then A and B can be properly separated by a nonzero linear functional on X.*

Proof: By Theorem 5.61 there is a nonzero linear functional f on Y that properly separates A and B. Now note that any linear extension of f to X is a nonzero linear functional on X that properly separates A and B. ∎

5.11 Separation by continuous functionals

Theorem 5.61 makes no mentions of any topology. In this section we impose topological hypotheses and draw topological conclusions. The next lemma gives a topological condition that guarantees the existence of internal points, which is a prerequisite for applying the Basic Separating Hyperplane Theorem 5.61. It is a consequence of the basic Structure Theorem 5.6 and although we have mentioned it before, we state it again in order to emphasize its importance.

5.63 Lemma *In a topological vector space, every neighborhood of zero is an absorbing set. Consequently, interior points are internal.*

Note that the converse of this is not true. In a topological vector space there can be absorbing sets with empty interior. For example, the unit ball in an infinite dimensional normed space is a very nice convex absorbing set, but it has empty interior in the weak topology, see Corollary 6.27.

The next lemma gives a handy criterion for continuity of a linear functional on a topological vector space. It generalizes the result for Banach spaces that linear functionals are bounded if and only if they are continuous.

5.64 Lemma *If a linear functional on a tvs is bounded either above or below on a neighborhood of zero, then it is continuous.*

Proof: If f is linear, then both f and $-f$ are convex, so the conclusion follows from Theorem 5.43. Or more directly, if $f \leqslant M$ on a symmetric neighborhood V of zero, then $x - y \in \frac{\varepsilon}{M}V$ implies $|f(x) - f(y)| = |f(x - y)| \leqslant \frac{\varepsilon}{M}M = \varepsilon$. ∎

The proof of the next result is left as an exercise.

5.65 Lemma *A nonzero continuous linear functional on a topological vector space properly separates two nonempty sets if and only if it properly separates their closures.*

Some more separation properties of linear functionals are contained in the next lemma.

5.66 Lemma *If A is a nonempty subset of a tvs X and a nonzero linear functional f on X satisfies $f(x) \geqslant \alpha$ for all $x \in A$, then $f(x) > \alpha$ for all $x \in A^\circ$ (and so if $A^\circ \neq \varnothing$, then f is continuous).*

In particular, in a tvs, if a nonzero linear functional separates two nonempty sets, one of which has an interior point, then it is continuous and properly separates the two sets.

Proof: Assume that $x_0 + V \subset A$, where V is a circled neighborhood of zero. If $f(x_0) = \alpha$, then for each $v \in V$ we have $\alpha \pm f(v) = f(x_0 \pm v) \geqslant \alpha$. Consequently, $\pm f(v) \geqslant 0$ or $f(v) = 0$ for all $v \in V$. Since V is absorbing, the latter yields $f(y) = 0$ for all $y \in X$, that is, $f = 0$, which is impossible. Hence $f(x) > \alpha$ holds for all $x \in A^\circ$. Now from $f(v) \geqslant \alpha - f(x_0)$ for all $v \in V$, it follows from Lemma 5.64 that f is continuous.

For the last part, let A and B be two nonempty subsets of a tvs X with $A^\circ \neq \varnothing$ and assume that there exist a linear functional f on X and some $\alpha \in \mathbb{R}$ satisfying $f(a) \geqslant \alpha \geqslant f(b)$ for all $a \in A$ and all $b \in B$. By the first part, f is continuous and $f(a) > \alpha$ for all $a \in A^\circ$. The latter shows that f properly separates A and B (so f also property separates $\overline{A}$ and $\overline{B}$). ∎

We now come to a basic topological separating hyperplane theorem.

5.67 Interior Separating Hyperplane Theorem *In any tvs, if the interiors of a convex set A is nonempty and is disjoint from another nonempty convex set B, then $\overline{A}$ and $\overline{B}$ can be properly separated by a nonzero continuous linear functional.*

Moreover, the pairs of convex sets $(A, \overline{B})$, $(\overline{A}, B)$, and (A, B) likewise can be properly separated by the same nonzero continuous linear functional.

Proof: Assume that A and B are two nonempty convex subsets of a tvs X such that $A^\circ \neq \varnothing$ and $A^\circ \cap B = \varnothing$. By Lemma 5.28 we know that $\overline{A^\circ} = \overline{A}$. Now, according to Theorem 5.61, there exists a nonzero linear functional f on X that properly separates A° and B. But then (by Lemma 5.66) f is continuous and properly separates $\overline{A^\circ} = \overline{A}$ and $\overline{B}$. ∎

5.68 Corollary *In any tvs, if the interior of two convex sets are nonempty and disjoint, then their closures (and so the convex sets themselves) can be properly separated by a nonzero continuous linear functional.*

The hypothesis that one of the sets must have a nonempty interior cannot be dispensed with. The following example, due to J. W. Tukey [332], presents two disjoint nonempty closed convex subsets of a Hilbert space that cannot be separated by a continuous linear functional.

5.69 Example (Inseparable disjoint closed convex sets) In ℓ_2, the Hilbert space of all square summable sequences, let

$$A = \{x=(x_1, x_2, \ldots) \in \ell_2 : x_1 \geqslant n|x_n - n^{-\frac{2}{3}}| \text{ for } n = 2, 3, \ldots\}.$$

The sequence v with $v_n = n^{-\frac{2}{3}}$ lies in ℓ_2 and belongs to A, so A is nonempty. Clearly A is convex. It is also easy to see that A is norm closed. Let

$$B = \{x = (x_1, 0, 0, \ldots) \in \ell_2 : x_1 \in \mathbb{R}\}.$$

The set B is clearly nonempty, convex, and norm closed. Indeed, it is a straight line, a one-dimensional subspace.

Observe that A and B are disjoint. To see this note that if x belongs to B, then $n|x_n - n^{-\frac{2}{3}}| = n^{\frac{1}{3}} \xrightarrow[n]{} \infty$, so x cannot lie in A.

We now claim that A and B cannot be separated by any nonzero continuous linear functional on ℓ_2. In fact, we prove the stronger result that $A - B$ is dense in ℓ_2. To see this, fix any $z = (z_1, z_2, \ldots)$ in ℓ_2 and let $\varepsilon > 0$. Choose k so that $\sum_{n=k+1}^{\infty} n^{-\frac{4}{3}} < \varepsilon^2/4$ and $\sum_{n=k+1}^{\infty} z_n^2 < \varepsilon^2/4$.

Now consider the vector $x = (x_1, x_2, \ldots) \in A$ defined by

$$x_n = \begin{cases} \max\limits_{1 \leqslant i \leqslant k} i|z_i - i^{-\frac{2}{3}}| & \text{if } n = 1, \\ z_n & \text{if } 2 \leqslant n \leqslant k, \\ n^{-\frac{2}{3}} & \text{if } n > k. \end{cases}$$

Let $y = (x_1 - z_1, 0, 0, \ldots) \in B$ and note that the vector $x - y \in A - B$ satisfies

$$\|z - (x - y)\| = \Big[\sum_{n=k+1}^{\infty} (z_n - n^{-\frac{2}{3}})^2\Big]^{\frac{1}{2}} \leqslant \Big[\sum_{n=k+1}^{\infty} z_n^2\Big]^{\frac{1}{2}} + \Big[\sum_{n=k+1}^{\infty} n^{-\frac{4}{3}}\Big]^{\frac{1}{2}} < \varepsilon.$$

That is, $A - B$ is dense, so A cannot be separated from B by a continuous linear functional. (Why?) ∎

As an application of the Interior Separating Hyperplane Theorem 5.67, we shall present a useful result on concave functions due to K. Fan, I. Glicksberg, and A. J. Hoffman [120]. It takes the form of an **alternative**, that is, an assertion that exactly one of two mutually incompatible statements is true. We shall see more alternatives in the sequel.

5.70 Theorem (The Concave Alternative) *Let $f_1, \dots, f_m \colon C \to \mathbb{R}$ be concave functions defined on a nonempty convex subset of some vector space. Then exactly one of the following two alternatives is true.*

1. *There exists some $x \in C$ such that $f_i(x) > 0$ for each $i = 1, \dots, m$.*
2. *There exist nonnegative scalars $\lambda_1, \dots, \lambda_m$, not all zero, such that*

$$\sum_{i=1}^{m} \lambda_i f_i(x) \leqslant 0$$

for each $x \in C$.

Proof: It is easy to see that both statements cannot be true. Now consider the subset of $\mathbb{R}^m$:

$$A = \{y \in \mathbb{R}^m : \exists\, x \in C \text{ such that } y_i \leqslant f_i(x) \text{ for each } i\}.$$

Clearly A is nonempty. To see that A is convex, let $y, z \in A$, and pick $x_1, x_2 \in C$ satisfying $y_i \leqslant f_i(x_1)$ and $z_i \leqslant f_i(x_2)$ for each i. Now if $0 \leqslant \alpha \leqslant 1$, then the concavity of the functions f_i implies

$$\alpha y_i + (1-\alpha) z_i \leqslant \alpha f_i(x_1) + (1-\alpha) f_i(x_2) \leqslant f_i\big(\alpha x_1 + (1-\alpha) x_2\big)$$

for each i. Since $\alpha x_1 + (1-\alpha)x_2 \in C$, the inequalities show that $\alpha y + (1-\alpha)z \in A$. That is, A is a convex subset of $\mathbb{R}^m$. Now notice that if (1) is not true, then the convex set A is disjoint from the interior of the convex set $\mathbb{R}^m_+$. So, according to Theorem 5.67 there exists a nonzero vector $\lambda = (\lambda_1, \dots, \lambda_m)$ such that

$$\lambda \cdot y = \sum_{i=1}^{m} \lambda_i y_i \geqslant \sum_{i=1}^{m} \lambda_i f_i(x)$$

for all $y \in \mathbb{R}^m_+$ and all $x \in C$. Clearly, $\sum_{i=1}^{m} \lambda_i f_i(x) \leqslant 0$ for all $x \in C$ and $\lambda \cdot y \geqslant 0$ for all $y \in \mathbb{R}^m_+$. The latter yields $\lambda_i \geqslant 0$ for each i and the proof is complete. ∎

5.12 Locally convex spaces and seminorms

To obtain a separating hyperplane theorem with a stronger conclusion than proper separation, we need stronger hypotheses. One such hypothesis is that the linear space be a locally convex space.

5.71 Definition *Recall that a topological vector space is **locally convex**, or is a **locally convex space**, if every neighborhood of zero includes a convex neighborhood of zero.*[7]

*A **Fréchet space** is a completely metrizable locally convex space.*

Since in a topological vector space the closure of a convex set is convex, the Structure Theorem 5.6 implies that in a locally convex space the closed convex circled neighborhoods of zero form a neighborhood base at zero. Next notice that the convex hull of a circled set is also circled. From this and the fact that the interior of a convex (resp. circled) neighborhood of zero is a convex (resp. circled) neighborhood of zero, it follows that in a locally convex space the collection of all open convex circled neighborhoods of zero is also a neighborhood base at zero.

In other words, we have the following result.

5.72 Lemma *In a locally convex space:*

1. *The collection of all the closed, convex and circled neighborhoods of zero is a neighborhood base at zero.*
2. *The collection of all open, convex and circled neighborhoods of zero is a neighborhood base at zero.*

It turns out that the locally convex topologies are precisely the topologies derived from families of seminorms. Let X be a vector space. For a seminorm $p\colon X \to \mathbb{R}$ and $\varepsilon > 0$, let us write

$$V_p(\varepsilon) = \{x \in X : p(x) \leqslant \varepsilon\},$$

the closed ε-ball of p centered at zero. Now let $\{p_i\}_{i\in I}$ be a family of seminorms on X. Then the collection $\mathcal{B}$ of all sets of the form

$$V_{p_1}(\varepsilon) \cap \cdots \cap V_{p_n}(\varepsilon), \quad \varepsilon > 0,$$

is a filter base of convex sets that satisfies conditions (1), (2), and (3) of the Structure Theorem 5.6. Consequently, $\mathcal{B}$ induces a unique locally convex topology on X having $\mathcal{B}$ as a neighborhood base at zero. This topology is called the locally convex topology **generated by the family of seminorms** $\{p_i\}_{i\in I}$. A family $\mathcal{F}$ of seminorms is **saturated** if $p, q \in \mathcal{F}$ implies $p \vee q \in \mathcal{F}$. If a family of seminorms is saturated, then it follows from Lemmas 5.50 and 5.49 (7) that a neighborhood base at zero is given by the collection of all $V_p(\varepsilon)$, no intersections required.

In the converse direction, let τ be a locally convex topology on a vector space X, and let $\mathcal{B}$ denote the neighborhood base at zero consisting of all circled convex closed neighborhoods of zero. Then, for each $V \in \mathcal{B}$ the gauge p_V is a seminorm on X. An easy argument shows that the family of seminorms $\{p_V\}_{V\in\mathcal{B}}$ is a saturated family generating τ. Thus, we have the following important characterization of locally convex topologies.

[7] Many authors define a locally convex space to be Hausdorff as well.

5.73 Theorem (Seminorms and local convexity) *A linear topology on a vector space is locally convex if and only if it is generated by a family of seminorms.*

In particular, a locally convex topology is generated by the family of gauges of the convex circled closed neighborhoods of zero.

Here is a simple example of a locally convex space.

5.74 Lemma *For any nonempty set X, the product topology on $\mathbb{R}^X$ is a complete locally convex Hausdorff topology.*

Proof: Note that the product topology is generated by the family of seminorms $\{p_x\}_{x\in X}$, where $p_x(f) = |f(x)|$. ∎

If X is countable, then $\mathbb{R}^X$ is a completely metrizable locally convex space, that is, $\mathbb{R}^X$ is a Fréchet space. The metrizable locally convex spaces are characterized by the following result whose proof follows from Theorem 5.10 and 5.73.

5.75 Lemma *A Hausdorff locally convex space (X, τ) is metrizable if and only if τ is generated by a sequence $\{q_n\}$ of seminorms—in which case the topology τ is generated by the translation invariant metric d given by*

$$d(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot \frac{q_n(x-y)}{1 + q_n(x-y)}.$$

Recall that a subset A of a topological vector space (X, τ) is (**topologically**) **bounded**, or more specifically **τ-bounded**, if for each neighborhood V of zero there exists some $\lambda > 0$ such that $A \subset \lambda V$.

The proof of the following simple lemma is left as an exercise.

5.76 Lemma *If a family of seminorms $\{p_i\}_{i\in I}$ on a vector space X generates the locally convex topology τ, then:*

1. *τ is Hausdorff if and only if $p_i(x) = 0$ for all $i \in I$ implies $x = 0$.*
2. *A net $\{x_\alpha\}$ satisfies $x_\alpha \xrightarrow{\tau} x$ if and only if $p_i(x_\alpha - x) \to 0$ for each i.*
3. *A subset A of X is τ-bounded if and only if $p_i(A)$ is a bounded subset of real numbers for each i.*

A locally convex space is **normable** if its topology is generated by a single norm.

5.77 Theorem (Normability) *A locally convex Hausdorff space is normable if and only if it has a bounded neighborhood of zero.*

Proof: If V is a convex, circled, closed, and bounded neighborhood of zero, then note that p_V is a norm that generates the topology. ∎

Here is a familiar example of a completely metrizable locally convex space that is not normable.

5.78 Example **($\mathbb{R}^{\mathbb{N}}$ is not normable)** According to Lemma 5.74 the product topology τ on $\mathbb{R}^{\mathbb{N}}$ is a Hausdorff locally convex topology that is generated by the countable collection $\{p_1, p_2, \ldots\}$ of seminorms, where $p_n(x) = |x_n|$ for each $x = (x_1, x_2, \ldots) \in \mathbb{R}^{\mathbb{N}}$. But then, by Lemma 5.75, the topology τ is also completely metrizable—and, indeed, is generated by the complete translation invariant metric $d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \frac{|x_n - y_n|}{1+|x_n - y_n|}$. In other words, it is a Fréchet space. However, the product topology τ is not normable: Let

$$V = \{x = (x_1, x_2, \ldots) \in \mathbb{R}^{\mathbb{N}} : \; |x_{n_i}| < \varepsilon \text{ for all } i = 1, \ldots, k\}$$

be a basic τ-neighborhood of zero and choose n such that $n \neq n_i$ for all $i = 1, \ldots, k$. Then it is easy to see that $\sup p_n(V) = \infty$. This shows that no τ-neighborhood of zero can be τ-bounded and therefore, by Theorem 5.77, τ is not normable. ∎

Not every tvs is locally convex. Theorems 13.31 and 13.43 show some of the surprises lurking in infinite dimensional spaces. Sometimes, zero is the only continuous linear functional!

5.13 Separation in locally convex spaces

In locally convex spaces, we have the following strong separating hyperplane theorem. (For a sharper version of this result holding for Banach spaces see Corollary 7.47.)

5.79 Strong Separating Hyperplane Theorem *For disjoint nonempty convex subsets of a (not necessarily Hausdorff) locally convex space, if one is compact and the other closed, then there is a nonzero continuous linear functional strongly separating them.*

Proof: Let A and B satisfy the hypotheses. By Lemma 5.3, $A - B$ is a nonempty closed convex set, and it does not contain zero. Thus its complement is an open neighborhood of zero, and since the space is locally convex, there is a circled convex open neighborhood V of zero disjoint from $A - B$. Since V is open, the Interior Separating Hyperplane Theorem 5.67 guarantees that there is a nonzero continuous linear functional f separating V and $A - B$. That is, $f(v) \leqslant f(a) - f(b)$ for all $v \in V$, $a \in A$, and $b \in B$. Since f is nonzero and V is absorbing, f cannot vanish on V. Therefore there exists some $v_0 \in V$ with $f(v_0) > 0$. Now if $\varepsilon = f(v_0)$ and $\alpha = \sup_{b \in B} f(b)$, then note that $f(a) \geqslant \alpha + \varepsilon > \alpha \geqslant f(b)$ for all a in A and b in B. That is, f strongly separates A and B. ∎

We state some easy consequences.

5.80 Corollary (Separating points from closed convex sets) *In a locally convex space, if K is a nonempty closed convex set and $z \notin K$, then there exists a nonzero continuous linear functional strongly separating K and z.*

5.81 Corollary (Non-dense vector subspaces) *A vector subspace of a locally convex space fails to be dense if and only if there exists a nonzero continuous linear functional that vanishes on it.*

5.82 Corollary (The dual separates points) *The topological dual of a locally convex space separates points if and only if the topology is Hausdorff.*

Proof: Let (X, τ) be a locally convex space. If the topological dual X' separates points and $x \neq y$ pick some $f \in X'$ satisfying $f(x) < f(y)$ and note that if $f(x) < c < f(y)$, then the open half spaces $[f < c]$ and $[f > c]$ are disjoint open neighborhoods of x and y.

Conversely, if τ is a Hausdorff topology, then singletons are closed and compact, so the separation of points follows immediately from Corollary 5.80. ∎

This last result stands in marked contrast to Theorem 13.31, where it is shown that zero is the only continuous linear functional on $L_p(\mu)$ for $0 < p < 1$. Of course, these spaces are not locally convex.

Closed convex sets can be characterized in terms of closed half spaces. Consequently they are determined by the dual space. (For a sharper version of the second part of the next theorem that is valid for Banach spaces see Corollary 7.48.)

5.83 Corollary (Closed convex sets) *In a locally convex space, if a convex set is not dense, then its closure is the intersection of all (topologically) closed half spaces that include it.*

In particular, in a locally convex space X, every proper closed convex subset of X is the intersection of all closed half spaces that include it.

Proof: Let A be a non-dense convex subset of a locally convex space. Recall that a closed half space is a set of the form $[f \leqslant \alpha] = \{x : f(x) \leqslant \alpha\}$, where f is a nonzero continuous linear functional. If $a \notin \overline{A}$, then according to Corollary 5.80 there exist a nonzero continuous linear functional g and some scalar α satisfying $\overline{A} \subset [g \leqslant \alpha]$ and $g(a) > \alpha$. This implies that $\overline{A}$ is the intersection of all closed half spaces including A. ∎

Note that if a convex set is dense in the space X, then its closure, X, is not included in any half space, so we cannot omit the qualification "not dense" in the theorem above. The last corollary takes the form of an alternative.

5.84 Corollary (The Convex Cone Alternative) *If C is a convex cone in a locally convex space (X, τ), then for each $x \in X$ one of the following two mutually exclusive alternatives holds.*

1. *The point x belongs to the τ-closure of C, that is, $x \in \overline{C}$.*

2. *There exists a τ-continuous linear functional f on X satisfying*

$$f(x) > 0 \qquad \text{and} \qquad f(c) \leqslant 0 \text{ for all } c \in C.$$

Proof: It is easy to check that statements (1) and (2) are mutually exclusive. Assume $x \notin \overline{C}$. Then, by the Strong Separating Hyperplane Theorem 5.79, there exist a nonzero τ-continuous linear functional f on X and some constant α satisfying $f(x) > \alpha$ and $f(c) \leqslant \alpha$ for all $c \in \overline{C}$. Since C is a cone, it follows that $\alpha \geqslant 0$ and $f(c) \leqslant 0$ for all $c \in \overline{C}$. Consequently, $f(x) > 0$ and $f(c) \leqslant 0$ for all $c \in C$. In other words, we have shown that if $x \notin \overline{C}$, then (2) is true and the proof is finished. ∎

A special case of this result is known as Farkas' Lemma. It and its relatives are instrumental to the study of linear programming and decision theory.

5.85 Corollary (Farkas' Lemma [121]) *If A is a real $m \times n$ matrix and b is a vector in $\mathbb{R}^m$, then one of the following mutually exclusive alternatives holds.*

1. *There exists a vector $\lambda \in \mathbb{R}^n_+$ such that $b = A\lambda$.*

2. *There exists a nonzero vector $a \in \mathbb{R}^m$ satisfying*

$$a \cdot b > 0 \qquad \text{and} \qquad A^t a \leqslant 0.$$

Here, as usual, λ is an n-dimensional column vector, and A^t denotes the $n \times m$ transpose matrix of A.

Proof: By Corollary 5.25 the convex cone C in $\mathbb{R}^m$ generated by the n columns of A is closed. Statement (1) is equivalent to $b \in C$. Corollary 5.84 says that either (1) holds or else there is a linear functional (represented by a nonzero vector a) such that $a \cdot b > 0$ and $a \cdot c \leqslant 0$ for all $c \in C$. But $a \cdot c \leqslant 0$ for all $c \in C$ if and only if $A^t a \leqslant 0$. But this is just (2). ∎

Recall that a seminorm p on a vector space X dominates a linear functional f if $f(x) \leqslant p(x)$ for each $x \in X$. This is equivalent to $|f(x)| \leqslant p(x)$ for each $x \in X$.

5.86 Lemma (Continuous linear functionals) *A linear functional on a tvs is continuous if and only if it is dominated by a continuous seminorm.*

Proof: Let (X, τ) be a tvs and let f be a linear functional on X. If $|f(x)| \leqslant p(x)$ for all $x \in X$ and some τ-continuous seminorm p, then it easily follows that $\lim_{x \to 0} f(x) = 0$, which shows that f is τ-continuous.

For the converse, simply note that if f is a τ-continuous linear functional, then $x \mapsto |f(x)|$ is a τ-continuous seminorm dominating f. ∎

5.87 Theorem (Dual of a subspace) *If (X, τ) is a locally convex space and Y is a vector subspace of X, then every τ-continuous linear functional on Y (endowed with the relative topology) extends to a (not necessarily unique) τ-continuous linear functional on X.*

In particular, the continuous linear functionals on Y are precisely the restrictions to Y of the continuous linear functionals on X.

Proof: Let $f : Y \to \mathbb{R}$ be a continuous linear functional. Pick some convex and circled τ-neighborhood V of zero satisfying $|f(y)| \leqslant 1$ for each y in $V \cap Y$. From part (3) of Lemma 5.50 we see that p_V is a continuous seminorm and it is easy to check that $\big|f(y)\big| \leqslant p_V(y)$ for all $y \in Y$. By the Hahn–Banach Theorem 5.53 there exists an extension $\hat{f}$ of f to all of X satisfying $|\hat{f}(x)| \leqslant p_V(x)$ for all $x \in X$. By Lemma 5.86, $\hat{f}$ is τ-continuous, and we are done. ∎

As an application of the preceding result, we shall show that every finite dimensional vector subspace of a locally convex Hausdorff space is complemented.

5.88 Definition *A vector space X is the **direct sum** of two subspaces Y and Z, written $X = Y \oplus Z$, if every $x \in X$ has a* unique *decomposition of the form $x = y + z$, where $y \in Y$ and $z \in Z$.*

*A closed vector subspace Y of a topological vector space X is **complemented** in X if there exists another closed vector subspace Z such that $X = Y \oplus Z$.*

5.89 Theorem *In a locally convex Hausdorff space every finite dimensional vector subspace is complemented.*

Proof: Let (X, τ) be a locally convex Hausdorff space and let Y be a finite dimensional vector subspace of X. Pick a basis $\{y_1, \dots, y_k\}$ for Y and consider the linear functionals $f_i : Y \to \mathbb{R}$ $(i = 1, \dots, k)$ defined by $f_i(\sum_{j=1}^k \lambda_j y_j) = \lambda_i$.

Clearly, each $f_i : (Y, \tau) \to \mathbb{R}$ is continuous. By Theorem 5.87, each f_i has a τ-continuous extension to all of X, which we again denote f_i. Now consider the continuous projection $P : X \to X$ defined by

$$P(x) = \sum_{i=1}^{k} f_i(x) y_i.$$

That is, P projects x onto the space spanned by $\{y_1, \dots, y_k\}$. Now define the closed vector subspace $Z = \{x - P(x) : x \in X\}$ of X, and note that Z satisfies $Y \oplus Z = X$. ∎