Proof. Let K be a barrel in X. Then, since K is absorbing, $X = \cup_{n=1}^{\infty} nK$. Therefore, by the Baire category theorem, some nK must have nonempty interior. But then $K = n^{-1}(nK)$ has nonempty interior. $\square$

Since any locally convex space has a neighborhood base at 0 of barrels, spaces with every barrel a neighborhood of 0 have lots of open sets, and so their topology is especially strong. We will make this precise in Section 8.5; see Theorem 8.5.24.

Definition. A *Montel space* is a barrelled space with the property that any closed bounded set is compact.

By Theorem 8.3.10, a Banach space with the norm topology is Montel if and only if it is finite-dimensional. The dual X^* of a Banach space has every closed bounded set compact in the $\sigma(X^*, X)$-topology, but it is not barrelled in this topology, and so not Montel. $\mathcal{S}(\mathbb{R}^\nu)$, $\mathcal{D}(\mathbb{R}^\nu)$, and $\mathcal{H}(\Omega)$ are all Montel spaces. For the first two, use equicontinuity ideas, and for the last, the Vitali convergence theorem.

8.4 Separation Theorems

What makes locally convex spaces special is not only that they have lots of continuous linear functionals and so closed hyperplanes, but enough to slip between disjoint convex sets. The hard work has already been done in Theorem 8.1.38 (the Hahn-Banach theorem). By a real linear functional, we mean a linear functional if the space is over $\mathbb{R}$ and a real-valued function linear on X as a real vector space if the space is over $\mathbb{C}$.

Remark. If ℓ is a real linear functional on a complex vector space, $L(x) = \ell(x) - i\ell(ix)$ is complex linear functional since $L(ix) = \ell(ix) - i\ell(-x) = i[\ell(x) - i\ell(ix)]$. Conversely, if L is complex linear, $\ell(x) = \operatorname{Re} L(x)$ is real linear. So we could talk about complex linear functionals and place Re[ ] in front of the functional to handle the complex case.

Definition. Two sets A and B in X, a topological vector space, are said to be *separated* if and only if there exists a nonzero continuous real linear functional ℓ on X and $\alpha \in \mathbb{R}$ so that

$$A \subset \{x \mid \ell(x) \le \alpha\} \qquad B \subset \{x \mid \ell(x) \ge \alpha\} \tag{8.4.1}$$

If the inequalities in (8.4.1) can be taken strict (i.e., = dropped), we say that A and B are strictly separated.

Theorem 8.4.1. *Let A and B be disjoint convex subsets of a locally convex space X with A open. Then A and B can be separated.*

Proof. $A-B=\cup_{x\in B}(A-x)$ is open. Pick $-x_0\in A-B$ and let $C=x_0+A-B$ so $0\in C$, C is open, convex $x_0\notin C$ (since $x_0\in C\Rightarrow 0\in A-B\Rightarrow A\cap B\neq\emptyset$). Let ρ_C be the gauge of C. Then, since $x_0\notin C$, $\rho_C(x_0)\geq 1$. Let $W=\{\lambda x_0\mid\lambda\in\mathbb{R}\}$ and $\ell: W\to\mathbb{R}$ by $\ell(\lambda x_0)=\lambda\rho_C(x_0)$. Then $\lambda>0$ implies that

$$\ell(\lambda x_0)=\lambda\rho_C(x_0)=\rho_C(\lambda x_0)$$

while $\lambda<0$ implies

$$\ell(\lambda x_0)=\lambda\rho_C(x_0)<0\leq\rho_C(\lambda x_0)$$

so (8.1.78) holds. Thus, by the Hahn-Banach theorem, there is $L: X\to\mathbb{R}$ so $L(x_0)=\rho_C(x_0)\geq 1$ and $L(x)\leq\rho_C(x)$ so for $x\in C$, $L(x)\leq 1$.

By the remark after the proof of Proposition 8.3.13, ρ_C is a continuous convex function. Thus $\tilde{\rho}_C(x)=\max(\rho_C(x),\rho_C(-x))$ is also continuous. Since $|L(x)|=\max(L(x),L(-x))$ obeys

$$|L(x)|\leq\tilde{\rho}_C(x)$$

we see that L is a continuous linear functional.

It follows if $a\in A$, $b\in B$, then

$$L(x_0+a-b)\leq L(x_0)$$

or

$$L(a)\leq L(b)$$

Since this holds for all pairs,

$$\sup_{a\in A}L(a)\equiv\alpha\leq\inf_{b\in B}L(b)$$

and L separates A and B. □

Lemma 8.4.2. *Let A be an open convex set and $L: X\to\mathbb{R}$ be a nonzero linear function. Then $L[A]$ is open.*

Proof. Pick x_0 with $L(x_0)\neq 0$. For any $y\in A$, $\{t\mid y+tx_0\in A\}$ is an open interval about 0, so $L[A]$ contains an open interval about $L(y)$. □

Theorem 8.4.3. *Let A and B be disjoint open convex subsets of a locally convex space X. Then A and B can be strictly separated.*

Proof. By Theorem 8.4.1, there is a nonzero linear functional ℓ and $\alpha \in \mathbb{R}$ so $\ell[A] \subset (-\infty, \alpha]$ and $\ell[B] \subset [\alpha, \infty)$. Since A and B are open, $\ell[A] \subset (-\infty, \alpha)$ and $\ell[B] \subset (\alpha, \infty)$ by the lemma. $\square$

Lemma 8.4.4. *Let A and B be disjoint, closed convex sets with B compact. Then there exist disjoint, open convex sets U and V with $A \subset U$ and $B \subset V$.*

Proof. Let $C = A - B$. If $x_\alpha = a_\alpha - b_\alpha$ is a net in C so $x_\alpha \to x$, then by passing to a subnet, we can suppose $b_\alpha \to b$ in B since B is compact. Thus $a_\alpha = x_\alpha + b_\alpha \to x + b = a \in A$ since A is closed. Thus $x = a - b \in C$, that is, C is closed. Since $0 \notin C$, we can find W open, balanced, and convex so $0 \in W$ and $W \cap C = \emptyset$. Let $U = A + \frac{1}{2}W$ and $V = B + \frac{1}{2}W$. Then $U \cap V$ is empty and U, V are open and convex. $\square$

Theorem 8.4.5. *Let A and B be disjoint, closed convex subsets of a locally convex space X with B compact. Then A and B can be strictly separated.*

Proof. This follows immediately from Theorem 8.4.3 and Lemma 8.4.4. $\square$

Remark. In $\mathbb{R}^2$, let $A = \{(x,y) \mid x \leq 0\}$ and $B = \{(x,y) \mid x \geq y^{-1}\}$ (see Figure TK). Then A and B are disjoint, closed convex sets. They cannot be strictly separated. This shows it is essential that B be compact. x-ref?

Corollary 8.4.6. *Let X be a locally convex vector space and $x, y \in X$ with $x \neq y$. Then there exists $\ell \in X$ with $\ell(x) \neq \ell(y)$.*

Proof. Take $A = \{x\}$ and $B = \{y\}$. $\square$

Corollary 8.4.7. *Let X be a locally convex vector space. Let $W \subset X$ be a closed subspace and $x \notin W$. Then there exists $\ell \in X^*$ so $\ell \restriction W = 0$ and $\ell(x) \neq 0$.*

Proof. Let $A = W$ and $B = \{x\}$. Let ℓ separate A and B. Since $\ell[W]$ is a subspace of $\mathbb{R}$ and it is semibounded, it must be 0, that is, $\ell[W] = \{0\}$. $\ell(x)$ is then nonzero. $\square$

Here is the section's final application of the separation theorem.

Definition. A *closed half-space* is a set of the form

$$\{x \mid \ell(x) \geq \alpha\} \tag{8.4.2}$$

for some continuous, nonzero linear functions and some $\alpha \in \mathbb{R}$.

Remark. One might also want to take sets of the form $\{x \mid \ell(x) \leq \alpha\}$, but since we can take $\ell \to -\ell$ and $\alpha \to -\alpha$, that is unnecessary!

Theorem 8.4.8. *A set A is a closed convex set if and only if it is an intersection of closed half-spaces. If $0 \in A$ and A is a closed convex set, it is the intersection of a family of half-spaces of the form* (8.4.2) *with $\alpha = -1$.*

Proof. An arbitrary intersection of closed sets is closed and an arbitrary intersection of convex sets is convex. Therefore, any intersection of closed half-spaces is a closed convex set.

Conversely, if A is a closed convex set and $x \notin A$, by Theorem 8.4.5, we can find ℓ_x and α_x so $\ell_x(x) < \alpha_x$ and

$$A \subset \{y \mid \ell_x(y) > \alpha_x\} \tag{8.4.3}$$

We claim

$$A \equiv \bigcap_{x \notin A} \{y \mid \ell_x(y) \geq \alpha_x\} \tag{8.4.4}$$

By (8.4.3), $A \subset \cap_{x \notin A}\{y \mid \ell_x(y) > \alpha_x\} \subset \cap_{x \notin A}\{y \mid \ell_x(y) \geq \alpha_x\}$ and by construction, for any $x \notin A$, $x \notin \{y \mid \ell_x(y) \geq \alpha_x\}$ so $x \notin \cap_{x \in A}\{y \mid \ell_x(y) \geq \alpha_x\}$.

Finally, if $0 \in A$, the α_x's above are negative. Replace ℓ_x by $|\alpha_x|^{-1}\ell_x = \tilde{\ell}_x$. We have

$$A = \bigcap_{x \in A} \{y \mid \tilde{\ell}_x(y) \geq -1\} \tag{8.4.5}$$

□

Corollary 8.4.9. *Let X be a locally convex space. Let $A \subset X$ be a closed convex set. Then A is closed in the $\sigma(X, X^*)$-topology. If it is compact in the original topology, it is also compact in the $\sigma(X, X^*)$-topology.*

Proof. Each closed half-space is weakly closed and so A is weakly closed by Theorem 8.4.8.

If $\mathcal{T}$ is the topology on X, the identity map $x : X_{\mathcal{T}} \to X_\sigma$ is continuous and a bijection on A. Therefore, $i[A]$ is σ-compact. □

8.8 Extreme Points and the Krein-Milman Theorem

The next four sections will focus on an important geometric aspect of compact sets, namely, the role of extreme points where:

Definition. An *extreme point* of a convex set, A, is a point $x \in A$, with the property that if $x = \theta y + (1-\theta)z$ with $y, z \in A$ and $\theta \in [0,1]$, then $y = x$ and/or $z = x$. $\mathcal{E}(A)$ will denote the set of extreme points of A.

In other words, an extreme point is a point that is not an interior point of any line segment lying entirely in A.

Example 8.8.1. The ν-simplex, Δ_ν is given by (8.5.3) as the convex hull in $\mathbb{R}^{\nu+1}$ of $\{\delta_1, \dots, \delta_{\nu+1}\}$, the coordinate vectors. It is easy to see its extreme points are precisely the $\nu+1$ points $\{\delta_j\}_{j=1}^{\nu+1}$. The hypercube $C_0 = \{x \in \mathbb{R}^\nu \mid |x_i| \le 1\}$ has the 2^ν points $(\pm1, \pm1, \dots, \pm1)$ as extreme points. The ball $B^\nu = \{x \notin \mathbb{R}^\nu \mid |x| \le 1\}$ has the entire sphere as extreme points, showing $\mathcal{E}(A)$ can be infinite.

An interesting example (see Figure TK) is the set $A \subset \mathbb{R}^3$, which is the x-ref?
convex hull of

$$A = \mathrm{ch}(\{(x,y,0) \mid x^2+y^2=1\} \cup \{(1,0,\pm1)\}) \tag{8.8.1}$$

Its extreme points are

$$\mathcal{E}(A) = \{(x,y,0) \mid x^2+y^2=1,\ x \neq 1\} \cup \{(1,0,\pm1)\}$$

$(1,0,0) = \frac{1}{2}(1,0,1) + \frac{1}{2}(1,0,-1)$ is not an extreme point. This example shows that even in the finite-dimensional case, the extreme points may not be closed. In the infinite-dimensional case, we will even see that the set of extreme points can be dense! □

If a point, x, in A is not extreme, it is an interior point of some segment

$$[y,z] = \{\theta y + (1-\theta)z \mid 0 \le \theta \le 1\} \tag{8.8.2}$$

with $y \neq z$. If y or z is not an extreme point, we can write them as extreme points and continue. (If A is compact and in $\mathbb{R}^\nu$, and if one extends the line segment to be maximal, one can prove this process will stop in finitely many steps. Indeed, that in essence is the method of proof we will use in

Theorem 8.8.11). If one thinks about writing y, z as convex combinations, one "expects" that any point in A is a convex linear combination of extreme points of A — and we will prove this precisely when A is compact and finite-dimensional. Indeed, if $A \subset \mathbb{R}^\nu$, we will prove that at most $\nu + 1$ extreme points are needed. This fails in infinite dimension, but we will find a replacement, the Krein-Milman theorem, which says that any point is a limit of convex combinations of extreme points. These are the two main results of this section.

Extreme points are a special case of a more general notion:

Definition. A *face* of a convex set is a nonempty subset, F, of A with the property that if $x, y \in A$, $\theta \in (0,1)$, and $\theta x + (1-\theta)y \in F$, then $x, y \in A$. A face, F, that is strictly smaller than A is called a *proper face*.

Thus a face is a subset so that any line segment $[xz] \subset A$, with interior points in F must lie in F. Extreme points are precisely one-point faces of A. (*Note*: See the remark before Proposition 8.8.6 for a later restriction of this definition.)

Example 8.8.2 (Example 8.8.1 continued). Δ_ν has lots of faces; explicitly, it has $2^{\nu+1} - 2$ faces, namely, $\nu + 1$ extreme points, $\binom{\nu+1}{2}$ facial lines, …, $\binom{\nu+1}{\nu}$ faces of dimension $(\nu - 1)$. The hypercube C_ν has $3^\nu - 1$ faces, namely, 2^ν extreme points, $\nu 2^{\nu-1}$ facial lines, $\binom{\nu}{2}2^{\nu-2}$ facial planes, …, $\binom{\nu}{\nu-1}2$ $\nu - 1$-dimensional faces. The only faces on the ball are its extreme points. The faces of the set A of (8.8.1) are its extreme points, the line $\{(1,0,y) \mid |y| \le 1\}$, and the lines $\{\theta(x_0, y_0, 0) + (1-\theta)(1,0,1)\}$ and $\{\theta(x_0, y_0, 0) + (1-\theta)(1,0,-1)\}$ where x_0, y_0 are fixed with $x_0^2 + y_0^2 = 1$ and $x_0 \neq 1$. $\square$

A canonical way proper faces are constructed is via linear functionals.

Theorem 8.8.3. *Let A be a convex subset of a real vector space. Let $\ell : A \to \mathbb{R}$ be a linear functional with*

(i)

$$\sup_{x \in A} \ell(x) = \alpha < \infty \tag{8.8.3}$$

(ii) *$\ell \restriction A$ is not constant.*

Then

$$\{y \mid \ell(y) = \alpha\} = F \tag{8.8.4}$$

if nonempty, is a proper face of A.

Remark. If A is compact and ℓ is continuous, of course, F is nonempty.

Proof. Since ℓ is linear, F is convex. Moreover, if $y, z \in A$ and $\theta \in (0,1)$ and $\theta y + (1-\theta)z \in F$, then $\theta\ell(y) + (1-\theta)\ell(z) = \alpha$ and $\ell(y) \leq \alpha$, $\ell(z) \leq \alpha$ implies $\ell(y) = \ell(z) = \alpha$, that is, $y, z \in F$. By (ii), F is a proper subset of A. $\square$

The hyperplane $\{y \mid \ell(y) = \alpha\}$ with α given by (8.8.3) is called a *tangent hyperplane* or *support hyperplane*. The set (8.8.4) is called an *exposed set*. If F is a single point, we call the point an *exposed point*.

Example 8.8.4. We have just seen that every exposed set is a face so, in particular, every exposed point is an extreme point. I'll bet if you think through a few simple examples like a disk or triangle in the plane or a convex polyhedron in $\mathbb{R}^3$, you'll conjecture the converse is true. But it is not! Here is a counterexample in $\mathbb{R}^2$ (see Figure TK): x-ref?

$$A = \{(x,y) \mid -1 \leq x \leq 1,\ -2 \leq y \leq 0\} \cup \{(x,y) \mid x^2 + y^2 \leq 1\}$$

The boundary of A above $y = -2$ is a C^1 curve, so there is a unique supporting hyperplane through each such boundary point. The supporting hyperplane through the extreme point $(1,0)$ is $x = 1$ so $(1,0)$ is not an exposed point, but it is an extreme point. $\square$

Proposition 8.8.5. *Any proper face F of A lies in the topological boundary of A. Conversely, if $A \subset X$, a locally convex space (and, in particular, in $\mathbb{R}^\nu$), and A^{int} is nonempty, any point $x \in A \cap \partial A$ lies in a proper face.*

Proof. Let $x \in F$ and pick $y \in A \backslash F$. The set of $\theta \in \mathbb{R}$ so $z(\theta) \equiv \theta x + (1-\theta)y \in A$ includes $[0,1]$, but it cannot include any $\theta > 1$ for if it did, $\theta = 1$ (i.e., x) would be an interior point of a line in A with at least one end point in $A \backslash F$. Thus $x = \lim_{n\downarrow 0} z(1+n^{-1})$ is a limit point of points not in A, that is, $x \in \bar{A} \cap \overline{X \backslash A} = \partial A$.

For the converse, let $x \in A \cap \partial A$ and let $B = A^{\text{int}}$. Since B is open, Theorem 8.4.1 implies there exists a continuous $L \neq 0$ with $\alpha = \sup_{y \in B} L(y) \leq L(x)$. Since $x \in A$, $L(x) = \alpha$. Since B is open, $L[B]$ is an open set (Lemma 8.4.2), so the supporting hyperplane $H = \{y \mid L(y) = \alpha\}$ is disjoint from B and so $H \cap A$ is a proper face. $\square$

To have lots of extreme points, we will need lots of boundary points, so it is natural to restrict ourselves to closed convex sets. The convex set $\mathbb{R}^\nu_+ = \{x \in \mathbb{R} \mid x_i \geq 0 \text{ all } i\}$ has a single extreme point, so we will also restrict to bounded sets. Indeed, except for some examples, we will restrict ourselves to compact convex sets in the infinite-dimensional case. Convex cones are interesting but can normally be treated as suspensions of compact convex sets; see the discussion in the Notes. So we will suppose A is a compact convex subset of a locally convex space. As noted in Corollary 8.4.9, A is weakly compact, so we will suppose henceforth that we are dealing with the weak topology.

x-ref?

Remark. Henceforth, we will also restrict the term "face" to indicate a *closed* set.

Proposition 8.8.6. *Let $F \subset A$ with A a compact convex set and F a face of A. Let $B \subset F$. Then B is a face of F if and only if it is a face of A. In particular, $x \in F$ is in $\mathcal{E}(F)$ if and only if it is also in $\mathcal{E}(A)$, that is,*

$$\mathcal{E}(F) = F \cap \mathcal{E}(A)$$

Proof. If B is a face of A, $x \in B$, and x is an interior point of $[y, z] \subset F$, it is an interior point of $[y, z] \subset A$, so $y, z \in B$ and thus B is a face of F.

Conversely, if B is a face of F, $x \in B$, and $x \in [y, z] \subset A$, since $x \in F$, the fact that F is a face implies $y, z \in F$ so $[y, z] \subset F$. Thus since B is a face of F, $y, z \in B$ and so B is a face of A. □

We turn next to a detailed study of the finite-dimensional case. We begin with some notions that involve finite dimension but which are useful in the infinite-dimensional case also. Since we will be discussing affine subspaces, affice spaces, affine independence, etc., we will temporarily use vector spaces, vector spaces, etc. to denote the usual notions in a vector space.

Let X be a vector space. An *affine subspace* is a set of the form $a + W$ where $a \in X$ and W is a vector subspace. The *affine span* of a subset $A \subset X$ is the smallest affine subspace containing A. If $A = \{e_1, \dots, e_n\}$, then its affine span is just

$$S(e_1, \dots, e_n) = \left\{ \theta_1 e_1 + \cdots + \theta_n e_n \;\middle|\; \theta \in \mathbb{R}^\nu,\ \sum_{i=1}^n \theta_i = 1 \right\} \tag{8.8.5}$$

as is easy to see since $\sum_{i=1}^n \theta_i = 1$ implies that

$$\theta_1 e_1 + \cdots + \theta_n e_n = e_1 + \sum_{j=2}^n \theta_j (e_j - e_1) \tag{8.8.6}$$

so the right-hand side of (8.8.5) is e_1 plus the vector span of $\{e_j - e_1\}_{j=2}^n$. The convex hull of $\{e_1, \dots, e_n\}$ is, of course,

$$\mathrm{ch}(e_1, \dots, e_n) = \left\{ \theta_1 e_1 + \cdots + \theta_n e_n \mid \theta \in \mathbb{R}^\nu,\ \sum_{i=1}^n \theta_i = 1,\ \theta_i \geq 0 \right\} \tag{8.8.7}$$

We call $\{e_1, \dots, e_n\}$ *affinely independent* if and only if $\sum_{i=1}^n \theta_i e_i = 0$ and $\sum_{i=1}^n \theta_i = 0$ implies $\theta \equiv 0$. By (8.8.6) this is true if and only if $\{e_j - e_1\}_{j=2}^n$ are vector independent.

Proposition 8.8.7. $\mathrm{ch}(e_1, \dots, e_n)$ *always has a nonempty interior as a subset of* $S(e_1, \dots, e_n)$.

Proof. By successively throwing out dependent vectors from $P = \{e_j - e_1\}_{j=2}^n$, find a maximal independent subset of P. By relabelling, suppose it is $P' \equiv \{e_j - e_1\}_{j=1}^k$ so $\{e_1, \dots, e_k\}$ are affinely independent, and each $e_\ell - e_1$ with $\ell > k$ is a linear combination of P'. Then $S(e_1, \dots, e_n) = S(e_1, \dots e_k)$.

Since $\mathrm{ch}(e_1, \dots, e_k) \subset \mathrm{ch}(e_1, \dots, e_n)$, it suffices to prove the result when $e_1, \dots, e_n$ are affinely independent. In that case, $\varphi : \Delta_{n-1} \to \mathrm{ch}(e_1, \dots, e_k)$ is a bijection and continuous, so a homeomorphism. Since Δ_{n-1} has a nonempty interior $(\{\theta_1, \dots, \theta_n) \mid \sum_{i=1}^n \theta_i = 1,\ 0 < \theta_i\})$, so does $\mathrm{ch}(e_1, \dots, e_k)$. ☐

Remark. The θ's are called *barycentric coordinates* for $S(e_1, \dots, e_\ell)$ and $\mathrm{ch}(e_1, \dots, e_\ell)$.

Theorem 8.8.8. *Let* $A \subset \mathbb{R}^\nu$ *be a convex set. Then there is a unique affine subspace* W *of* $\mathbb{R}^\nu$ *so that* $A \subset W$*, and as a subset of* W*,* A *has a nonempty interior.*

Proof. Pick $e_1 \in A$ and consider $B = A - e_1 \ni 0$. Let W be the subspace generated by B, that is, let $f_1, \dots, f_{\ell-1}$ be a maximal linear independent subset of B, and let X be the vector span of $\{f_j\}_{j=1}^{\ell-1}$. Let $e_j = f_{j-1} + e_1$ for $j = 2, \dots, \ell$ so $e_1 + X \equiv W$ is the affine span of $\{e_j\}_{j=1}^\ell$. By construction $B \subset X$ so $A \subset W = S(e_1, \dots, e_\ell)$. By Proposition 8.8.7, $\mathrm{ch}(e_1, \dots, e_\ell) \subset A$ is open in S, so A has no nonempty interior as a subset of S.

W is unique because any affine subspace containing A must contain $e_1, \dots, e_\ell$ and so $S(e_1, \dots, e_\ell)$. If its dimension were larger than W, W would have empty interior in it and so would A. Thus the condition that A have nonempty interior uniquely determines W. ☐

Definition. The *dimension* of a convex set $A \subset \mathbb{R}^\nu$ is the dimension of the unique affine subspace given by Theorem 8.8.8. The interior of A as a subset of W is written A^{iint} and called the *intrinsic interior* of A. $\partial^i A$, the *intrinsic boundary* of $A = \bar{A}\backslash A^{\text{iint}}$.

Proposition 8.8.9. *Let A be a compact convex subset of $\mathbb{R}^\nu$. Then*

(i) *$\partial^i A$ is the union of the proper faces of A.*

(ii) *If $x \in \partial^i A$ and y is any point in A^{iint}, $\{\theta \mid (1-\theta)x + \theta y \in A\} = [0, \alpha]$ for some $\alpha > 1$.*

(iii) *If $x \in A^{\text{iint}}$ and $y \in A$, $\{\theta \mid (1-\theta)x + \theta y \in A\} \cap (-\infty, 0) \neq \emptyset$.*

Remark. This gives us an intrinsic definition of A^{iint}. $x \in A^{\text{iint}}$ if and only if for any $y \in A$, the line $[y, x]$ continued past x lies in A for at least a while. Similarly, $\partial^i A$ is determined by the condition that any line that intersects A in more than one point enters and leaves A at points in $\partial^i A$ and any $x \in \partial^i A$ lies on such a line as an extreme point.

Proof. (i) This follows from Proposition 8.8.5 if we view A as a subset of W.

(ii) We know $x \in \partial^i A$ lies in some face F. Since A^{iint}, viewed as a subset of W, is disjoint from the boundary, $y \notin F$. As in the proof of Proposition 8.8.5, $\{\theta \mid (1-\theta)x + \theta y \in A\} \cap (-\infty, 0) = \emptyset$. Since this set is connected and compact and contains $[0, 1]$, it must be the requisite form. That $\alpha > 1$ and $\alpha \neq 1$ follows from (iii).

(iii) $[x, y]$ lies in A, so in W, so since A^{iint} is open in W, $\{\theta \mid (1-\theta)x+\theta y \in A^{\text{iint}}\}$ is open. Since it contains 0, it must contain an interval $(-\varepsilon, \varepsilon)$ about 0. ☐

Proposition 8.8.10. *Let $A \subset \mathbb{R}^\nu$ be a compact convex set. Let $\ell = \dim(A)$ and let F be a proper face of F. Then $\dim(F) < \ell$.*

Proof. Let $A \subset W$ where W is the unique ℓ-dimensional space containing A. If $\dim(F) = \ell$, then W must also be the unique ℓ-dimensional space containing F, and so F has not empty interior. But as a set in W, $F \subset \partial A$, which contradicts $F^{\text{int}} \neq \emptyset$. Thus $\dim(F) < \ell$. ☐

We are now ready for the main finite-dimensional result:

Theorem 8.8.11 (Minkowski-Carathéodory Theorem). *Let A be a compact convex subset of $\mathbb{R}^\nu$ of dimension n. Then any point in A is a convex combination of at most $n + 1$ extreme points. In fact, for any x, one* *can fix* $e_0 \in \mathcal{E}(A)$ *and find* $e_1, \dots, e_n \in \mathcal{E}(A)$ *so* x *is a convex combination of* $\{e_j\}_{j=0}^n$ *If* $x \in A^{\text{iint}}$, *then* $x = \sum_{j=0}^n \theta_j e_j$ *with* $\theta_0 > 0$. *In particular,*

$$A = \text{ch}(\mathcal{E}(A)) \tag{8.8.8}$$

Remarks. 1. It pays to think of the square in $\mathbb{R}^2$ which has four extreme points, but where any point is in the convex hull of three points (indeed, for most interior points in exactly two ways).

2. The example of the n simplex Δ_n shows that for general A's, one cannot do better than $n+1$ points. Of course, for some sets, one can do better. No matter what value of ν, the ball B^ν has the property that any point is a convex combination of at most two extreme points.

Proof. We use induction on n. $n = 0$, that is, single-point sets, is trivial. Suppose we have the result for all sets, B, with $\dim(B) \le n-1$. Let A have dimension n and $x \in A$ and $e_0 \in \mathcal{E}(A)$. Thake the line segment $[e_0, x]$ and extend it $\{\theta \mid (1-\theta)e_0 + \theta x \in A\} = [0, \alpha]$ for some α by Proposition 8.8.9. Let $y = (1-\alpha)e_0 + \alpha x$. Since $\alpha \ge 1$,

$$x = \theta_0 e_0 + (1-\theta_0)y \tag{8.8.9}$$

where $\theta_0 = 1 - \alpha^{-1} \ge 0$.

By construction, $y \in \partial^i A$ and so, by Proposition 8.8.9, $y \in F$, some proper face of A. By Proposition 8.8.10, $\dim(F) \le n-1$, so by the induction hypothesis, $y = \sum_{j=1}^n \varphi_j e_j$ where $\varphi_j \ge 0$, $\sum_{j=1}^n \varphi_1 = 1$, and $\{e_1, \dots, e_n\} \subset \mathcal{E}(F)$. By Proposition 8.8.6, $\mathcal{E}(F) \subset \mathcal{E}(A)$. Thus

$$x = \sum_{j=0}^n \theta_j e_j$$

where $\theta_j = (1-\theta_0)\varphi_j$ for $j = 1, \dots, n$.

If $\theta_0 = 0$, by (8.8.9), $x = y$ and $x \in \partial^i A$. Thus if $x \in A^{\text{iint}}$, $\theta_0 \ne 0$. □

We will have more to say about extreme points of finite-dimensional convex sets in Section 8.15 when we discuss a particular convex set, the set of all doubly stochastic matrices. In particular, we will show that a compact, convex set, K, in $\mathbb{R}^\nu$ has finitely many extreme points if and only if it is a finite intersection of closed half-spaces (Corollary 8.15.3).

In the infinite-dimensional case, it is not clear that $\mathcal{E}(A)$ is nonempty — we will go through the main construction in two phases. We will first show that $\mathcal{E}(A) \neq \emptyset$ for A a compact convex subset of a locally convex space and then, fairly easily, we will be able to show that

$$A = \operatorname{cch}(\mathcal{E}(A))$$

which is the Krein-Milman theorem. The following illustrates that the infinite-dimensional case is subtle.

Example 8.8.12. Let A be the closed unit ball in $L^1(0,1)$. Let $f \in A$ with $f \neq 0$. Then $H_f(s) = \int_0^s |f(t)|\,dt$ is a continuous function with $H_f(0) = 0$ and $H_f(1) = \alpha \leq 1$. Thus there exists s_0 with $H_f(s_0) = \alpha/2$. Let

$$g = 2f\chi_{(0,s_0)}$$
$$h = 2f\chi_{(s_0,1)}$$

Then $\|g\|_1 = \|h\|_1 = \|f\|_1 = \alpha \leq 1$ and $f = \frac{1}{2}h + \frac{1}{2}g$. Since $h \neq g$, f is not an extreme point. Clearly, $0 = \frac{1}{2}(f - f)$ is not extreme either. Thus A has no extreme points!

We will show below that any compact convex subset, A, of a locally convex space has $\mathcal{E}(A) \neq \emptyset$. This means that the unit ball in $L^1(0,1)$ cannot be compact in any topology making it into a locally convex space. In particular, because of the Bourbaki-Alaoglu theorem, $L^1(0,1)$ cannot be the dual of any Banach space. This is subtle because $\ell^1(\mathbb{Z})$ is a dual (of $c_0(\mathbb{Z})$, the bounded sequences vanishing at infinity). Of course, the unit ball in $\ell^1(\mathbb{Z})$ has lots of extreme points in each $\pm\delta_n$. ☐

Proposition 8.8.13. *Let A be a compact convex subset of a locally convex space, X. Then $\mathcal{E}(A) \neq \emptyset$.*

Proof. Extreme points are one-point faces. We will find them as minimal faces. So let $\mathcal{F}$ be the family of proper faces of A with $F_1 > F_2$ if $F_1 \subset F_2$. This is a partially ordered set and it has the chain property, that is, if $\{F_\alpha\}_{\alpha \in I}$ is linearly ordered, then it has an "upper" bound ("upper" here means small since a "larger than" means contained in), namely, $\cap_{\alpha \in I} F_\alpha$. This is closed, a face (by a simple argument), and nonempty because of the intersection property for compact sets (see TK). x-ref?

Thus, by Zorn's lemma, there exist minimal faces. Suppose F is such a minimal face and F has at least two distinct points x and y. By Corollary 8.4.6, there is a linear functional on X and so on F with $\ell(x) \neq \ell(y)$. Since F is compact,

$$\tilde{F} = \left\{ z \in F \mid \ell(z) = \sup_{w \in F} \ell(w) \right\}$$

is nonempty. It is a face of F and so, by Proposition 8.8.6, $\tilde{F}$ is a face of A. Since $\ell(x) \neq \ell(y)$, it cannot be that both x and y lie in $\tilde{F}$, so $\tilde{F} \subsetneqq F$, violating minimality. It follows that F has a single point and that point must be an extreme point. $\square$

Remark. In $L^1(0,1)$, $F_\alpha = \{f \in L^1 \mid \|\|f\|_1\| = 1$ and $f(x) = 0$ on $(0,\alpha)$, $f \geq 0\}$ is a face and it is linearly ordered (since $\alpha > \beta \Rightarrow F_\alpha \subset F_\beta$), but $\cap_\alpha F_\alpha$ is empty. This proves the lack of compactness directly.

Theorem 8.8.14 (The Krein-Milman Theorem). *Let A be a compact convex subset of a locally convex vector space, X. Then*

$$A = \mathrm{cch}(\mathcal{E}(A)) \tag{8.8.10}$$

Proof. Since $\mathcal{E}(A) \subset A$ and A is closed and convex, $B \equiv \mathrm{cch}(\mathcal{E}(A)) \subset A$. Suppose $B \neq A$ so there exists $x_0 \in A \backslash B$. Since B is closed and convex, by Theorem 8.4.5, there exists $\ell \in X^*$ so

$$\ell(x_0) > \sup_{y \in B} \ell(y) \tag{8.8.11}$$

Let $F = \{x \in A \mid \ell(x) = \sup_{z \in A} \ell(z)\}$. Then F is nonempty since A is compact, a face, and by (8.8.11),

$$F \cap B = \emptyset \tag{8.8.12}$$

By Proposition 8.8.13, F has an extreme point, y_0, and then, by Proposition 8.8.6, $y_0 \in \mathcal{E}(A)$. Thus, $y_0 \in B$, contradicting (8.8.12). $\square$

Remark. In the next section (see Theorem 8.9.4), we will prove a sort of converse of this theorem.

Example 8.8.15. Let $X = C_{\mathbb{R}}([0,1])$ and let A be the unit ball in $\|\cdot\|_\infty$. If $|f(x)| < 1$ for some x_0 in $[0,1]$, then by continuity for some ε, $|f(y)| < 1$ for $|y - x_0| < \varepsilon$ and we can find $g \neq 0$ supported in $(x_0 - \varepsilon, x_0 + \varepsilon)$, so both $f + g$ and $f - g$ lie in A. Since $f = \frac{1}{2}(f+g) + \frac{1}{2}(f-g)$, f is not an extreme point. Thus, extreme points have $|f(x)| = 1$. By continuity and reality, A has precisely two extreme points $f \equiv \pm 1$. $\mathrm{cch}(\mathcal{E}(A))$ is the constant functions in A so $A \neq \mathrm{cch}(\mathcal{E}(A))$. Thus, $C_{\mathbb{R}}([0,1])$ is not a dual space. $\square$

Example 8.8.16. This is an important example. Let X be a compact Hausdorff space and let $A = \mathcal{M}_{+1}(X)$ be the set of probability measures on X. The extreme points of A are precisely the single-point pure points, δ_x, since if $C \subset X$ has $0 < \mu(C) < 1$ and

$$\begin{aligned} \mu_C(B) &= \mu(C)^{-1}\mu(B \cap C) \\ \mu_{X\backslash C} &= \mu(X\backslash C)^{-1}\mu(B\backslash C) \end{aligned}$$

then with $\theta = \mu(C)$, $\mu = \theta\mu_C + (1-\theta)\mu_{X\backslash C}$ so μ is not an extreme point.

Suppose μ has the property that $\mu(A)$ is 0 or 1 for each $A \subset X$. If $x \neq y$ are both in $\operatorname{supp}(\mu)$, we can find disjoint open sets B, C with $x \in B$ and $y \in C$. By the $0, 1$ law, either $\mu(B) = 0$ or $\mu(C) = 0$ or both. But that would mean x and y cannot both be in $\operatorname{supp}(\mu)$. Thus $\operatorname{supp}(\mu)$ is a single point and $\mu = \delta_x$ for some x, that is, the only extreme points are among the $\{\delta_x\}$. But each δ_x is an extreme point since $\delta_x = \frac{1}{2}\mu + \frac{1}{2}\nu$ implies $\operatorname{supp}(\mu) \subset \{x\}$ so $\mu = \delta_x$. Thus $\mathcal{E}(A) = \{\delta_x \mid x \in X\}$.

$\operatorname{ch}(\mathcal{E}(A))$ is the pure point measures. A is compact in the $\sigma(\mathcal{M}(X), C(X))$-topology and so the Krein-Milman theorem says that the pure point measures are weakly dense — something that is easy to prove directly. □

Example 8.8.17. In some ways, this is an extension of the last example. Let X be a compact Hausdorff space and let $T : X \to X$ be a continuous bijection. A probability measure μ on X is called *invariant* if and only if $\mu(T^{-1}[A]) = A$ for all $A \subset X$. This is equivalent to

$$\int f(Tx)\, d\mu(x) = \int f(x)\, d\mu(x) \tag{8.8.13}$$

for all $f \in C(X)$. An invariant measure, μ, is called *ergodic* if and only if $\mu(A \triangle T[A]) = 0$ (i.e., $A = T[A]$ μ a.e.) implies $\mu(A)$ is 0 or 1. Ergodic measures will be the theme of Chapter 21. x-ref?

Let T^* map $\mathcal{M}_{+,1}(X) \to \mathcal{M}_{+,1}(X)$ by

$$\int f(x)\, d(T^*\mu)(x) = \int f(Tx)\, d\mu(x)$$

Pick any $\mu \in \mathcal{M}_{+,1}(X)$ and let

$$\mu_n = \frac{1}{n}\sum_{j=0}^{n-1}(T^*)^j(\mu)$$

Then for any $f \in C(X)$,

$$|\mu_n(f) - \mu_n(Tf)| = \left|\frac{1}{n}[((T^*)^n\mu)(f) - \mu(f)]\right| \leq \frac{2}{n}\,\|f\|_\infty \tag{8.8.14}$$

Thus, if μ_∞ is any weak-$*$ limit point of μ_n, $\mu_\infty(Tf) = \mu_\infty(f)$ for all f, that is, $T^*\mu_\infty = \mu_\infty$. Since $\mathcal{M}_{+,1}(X)$ is compact in the weak-$*$ topology, we conclude

$$\mathcal{M}^I_{+,1}(T) = \{\mu \in \mathcal{M}_{+,1} \mid T^*\mu = \mu\}$$

is not empty.

We claim $\mu \in \mathcal{M}^I_{+,1}(T)$ is ergodic if and only if $\mu \in \mathcal{E}(\mathcal{M}^I_{+,1}(T))$. Suppose μ is not ergodic. Then there exists an almost invariant set A with $0 < \mu(A) < 1$. μ can be decomposed $\mu = \theta\mu_A + (1-\theta)\mu_{X\setminus A}$ with $\theta = \mu(A)$ and $\mu_C(B) = \mu(C)^{-1}\mu(B \cap C)$.

Conversely, suppose μ is ergodic. Then in $L^2(X, d\mu)$, define $(Uf)(x) = f(Tx)$. Then U is unitary. Since as functions on ∂D

$$\frac{1}{n}\sum_{j=0}^{n-1} e^{in\theta} \to \begin{cases} 1 & \theta = 0 \\ 0 & \theta \in (0, 2\pi) \end{cases}$$

the continuity of the functional calculus (see Theorem 3.3.2) implies

$$\frac{1}{n}\sum_{j=1}^{n-1} U^n f \xrightarrow{L^2} P_{\{1\}} f \tag{8.8.15}$$

where $P_{\{1\}}$ is the projection onto the invariant functions, that is, those g with $Ug = g$. We claim that, since μ is ergodic, any such g is constant. For clearly, $\operatorname{Re} g$ and $\operatorname{Im} g$ obey $Ug = g$ so we can suppose g is real. But then, for all rational (α, β), $\{x \mid \alpha < g(x) < \beta\}$ is almost T-invariant and so it has measure 0 or 1. This implies g is a.e. constant. Since $\langle 1, U^n f\rangle = \langle 1, f\rangle = \mu(f)$, we see the constant must be $\mu(f) = \int f(x)\,d\mu(x)$.

We have thus shown that if μ is ergodic, then

$$\int \left|\frac{1}{n}\sum_{j=0}^{n-1} f(T^{n-1}x) - \mu(f)\right|^2 d\mu(x) = 0 \tag{8.8.16}$$

Suppose now $\mu = \theta\nu + (1-\theta)\eta$ with $0 < \theta < 1$. Since (8.8.16) has a positive integrand, we see that (8.8.16) holds if μ is replaced by ν or η (but $\mu(f)$ is left fixed). Thus

$$\int \frac{1}{n} \sum_{j=0}^{n-1} f(T^{n-1}x)\, d\nu(x) \to \mu(f) \tag{8.8.17}$$

But since ν is invariant, the left side of (8.8.17) is $\nu(f)$ for any n. Thus $\nu(f) = \mu(f)$, and similarly, $\eta(f) = \mu(f)$. It follows that $\nu = \eta = \mu$, that is, μ is an extreme point.

We have therefore shown that ergodic measures are precisely the extreme points of $\mathcal{M}^I_{+,1}(T)$. The Krein-Milman theorem therefore implies the existence of ergodic measures. If $\mathcal{M}^I_{+,1}(T)$ has more than one point, there must be multiple extreme points.

Now suppose that $\{T_\alpha\}_{\alpha\in I}$ is an arbitrary family of commuting maps of X to X. Invariant measures for all the T_α's at once are defined in the obvious way, and μ is called ergodic if $\mu(A\triangle T_\alpha[A]) = 0$ for all α implies $\mu(A)$ is 0 or 1. Since the T's commute, T^*_α maps each $\mathcal{M}^I_{+,1}(T_\beta)$ to itself, and so by repeating the proof that $\mathcal{M}_{+,1}(X)$ has invariant measures, we see $\mathcal{M}^I_{+,1}(T_\beta)$ has a T^*_α-invariant point. By induction, there are invariant measures for any finite set $\{T^*_{\alpha_i}\}_{i=1}^{\ell}$, and then by compactness and the fact that invariant measures are closed, invariant measures for all $\{T_\alpha\}_{\alpha\in I}$. We summarize in the following theorem. This example is discussed further in Example 8.9.7. □

Theorem 8.8.18. *Let X be a compact Hausdorff space and let $\{T_\alpha\}_{\alpha\in I}$ be a family of commuting bijections of X to itself. Then $\mathcal{M}^I_{+,1}(\{T_\alpha\})$, the set of common invariant measures, is nonempty. The ergodic measures are precisly $\mathcal{E}(\mathcal{M}^I_{+,1}(\{T_\alpha\}))$, the extreme points, and are therefore also nonempty.*

As an example, if X is a compact abelian group, and for each $x \in X$, $T_x : X \to X$ by $T_x(y) = xy$, then there is an invariant measure. We have therefore constructed a Haar measure in this case, which is unique (see TK). Similar ideas can be used to construct what are invariant means on noncompact abelian groups. See the Notes.

x-ref?

x-ref?

(8.8.15) provides a useful criterion for ergodicity.

Theorem 8.8.19. *Let μ be an invariant measure for a continuous bijection T on a compact Hausdorff space. For any function $f \in L^2(X, d\mu)$ and $n = 0, 1, \dots,$ define*

$$(Av_n f)(x) = \frac{1}{2n+1} \sum_{j=-n}^{n} f(T^j x) \tag{8.8.18}$$

Then μ is ergodic if and only if

$$\lim_{n\to\infty} \mu(|Av_n f|^2) = |\mu(f)|^2 \tag{8.8.19}$$

For (8.8.19) *to hold, it suffices that it holds for a dense set, S, in $L^2(X, d\mu)$.*

Proof. (8.8.19) is equivalent to weak convergence

$$(Av_n)^*(Av_n) \xrightarrow{w} (1,\,\cdot\,)1$$

the projection onto 1, so since $\|Av_n\| \le 1$, it suffices to prove it for a dense set.

If T is ergodic, then (8.8.15) implies (8.8.19). Conversely, if (8.8.19) holds, A is an invariant set, and χ_A is its characteristic function, then $Av_n(\chi_A) = \chi_A$ so (8.8.19) implies $\mu(A) = \mu(A)^2$, that is, $\mu(A)$ is 0 or 1. Thus μ is ergodic. □

Example 8.8.20. Let $X = \partial D$, the unit circle. Let α be an irrational number and let

$$T(e^{i\theta}) = e^{i(\theta+2\pi\alpha)}$$

Let $d\mu = d\theta/2\pi$ and $f_m = e^{im\theta} \in L^2(\partial D, d\mu)$. Then, for $m \neq 0$,

$$\begin{aligned} Av_n(f_m) &= (2n+1)^{-1}\biggl(\sum_{j=-n}^{j} e^{2\pi i j\alpha m}\biggr) f_m \\ &= (2n+1)^{-1}\,\frac{\sin(2\pi(n+\frac{1}{2})m\alpha)}{\sin(\pi m\alpha)}\, f_m \end{aligned}$$

so $\|Av_n(f_m)\| \to 0$ if $m \neq 0$. Since $\{f_m\}_{m=0,\pm1,\dots}$ are a basis of $L^2(\partial D, d\mu)$, (8.8.19) holds, so μ is ergodic. Notice $A = \{e^{2\pi i\alpha m}\}_{m=-\infty}^{\infty}$ is an invariant set but it has measure 0. It can be shown that μ is the only invariant measure in this case. □

Example 8.8.21. This is really an ad for a later discusion in Section 12.5. x-ref? Given a locally compact group, G, a unitary representation is a continuous map U taking G to the unitary operators on a Hilbert space, $\mathcal{H}$. Given such a representation, one can form the functions $F_{\varphi,U}(g) = \langle\varphi, U(g)\varphi\rangle$ for each $U \in \mathcal{H}$. We will see that as φ runs over all unit vectors and U over all representations, $\{F_{\varphi,U}\}$ forms a compact convex subset in $C(G)$ in the $\|\cdot\|_\infty$-topology. Its extreme points will correspond to what we called irreducible representations, and we will use the Krein-Milman theorem to prove the existence of such representations. □

We end this section with a few results relating extreme points and linear or affine maps between spaces and sets. These will be needed in the next section.

Proposition 8.8.22. *Let X and Y be locally convex spaces and let A, B be compact convex subsets of X and Y, respectively. Let $T : X \to Y$ be a continuous linear map. Then if $T[\mathcal{E}(A)] \subset B$, we have that $T[A] \subset B$.*

Proof. Since T is linear and B is convex, each $T(\sum_{i=1}^n \theta_i x_i)$ with $\sum_{i=1}^n \theta_i = 1$ and $x_i \in \mathcal{E}(A)$ lies in B. Then, since B is closed and T is continuous, the same is true of limits. Since $A = \operatorname{cch}(\mathcal{E}(A))$, we see $T[A] \subset B$. $\square$

Definition. Let X and Y be locally convex spaces and let A, B be convex subsets of X and Y, respectively. A map $T : A \to B$ is called affine if and only if for all $x, y \in A$ and $\theta \in [0, 1]$, $T(\theta x + (1-\theta)y) = \theta T(x) + (1-\theta)T(y)$.

Proposition 8.8.23. *Let A and B be compact convex subsets of locally convex spaces and let $T : A \to B$ be a continuous affine map. Then for any face, F, of B, $G \equiv T^{-1}[F]$, if nonempty, is a face of A.*

Proof. G is closed since F is closed and T is continuous. If $x \in G$, $y, z \in A$, and $x = \theta y + (1-\theta)z$ with $\theta \in (0, 1)$, then $T(x) \in F$, $T(y), T(z) \in B$, and $T(x) = \theta T(y) + (1-\theta)T(z)$. Since F is a face, $T(y), T(z) \in F$, that is, $y, z \in G$. Thus G is a face. $\square$

8.9 The Strong Krein-Milman Theorem

The representation theorem for points in a compact convex set in terms of extreme points is clean in the finite-dimensional case — a point is a convex combination of finitely many extreme points. But in the form we have it so far, the infinite-dimensional case is murky — points are only limits of convex sums of extreme points. An attractive thought is that somehow this limit of sums is just an integral. We will take a first stab at this idea in this section, a stab that is often fine and which we will raise to high art in the next section.

While the main result (Theorem 8.9.2) in this section is somewhat mathematically unsatisfying since it only asserts any point in A, a compact convex subset, is an integral of points in $\overline{\mathcal{E}(A)}$ (and we will show lots of examples where $\overline{\mathcal{E}(A)}$ is all of A!), it is powerful and includes many classical integral