A

Elements of functional analysis

A.1 Separating hyperplane theorem

Let $v \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$ be given and consider the set $H = \{x \in \mathbb{R}^n : \langle v, x\rangle = \gamma\}$. For $x \in H$ we have

$$\langle v, x - (\gamma/\|v\|^2)v\rangle = 0,$$

so $H = v^\perp + (\gamma/\|v\|^2)v$. The complement of H consists of the two sets $\{x : \langle v, x\rangle < \gamma\}$ and $\{x : \langle v, x\rangle > \gamma\}$ on the two "sides" of the hyperplane.

The following theorem says that for two disjoint, convex sets, one compact and one closed, there exists two "parallel" hyperplanes such that the sets lie strictly one different sides of those hyperplanes.

The assumption that one of the sets is compact can not be dropped (see Exercise 1)

Theorem A.1.1 (Separating hyperplane theorem). *Let K and C be disjoint, convex subsets of $\mathbb{R}^n$, K compact and C closed. There exist $v \in \mathbb{R}^n$ and $\gamma_1, \gamma_2 \in \mathbb{R}$ such that*

$$\langle v, x\rangle < \gamma_1 < \gamma_2 < \langle v, y\rangle$$

for all $x \in K$ and $y \in C$.

Proof. Consider the function $f : K \to \mathbb{R}$ defined by $f(x) = \inf\{\|x - y\| : y \in C\}$, i.e. $f(x)$ is the distance of x to C. The function f is continuous (check) and since K is compact, there exists $x_0 \in K$ such that f attains its minimum at x_0. Let $y_n \in C$ be such that $\|x_0 - y_n\| \to f(x_0)$. By the parallelogram law we have

$$\begin{aligned}\left\|\frac{y_n - y_m}{2}\right\|^2 &= \left\|\frac{y_n - x_0}{2} - \frac{y_m - x_0}{2}\right\|^2 \\ &= \tfrac{1}{2}\|y_n - x_0\|^2 + \tfrac{1}{2}\|y_m - x_0\|^2 - \left\|\frac{y_n + y_m}{2} - x_0\right\|^2.\end{aligned}$$

By convexity $(y_n + y_m)/2 \in C$, so that $\|(y_n + y_m)/2 - x_0\| \geq f(x_0)$. Hence, we have

$$\left\|\frac{y_n - y_m}{2}\right\|^2 \leq \tfrac{1}{2}\|y_n - x_0\|^2 + \tfrac{1}{2}\|y_m - x_0\|^2 - f^2(x_0).$$

The right-hand side of this display converges to 0 as $n, m \to \infty$. So the y_n form a Cauchy sequence and hence they converge to some $y_0 \in \mathbb{R}^n$. Since C is closed, $y_0 \in C$. Let $v = y_0 - x_0$. Since K and C are disjoint, $v \neq 0$. It follows that $0 < \|v\|^2 = \langle v, y_0 - x_0\rangle = \langle v, y_0\rangle - \langle v, x_0\rangle$. It remains to show that $\langle v, x\rangle \leq \langle v, x_0\rangle$ and $\langle v, y_0\rangle \leq \langle v, y\rangle$ for all $x \in K$ and $y \in C$.

Take $y \in C$. Since C is convex, the line segment $y_0 + \lambda(y - y_0)$, $\lambda \in [0, 1]$, belongs to C. Since y_0 minimizes the distance to x_0, we have

$$\|y_0 - x_0\| \leq \|y_0 - x_0 + \lambda(y - y_0)\|$$

for every λ. By squaring this we find that

$$0 \leq 2\lambda \langle y_0 - x_0, y - y_0\rangle + \lambda^2\|y - y_0\|^2.$$

Dividing by λ and then letting $\lambda \to 0$ gives $\langle v, y - y_0\rangle \geq 0$, as desired.

A similar argument shows that $\langle v, x\rangle \leq \langle v, x_0\rangle$ for $x \in K$. □

The *polar* C^0 of a set $C \subseteq \mathbb{R}^n$ is defined as

$$C^0 = \{y \in \mathbb{R}^n : \langle x, y\rangle \leq 1 \text{ for all } x \in C\}.$$

Note that in the special case that C is closed under multiplication with positive scalars, we have $C^0 = \{y \in \mathbb{R}^n : \langle x, y\rangle \leq 0 \text{ for all } x \in C\}$ (check). For a given x, the set $C_x^0 = \{z : \langle x, z\rangle \leq 0\}$ is the set of all vectors that lie on the same side of $x^\perp$ as $-x$. The polar is in this case the intersection of all the sets C_x^0 for $x \in C$.

To illustrate the *bipolar theorem* geometrically, consider a V-shaped set: C the union of two rays emanating from the origin. Then one readily sees that the polar of the polar of C precisely equals the convex hull of C. The general result is as follows.

Theorem A.1.2 (Bipolar theorem). *Let $C \subseteq \mathbb{R}^n$ contain 0. Then the bipolar $C^{00} = (C^0)^0$ equals the closed convex hull of C.*

Proof. It is clear that C^{00} is a closed, convex set containing C, so the closed convex hull A of C is a subset of C^{00}. Suppose that the converse inclusion does not hold. Then there exists a point $x_0 \in C^{00}$ that is not in A. By the separating hyperplane theorem there then exists a vector $v \in \mathbb{R}^n$ and $\gamma_1, \gamma_2 \in \mathbb{R}$ such that $\langle x_0, v\rangle > \gamma_1 > \gamma_2 > \langle y, v\rangle$ for all $y \in A$. Since $0 \in C \subseteq A$ we have $\gamma_1 > 0$. Dividing by γ_1 shows there exists a vector $v \in \mathbb{R}^n$ such that $\langle x_0, v\rangle > 1 > \langle y, v\rangle$ for all $y \in A$. The second inequality implies that $v \in C^0$, and then the first one implies that $x_0 \notin C^{00}$, which gives a contradiction. □

A.2 Topological vector spaces

A vector space X is called a *topological vector space* if it is endowed with a topology which is such that every point of X is a closed set and the addition and scalar multiplication operations are continuous.

It is easy to see that translation by a fixed vector and multiplication by a nonzero scalar are homeomorphisms of a topological vector space. This implies in particular that the topology is *translation-invariant*, meaning that a set $E \subseteq X$ is open if and only if each of its translates $x + E$ is open.

Topological vector spaces have nice separation properties. Combined with the fact that points are closed sets, the next theorem implies for instance that they are always Hausdorff.

Theorem A.2.1. *Suppose that K and C are disjoint subsets of a topological vector space X, K compact and C closed. Then there exits a neighborhood V of 0 such that $K + V$ and $C + V$ are disjoint.*

Proof. The continuity of addition implies that for every neighborhood W of 0 there exist neighborhoods V_1 and V_2 of 0 such that $V_1 + V_2 \subseteq W$ (check). Now put $U = V_1 \cap V_2 \cap (-V_1) \cap (-V_2)$. Then U is symmetric (i.e. $U = -U$) and $U + U \subseteq W$. Applying the same procedure to the neighborhood U we see that for every neighborhood W of 0 there exists a symmetric neighborhood U such that $U + U + U \subseteq W$ (etc.).

Pick an $x \in K$. Then $X \backslash C$ is an open neighborhood of x. By translation invariance and the preceding paragraph there exists a symmetric neighborhood V_x of 0 such that $x + V_x + V_x + V_x$ does not intersect C. By the symmetry of V_x this implies that $x + V_x + V_x$ and $C + V_x$ are disjoint (check). Since K is compact, it is covered by finitely many sets $x_1 + V_{x_1}, \ldots, x_n + V_{x_n}$. Put $V = V_{x_1} \cap \cdots \cap V_{x_n}$. Then

$$K + V \subseteq \bigcup (x_i + V_{x_i} + V) \subseteq \bigcup (x_i + V_{x_i} + V_{x_i})$$

and none of the terms in the last union intersects $C + V$. □

The following lemma implies that if V is a neighborhood of 0 in a topological vector space X, then for every $x \in X$ it holds that $x \in rV$ if r is large enough. A set V with this property is called *absorbing.*

Lemma A.2.2. *Suppose V is a neighborhood of 0 in a topological vector space X and r_n is a sequence of positive numbers tending to infinity. Then*

$$\bigcup r_n V = X.$$

Proof. Fix $x \in X$. Then since V is open in X and $\lambda \mapsto \lambda x$ from $\mathbb{R}$ to X is continuous, $\{\lambda : \lambda x \in V\}$ is open in $\mathbb{R}$. The set contains 0, and hence it contains $1/r_n$ for n large enough. This completes the proof. □

For an arbitrary absorbing subset A (for instance a neighborhood of 0) of a topological vector space X we define the *Minkowsky functional* $\mu_A : X \to [0, \infty)$ by

$$\mu_A(x) = \inf\{t > 0 : x/t \in A\}.$$

Note that μ_A is indeed finite-valued, since A is absorbing. The following lemma collects properties that we need later.

Lemma A.2.3. *Let A be a convex, absorbing subset of a topological vector space X and let μ_A be its Minkowsky functional.*

(i) $\mu_A(x + y) \le \mu_A(x) + \mu_A(y)$ *for all* $x, y \in X$.

(ii) $\mu_A(tx) = t\mu_A(x)$ *for all* $x \in X$ *and* $t \ge 0$.

Proof. For $x, y \in X$ and $\varepsilon > 0$, consider $t = \mu_A(x) + \varepsilon$, $s = \mu_A(y) + \varepsilon$. Then by definition of μ_A, $x/t \in A$ and $y/s \in A$. Hence, the convex combination

$$\frac{x + y}{s + t} = \frac{t}{s + t}\frac{x}{t} + \frac{s}{s + t}\frac{y}{s}$$

belongs to A as well. This proves (i). The proof of (ii) is easy. □

For the proof of the following characterization of continuous linear functionals we need the notion of a *balanced neighborhood*. A set $B \subseteq X$ is said to be *balanced* if $\alpha B \subseteq B$ for every scalar $\alpha \in \mathbb{R}$ with $|\alpha| \le 1$.

Lemma A.2.4. *Every neighborhood of* 0 *contains a balanced neighborhood of* 0.

Proof. Let U be a neighborhood of 0. Since scalar multiplication is continuous, there exists a $\delta > 0$ and a neighborhood V of 0 in X such that $\alpha V \subseteq U$ whenever $|\alpha| < \delta$. Then $W = \cup_{|\alpha|<\delta}\, \alpha V$ is a balanced neighborhood of 0. □

A linear map $\Lambda : X \to \mathbb{R}$ is called a *linear functional* on the space X. A linear functional on X is called *bounded* on a subset $A \subseteq X$ if there exists a number $K > 0$ such that $|\Lambda x| \le K$ for all $x \in A$.

Theorem A.2.5. *Let Λ be a nontrivial linear functional on a topological vector space X. Then Λ is continuous if and only if Λ is bounded on a neighborhood of* 0.

Proof. Suppose Λ is continuous. Then the null space $N = \{x \in X : \Lambda x = 0\}$ is closed. Since Λ is nontrivial, there exists $x \in X \backslash N$. By Theorem A.2.1 there exists a balanced neighborhood V of 0 such that $x + V$ and N are disjoint. Then $\Lambda(V)$ is a balanced subset of $\mathbb{R}$. Suppose it is not bounded. Then since it is balanced, it most be all of $\mathbb{R}$. In particular, there then exists a $y \in V$ such that

$\Lambda y = -\Lambda x$. But then $x + y \in N$, a contradiction. Hence, $\Lambda(V)$ is bounded, i.e. Λ is bounded on V.

Conversely, suppose that $|\Lambda x| \leq M$ for all $x \in V$. For $r > 0$, put $W = (r/M)V$. Then for $x \in W$, say $x = (r/M)y$ for $y \in V$, we have $|\Lambda x| = (r/M)|\Lambda y| \leq r$. Hence, Λ is continuous at 0. By translation invariance, it is continuous everywhere. □

A.3 Hahn-Banach theorem

The proof of the following version of the Hahn-Banach theorem relies on the axiom of choice, in the form of the *Hausdorff maximality theorem*:

Every nonempty partially ordered set $\mathcal{P}$ contains a totally ordered subset $\mathcal{Q}$ which is maximal with respect to the property of being totally ordered.

A proof of this fact can for instance be found in Rudin (1987), pp. 395–396.

Theorem A.3.1 (Hahn-Banach theorem). *Suppose X is a (real) vector space and $p : X \to \mathbb{R}$ satisfies $p(x+y) \leq p(x) + p(y)$ and $p(tx) = tp(x)$ for $x, y \in X$ and $t \geq 0$. Then if f is a linear functional on a subspace M of X such that $f(x) \leq p(x)$ for all $x \in M$, f extends to a linear functional Λ on the whole space X such that*

$$-p(-x) \leq \Lambda x \leq p(x)$$

for all $x \in X$.

Proof. Suppose M is a proper subspace of X and pick $x_1 \in X \backslash M$. For $x, y \in M$ we have

$$f(x) + f(y) = f(x+y) \leq p(x+y) \leq p(x - x_1) + p(y + x_1),$$

hence $f(x) - p(x - x_1) \leq p(y + x_1) - f(y)$. So there exists an α such that

$$f(x) - \alpha \leq p(x - x_1), \quad f(y) + \alpha \leq p(y + x_1) \tag{A.1}$$

for all $x, y \in M$. Now let M_1 be the vector space spanned by M and x_1. An element of M_1 is of the form $x + \lambda x_1$ for some $\lambda \in \mathbb{R}$. So we can extend f to M_1 by setting $f_1(x + \lambda x_1) = f(x) + \lambda\alpha$. Then f_1 is a well-defined linear functional on M_1 and the inequalities in (A.1) imply that $f_1(x) \leq p(x)$ for all $x \in M_1$ (check).

Let $\mathcal{C}$ be the collection of pairs (M', f'), where M' is a subspace of X containing M and f' is a linear extension of f to M' such that $f \leq p$ on M'. Put an ordering on $\mathcal{C}$ by saying that $(M', f') \leq (M'', f'')$ if $M' \subseteq M''$ and $f''|_{M'} = f'$. This is a partial ordering and $\mathcal{C}$ is not empty. Hence, by the Hausdorff maximality theorem, we can extract a maximal totally ordered subcollection $\mathcal{C}'$. Let $\tilde{M}$ be the union of all M' for which $(M', f') \in \mathcal{C}'$. Then

$\tilde{M}$ is a subspace of X (check). If $x \in \tilde{M}$ then $x \in M'$ for some M' such that $(M', f') \in \mathcal{C}'$. We then put $\Lambda x = f'(x)$. This defines a linear function Λ on $\tilde{M}$ and we have that $\Lambda \leq p$ on $\tilde{M}$ (check). If $\tilde{M}$ were a proper subspace of X the construction of the preceding paragraph would give us a further extension of Λ, contradicting the maximality of $\mathcal{C}'$. Hence, $\tilde{M} = X$. This completes the proof, upon noting that $\Lambda \leq p$ implies that $-p(-x) \leq -\Lambda(-x) = \Lambda x$ for all $x \in X$. □

Before we use the Hahn-Banach theorem to prove the infinite-dimensional version of the separating hyperplane theorem we introduce some more concepts and notation.

A topological vector space X is called *locally convex* if for every neighborhood V of 0 there exists a convex neighborhood U of 0 such that $U \subseteq V$. The space of continuous linear maps from X to $\mathbb{R}$ is denoted by X^*. It is called the *dual* of X, and is treated in more detail in the next section.

Theorem A.3.2 (Separation theorem). *Let A and B be disjoint, nonempty, convex subsets of a topological vector space X.*

(i) If A is open there exist $\Lambda \in X^*$ and $\gamma \in \mathbb{R}$ such that

$$\Lambda x < \gamma \leq \Lambda y$$

for every $x \in A$ and $y \in B$.

(ii) If X is locally convex, A is compact and B is closed, there exist $\Lambda \in X^*$ and $\gamma_1, \gamma_2 \in \mathbb{R}$ such that

$$\Lambda x < \gamma_1 < \gamma_2 < \Lambda y$$

for every $x \in A$ and $y \in B$.

Proof. (i). Pick $a_0 \in A$ and $b_0 \in B$ and put $x_0 = b_0 - a_0$. Define $C = A - B + x_0$ and note that C is a convex, open neighborhood of 0. Let μ_C be the Minkowsky functional of C.

Let M be the linear subspace generated by x_0 and define a linear functional f on M by putting $f(\lambda x_0) = \lambda$. Since A and B are disjoint, $x_0 \notin C$ so we have $\mu_C(x_0) \geq 1$ and hence, for $\lambda \geq 0$, $f(\lambda x_0) = \lambda \leq \lambda \mu_C(x_0) = \mu_C(\lambda x_0)$. For $\lambda < 0$ we have $f(\lambda x_0) < 0 \leq \mu_C(\lambda x_0)$. By Lemma A.2.3 and the Hahn-Banach theorem, Theorem A.3.1, the functional f extends to a linear functional Λ on X, and the extension satisfies $\Lambda x \leq \mu_C(x)$ for all $x \in X$. In particular $\Lambda \leq 1$ on C, so that $|\Lambda| \leq 1$ on the neighborhood $C \cap -C$ of 0. By Theorem A.2.5 this implies that Λ is continuous, i.e. $\Lambda \in X^*$.

Now for $a \in A$ and $b \in B$ we have that

$$\Lambda a - \Lambda b + 1 = \Lambda(a - b + x_0) \leq \mu_C(a - b + x_0) < 1,$$

since $a - b + x_0 \in C$ and C is open (Exercise 2), so $\Lambda a < \Lambda b$. It follows that $\Lambda(A)$ and $\Lambda(B)$ are disjoint, convex subsets of $\mathbb{R}$, the first one lying on the left of the second one. Since A is open and Λ is nonconstant, $\Lambda(A)$ is open as well

(Exercise 3). Letting γ be the right end point of $\Lambda(A)$ completes the proof of (i).

(ii). By Theorem A.2.1 and the local convexity of X there exists a convex neighborhood V of 0 such that $(A+V)\cap B = \varnothing$. By the proof of part (i) there exists $\Lambda \in X^*$ such that $\Lambda(A+V)$ and $\Lambda(B)$ are disjoint, convex subsets of $\mathbb{R}$, the first one lying on the left of the second one, the first one being open. Moreover, $\Lambda(A)$ is a compact subset of $\Lambda(A+V)$. The proof is now easily completed. $\square$

Corollary A.3.3. *If X is a locally convex topological vector space, X^* separates the points of X.*

Proof. given distinct points $x, y \in X$, apply the separation theorem with $A = \{x\}$ and $B = \{y\}$. $\square$

For $x \in X$ and $\Lambda \in X^*$ we define, in analogy with the finite-dimensional situation, $\langle x, \Lambda\rangle = \Lambda x$. The *polar* C^0 of a set $C \subseteq X$ is defined as

$$C^0 = \{\Lambda \in X^* : \langle x, \Lambda\rangle \leq 1 \text{ for all } x \in C\}.$$

Similarly, the *bipolar* is defined as

$$C^{00} = (C^0)^0 = \{x \in X : \langle x, \Lambda\rangle \leq 1 \text{ for all } \Lambda \in C^0\}.$$

Theorem A.3.4 (Bipolar theorem). *The bipolar C^{00} of a subset C of a locally convex topological vector space X equals the closed convex hull of C.*

Proof. It is clear that C^{00} is a convex set containing C, so the closed convex hull A of C is a subset of C^{00}. Suppose that the reverse inclusion does not hold. Then there exists a point $x_0 \in C^{00}$ that is not in A. By the separation theorem there then exists a functional $\Lambda \in X^*$ such that $\Lambda x_0 > 1 > \Lambda y$ for all $y \in A$ (check). The second inequality implies that $\Lambda \in C^0$, and then the first one implies that $x_0 \notin C^{00}$, which is a contradiction. $\square$

A.4 Dual space

The *dual* of a topological vector space X is the space X^* of continuous linear functionals on X. By Theorem A.2.5 this is the same as the space of linear functionals that are bounded on a neighborhood of 0.

It is easy to see that if the topology on X is induced by a *norm* $\|\cdot\|$, a linear functional Λ belongs to X^* if and only if the unit ball in X is mapped into a bounded subset of $\mathbb{R}$. In that case we define the *norm of* Λ by

$$\|\Lambda\| = \sup_{\|x\|\leq 1} |\Lambda(x)|$$

and we have the relation $|\Lambda(x)| \leq \|\Lambda\| \, \|x\|$ for every $x \in X$.

Example A.4.1. Let $(E, \mathcal{E}, \mu)$ be a measure space with μ a finite measure, $p \in [1, \infty)$ and $X = L^p(E, \mathcal{E}, \mu)$. Consider a continuous linear functional Λ on X. Then the map $\nu : \mathcal{E} \to \mathbb{R}$ defined by $\nu(B) = \Lambda(1_B)$ is a signed measure (note that the finiteness of μ implies that ν is well-defined). Indeed, if B_n are disjoint elements of $\mathcal{E}$ and $B = \cup B_n$, then $1_{\cup_{k \leq n} B_k} \to 1_B$ in L^p. Since Λ is continuous, this implies that ν is countably additive. If $\mu(B) = 0$ then 1_B vanishes in L^p and hence $\nu(B) = 0$, so $\nu \ll \mu$. Hence, by the Radon-Nikodym theorem, there exists a $g \in L^1$ such that

$$\Lambda(1_B) = \nu(B) = \int_B g \, d\mu$$

for all $B \in \mathcal{E}$. By linearity we then have

$$\Lambda(f) = \int fg \, d\mu \tag{A.2}$$

for all simple functions f. Every bounded measurable function f is the uniform limit of simple functions and since μ is finite, uniform convergence implies convergence in L^p. It follows that (A.2) holds for all $f \in L^\infty$.

Suppose that $p > 1$ and let q be the conjugate exponent. For $E_n = \{x : |g(x)| \leq n\}$ we have, since g is bounded on E_n and Λ is continuous and hence bounded,

$$\int_{E_n} |g|^q \, d\mu = \int_{E_n} |g|^{q-1} \mathrm{sign}(g) g \, d\mu = \Lambda(1_{E_n} |g|^{q-1} \mathrm{sign}(g)) \leq \|\Lambda\| \Big(\int_{E_n} |g|^q \, d\mu \Big)^{1/p}.$$

It follows that

$$\Big(\int_{E_n} |g|^q \, d\mu \Big)^{1/q} \leq \|\Lambda\| \tag{A.3}$$

and letting $n \to \infty$ shows that $g \in L^q$. If $p = 1$ then for every $B \in \mathcal{E}$ we have

$$\Big| \int_B g \, d\mu \Big| = |\Lambda(1_B)| \leq \|\Lambda\| \mu(B).$$

But this implies that $|g| \leq \|\Lambda\|$ a.e. (indeed: if not there would exist an $\varepsilon > 0$ such that the set $B = \{x : |g(x)| > \|\Lambda\| + \varepsilon\}$ has positive μ-measure, leading to a contradiction), hence $g \in L^\infty$.

So in all cases the function g in (A.2) belongs to L^q. We proved already that (A.2) holds for all bounded functions f. Now Λ is continuous on L^p by assumption and Hölders inequality implies that the right-hand side is continuous for $f \in L^p$ as well. This shows that the relation holds in fact for all $f \in L^p$. Uniqueness of g is easy to prove. We conclude that we may identify the dual of L^p with L^q. Moreover, using (A.3) it is easy to see that for $\Lambda \in (L^p)^*$ given by (A.2), we have $\|\Lambda\| = \|g\|_{L^q}$ (Exercise 4). ■

Let X be a topological vector space with dual X^*. Every point $x \in X$ induces a linear functional on X^*, defined by $\Lambda \mapsto \Lambda x$. The *weak*-topology* of X^* is the weakest (i.e. smallest) topology making all these maps continuous.

The following theorem states that X^* with the weak*-topology is a locally convex topological vector space. This implies for instance that we can apply the separation theorem to it. In general, the space X^* endowed with the weak*-topology is not a Banach space. (In fact, it is not even metrizable if X is an infinite-dimensional Banach space.)

Theorem A.4.2. *The dual X^* of a topological vector space X, endowed with the weak*-topology, is a locally convex topological vector space. Its dual is given by $\{\Lambda \mapsto \Lambda x : x \in X\}$.*

Proof. Denote by f_x be the linear functional $\Lambda \mapsto \Lambda x$. If $\Lambda \neq \Lambda'$ in X^*, there exists an $x \in X$ such that $f_x\Lambda \neq f_x\Lambda'$. Hence, in $\mathbb{R}$ there exist disjoint neighborhoods U of $f_x\Lambda$ and U' of $f_x\Lambda'$. Since f_x is continuous, $f_x^{-1}(U)$ and $f_x^{-1}(U')$ are disjoint neighborhoods of Λ and Λ'. This shows that X^* is Hausdorff, and in particular that points are closed.

To show that the weak*-topology is translation invariant, consider an open base set

$$U = \{\Lambda : \Lambda x_1 \in B_1, \dots, \Lambda x_n \in B_n\}$$

and $\Lambda' \in X^*$. Then $\Lambda' + U = \{\Lambda : \Lambda x_1 \in B_1 + \Lambda' x_1 \dots, \Lambda x_n \in B_n + \Lambda' x_n\}$ is an open base set as well. It follows that the topology is translation invariant. Note that the open sets V of the form

$$V = \{\Lambda : |\Lambda x_1| < r_1, \dots, |\Lambda x_n| < r_n\} \tag{A.4}$$

for $x_1, \dots, x_n \in X$ and $r_1, \dots, r_n > 0$ form a local base at 0. Every such set V is convex, balanced and absorbing (check). In particular, X^* is a locally convex space.

For the set V in the preceding display we have $V/2 + V/2 = V$ and hence addition is continuous at $(0,0)$. As for scalar multiplication, suppose that $\alpha\Lambda \in V$ for some scalar $\alpha \in \mathbb{R}$ and $\Lambda \in X^*$. By Exercise 2, there exists $t > 0$ such that $t < 1/|\alpha|$ and $\Lambda \in tV$. For $\varepsilon > 0$ and $\Lambda' \in tV$ we have that $(\alpha+\varepsilon)\Lambda' \in (\alpha+\varepsilon)tV$. Hence, since V is balanced, $(\alpha+\varepsilon)\Lambda' \in V$ for all ε such that $|\alpha|t + |\varepsilon|t \leq 1$. Since $|\alpha|t < 1$ there is a nonempty interval around 0 of ε satisfying this condition. Hence, scalar multiplication is continuous.

It remains to identify the dual of X^* (endowed with the weak*-topology). If $x \in X$, the linear map $\Lambda \mapsto \Lambda(x)$ is weak*-continuous by definition of the weak*-topology. Conversely, let $f : X^* \to \mathbb{R}$ be weak*-continuous. By Theorem A.2.5, f is bounded on a neighborhood of 0, and hence also on a base set V of the form (A.4). This implies that f vanishes on the set $N = \{\Lambda : \Lambda x_1 = \cdots = \Lambda x_n = 0\}$ (Exercise 5). Now N is the kernel of the linear map $\pi : X^* \to \mathbb{R}^n$ defined by $\pi(\Lambda) = (\Lambda x_1, \dots, \Lambda x_n)$. It follows that the linear map $F : \pi(X^*) \to \mathbb{R}$ given by $F(\pi(\Lambda)) = f(\Lambda)$ is well defined (check). We can extend F to a linear functional on $\mathbb{R}^n$. It is then necessarily of the form $F(z_1, \dots, z_n) = \sum \alpha_i z_i$ for certain real numbers α_i. In particular,

$$f(\Lambda) = F(\Lambda x_1, \dots, \Lambda x_n) = \sum \alpha_i \Lambda x_i.$$

So indeed, $f(\Lambda) = \Lambda x$, with $x = \sum \alpha_i x_i$. $\square$

If X is a Banach space its dual X^* is endowed with a norm, and the unit ball in X^* is the set $\{\Lambda \in X^* : |\Lambda x| \leq \|x\| \text{ for all } x \in X\}$. In the norm-topology this set is not compact in general (think of an infinite-dimensional Hilbert space). In the weak*-topology however, it is always compact.

Theorem A.4.3 (Banach-Alaoglu). *The unit ball of the dual of a Banach space is weak*-compact.*

Proof. Denote the Banach space by X and let B^* be the unit ball in its dual. By Tychonov's theorem, $P = \Pi_{x\in X}[-\|x\|, \|x\|]$ is compact (relative to the product topology). We can view P as a collection of functions on X, with $f \in P$ if and only if $|f(x)| \leq \|x\|$ for all $x \in X$. As such, we have $B^* \subseteq X^* \cap P$. Hence, B^* inherits two topologies: the weak*-topology from X^* and the product topology from P. These two topologies on B^* coincide. To see this, take $\Lambda_0 \in B^*$. The sets of the form

$$V_1 = \{\Lambda \in X^* : |\Lambda x_1 - \Lambda_0 x_1| < r_1, \dots, |\Lambda x_n - \Lambda_0 x_1| < r_n\}$$

and

$$V_2 = \{f \in P : |f(x_1) - \Lambda_0 x_1| < r_1, \dots, |f(x_n) - \Lambda_0 x_1| < r_n\}$$

form a local base for the weak*-topology and, respectively, the product topology at Λ_0. Since $B^* \subseteq X^* \cap P$ we have $V_1 \cap B^* = V_2 \cap B^*$ and hence the two relative topologies coincide.

Next we show that B^* is closed in P. Take f_0 in the closure of B^* (with respect to the product topology). For $x, y \in X$, $\alpha, \beta \in \mathbb{R}$ and $\varepsilon > 0$ we have that the set

$$U = \{f \in P : |f(x) - f_0(x)| < \varepsilon, |f(y) - f_0(y)| < \varepsilon, |f(\alpha x + \beta y) - f_0(\alpha x + \beta y)| < \varepsilon\}$$

is an open neighborhood of f_0. Hence, there exist an $f \in U \cap B^*$. Since f is linear we have

$$\begin{aligned} &f_0(\alpha x + \beta y) - \alpha f_0(x) - \beta f_0(y) \\ &\quad = (f_0 - f)(\alpha x + \beta y) - \alpha(f_0 - f)(x) - \beta(f_0 - f)(y) \end{aligned}$$

and hence

$$|f_0(\alpha x + \beta y) - \alpha f_0(x) - \beta f_0(y)| \leq (1 + |\alpha| + |\beta|)\varepsilon.$$

Since ε was arbitrary, it follows that f_0 is linear. By definition of P we have that $|f_0(x)| \leq \|x\|$ for every $x \in X$, so indeed $f_0 \in B^*$.

The proof is now completed upon noting that by the preceding paragraph, B^* is compact with respect to the product topology. But by the first part of the proof, the latter topology coincides on B^* with the weak*-topology. $\square$

Example A.4.4. Although the weak*-topology has some nice properties according to Theorem A.4.2, it is good to note that it is typically "strange". Consider for instance a finite measure μ on the line and view $L^\infty(\mu)$ as the dual of $L^1(\mu)$. Then from the form of the local base at 0 given in the proof of the theorem one sees that a sequence f_n in L^∞ converges in the weak*-topology to 0 if $\int f_n g \, d\mu \to 0$ for every $g \in L^1$. By dominated convergence, this holds for instance for $f_n = 1_{(-n,n)^c}$. This sequence does however not converge to 0 in the ordinary, uniform topology on L^∞. More generally, to say that a function $f \in L^\infty$ belongs to the weak*-closure of a set $C \subseteq L^\infty$ does not necessarily mean that f is well-approximated by elements of C in a uniform or any other intuitively reasonable way. ■

A.5 Exercises

1. Give an example which shows that the separation theorem does not hold in general if the assumption of compactness of one of the sets in dropped.

2. Suppose that C is an open neighborhood of 0 in a topological vector space and let μ_C be its Minkowsky functional. Show that for all $x \in C$ it holds that $\mu_C(x) < 1$.

3. Show that a non-constant linear functional on a topological vector space maps open sets to open sets.

4. In Example A.4.1, show that for the functional Λ on L^p defined by (A.2) we have $\|\Lambda\| = \|g\|_{L^q}$.

5. In the last part of the proof of Theorem A.4.2, show that the functional f vanishes on the set N.

B

Elements of martingale theory

B.1 Basic definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A collection of $\mathbb{R}^d$-valued random variables $X = (X_t)_{t \in T}$ indexed by a set $T \subseteq \mathbb{R}$ is called a *(d-dimensional) stochastic process.* We call the process *continuous* (or *cadlag*), it its trajectories $t \mapsto X_t(\omega)$ are continuous (or cadlag). The process is called *bounded* if there exists a finite number K such that a.s. $\|X_t\| \leq K$ for all t.

A *filtration* is a collection $(\mathcal{F}_t)_{t \in T}$ of sub-σ-fields of $\mathcal{F}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$. It is said to satisfy the *usual conditions* if it is *right-continuous*, i.e. $\cap_{s>t} \mathcal{F}_s = \mathcal{F}_t$ for all t and $\mathcal{F}_0$ contains all the $\mathbb{P}$-null sets in $\mathcal{F}$. A process X is called *adapted* to $(\mathcal{F}_t)$ is for every t, X_t is $\mathcal{F}_t$-measurable. For a process X and $t \in T$ we define $\mathcal{F}_t^X$ to be the σ-field generated by the collection of random variables $\{X_s : s \leq t\}$. The filtration $(\mathcal{F}_t^X)$ is called the *natural filtration* of the process X. It is the smallest filtration to which it is adapted. A process $X = (X_t)_{t \in [0,T]}$ is called *progressively measurable* relative to the filtration $(\mathcal{F}_t)_{t \in [0,T]}$ if for all t, the map $(\omega, s) \mapsto X_s(\omega)$ on $\Omega \times [0,t]$ is $\mathcal{F}_t \otimes \mathcal{B}([0,t])$-measurable.

A $[0,\infty]$-valued random variable τ is called a *stopping time* relative to the filtration $(\mathcal{F}_t)$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for every t. If τ is a stopping time and X a process, the *stopped process* X^τ is defined by $X_t^\tau = X_{\tau \wedge t}$. A *localizing sequence* is a sequence of stopping times τ_n increasing a.s. to infinity. A process X is said to have a property P *locally* if there exists a localizing sequence τ_n such that for every n, the stopped process X^{τ_n} has the property P.

A process M is called a *martingale* relative to the filtration $(\mathcal{F}_t)$ if every M_t is integrable and for all $s \leq t$ it holds that $\mathbb{E}(M_t \,|\, \mathcal{F}_s) = M_s$ a.s.. In accordance with the previously introduced notation the process M is called a *local martingale* if there exists a localizing sequence τ_n such that for every n, the stopped process M^{τ_n} is a martingale. Every martingale is a local martingale, but not vice versa..

B.2 Theorems

For a filtration $(\mathcal{F}_t)$ and a stopping time τ we define

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t\}.$$

The set $\mathcal{F}_\tau$ is always a σ-field and should be thought of as the collection of events describing the history before time τ.

Theorem B.2.1 (Optional stopping theorem). *Let M be a cadlag, uniformly integrable martingale. Then for all stopping times $\sigma \leq \tau$,*

$$\mathbb{E}(M_\tau \mid \mathcal{F}_\sigma) = M_\sigma.$$

Theorem B.2.2 (Kakutani's theorem). *Let $X_1, X_2, \ldots$ be independent non-negative random variables with mean 1. Define $M_0 = 1$ and $M_n = X_1 X_2 \cdots X_n$. It holds that M is uniformly integrable if and only if $\sum(1 - \mathbb{E}\sqrt{X_n}) < \infty$. If M is not uniformly integrable, then $M_n \to 0$ a.s..*

Corollary B.2.3. *Let $X = (X_1, X_2, \ldots)$ and $Y = (Y_1, Y_2, \ldots)$ be two sequences of independent random variables. Assume X_i has a positive density f_i with respect to a dominating measure μ, and Y_i has a positive density g_i with respect to μ. Then the laws of the sequences X and Y are equivalent probability measures on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ if and only if*

$$\sum_{i=1}^{n} \int (\sqrt{f_i} - \sqrt{g_i})^2 \, d\mu < \infty.$$

If the laws are not equivalent, they are mutually singular.

Proof. Let $(\Omega, \mathcal{F}) = (\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ and $Z = (Z_1, Z_2, \ldots)$ the coordinate process on $(\Omega, \mathcal{F})$, so $Z_i(\omega) = \omega_i$. Let $\mathcal{F}_n \subseteq \mathcal{F}$ be the σ-field generated by $Z_1, \ldots, Z_n$. Since the densities f_i and g_i are all positive, the distributions $\mathbb{P}_X$ and $\mathbb{P}_Y$ of the sequences X and Y are equivalent on $\mathcal{F}_n$. For $A \in \mathcal{F}_n$ we have

$$\mathbb{P}_X(A) = \int_A M_n \, d\mathbb{P}_Y,$$

where the Radon Nikodym derivative is defined by $M_n = \prod_{i=1}^n f_i(Z_i)/g_i(Z_i)$. Observe that under $\mathbb{P}_Y$, the process M is a martingale to which the preceding theorem applies. It is readily verified that the measures $\mathbb{P}_X$ and $\mathbb{P}_Y$ are equivalent on the whole σ-field $\mathcal{F}$ if and only if M is uniformly integrable with respect to $\mathbb{P}_Y$ (Exercise 1). Hence, by the preceding theorem, the measures are equivalent if and only if

$$\sum_{i=1}^{n}\Big(1-\int\sqrt{f_i g_i}\,d\mu\Big)<\infty.$$

The proof of the first part is completed by noting that $\int(\sqrt{f_i}-\sqrt{g_i})^2\,d\mu = 2-2\int\sqrt{f_i g_i}\,d\mu$.

We noted that if $\mathbb{P}_X$ and $\mathbb{P}_Y$ are not equivalent, then M is not uniformly integrable relative to $\mathbb{P}_Y$. Hence, by the preceding theorem, $M_n \to 0$, $\mathbb{P}_Y$-a.s.. We can reverse the roles of X and Y, which amounts to replacing M by $1/M$. Then we find that if $\mathbb{P}_X$ and $\mathbb{P}_Y$ are not equivalent, $1/M_n \to 0$, $\mathbb{P}_X$-a.s.. It follows that for the event $A = \{M_n \to 0\}$ we have $\mathbb{P}_Y(A) = 1$ and $\mathbb{P}_X(A) = 0$. □

Example B.2.4. Let $X = (X_1, X_2, \dots)$ and $Y = (Y_1, Y_2, \dots)$ be two sequences of independent random variables. Suppose that $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$ and $\mathbb{P}(Y_i = 1) = 1 - \mathbb{P}(Y_i) = -1 = 1/2 + \varepsilon_i$ for some $\varepsilon_i \in (-1/2, 1/2)$. By the corollary, applied with μ the counting measure, $f_i(1) = f_i(-1) = 1/2$, $g_i(1) = 1 - g_i(-1) = 1/2 + \varepsilon_i$, the laws of the sequences X and Y are equivalent if and only if

$$\sum\Big((\sqrt{1/2}-\sqrt{1/2+\varepsilon_i})^2+(\sqrt{1/2}-\sqrt{1/2-\varepsilon_i})^2\Big)<\infty.$$

By Taylor's formula the function $h(x) = (\sqrt{1/2}-\sqrt{1/2+x})^2 + (\sqrt{1/2}-\sqrt{1/2-x})^2$ behaves like a multiple of x^2 near $x = 0$ (check!). It follows that the sequences are equivalent if and only if $\sum \varepsilon_i^2 < \infty$.

B.3 Exercises

1. In the proof of Corollary B.2.3, show that the measures $\mathbb{P}_X$ and $\mathbb{P}_Y$ are equivalent on the whole σ-field $\mathcal{F}$ if and only if M is uniformly integrable.

References

Conway, J. B. (1990). *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.

Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.* **300**(3), 463–520.

Delbaen, F. and Schachermayer, W. (2006). *The mathematics of arbitrage*. Springer Finance. Springer-Verlag, Berlin.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Process. Appl.* **11**(3), 215–260.

Kreps, D. M. (1981). Arbitrage and equilibrium in economies with infinitely many commodities. *J. Math. Econom.* **8**(1), 15–35.

Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition.

Rudin, W. (1991). *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, second edition.