Lecture Notes for Week 6

Convex Sets, Internal Points, and Separation by Linear Functionals (Continued)

We now state and prove a basic result concerning separation of convex sets in a linear space.

Theorem 6.1: Let X be a linear space over $\mathbb{K}$ and K_1, K_2 be nonempty convex subsets of X. Assume that $K_1 \cap K_2 = \emptyset$ and that K_1 has an internal point. Then there is a nonzero linear functional that separates K_1 and K_2.

Proof: Assume first that $\mathbb{K} = \mathbb{R}$. Choose an internal point $x_1 \in K_1$ and any point $x_2 \in K_2$. Put $x_0 = x_1 - x_2$. Then $K_1 - K_2$ is convex and has x_0 as an internal point. (The convexity of $K_1 - K_2$ follows from Proposition 5.12. Notice that $x_1 - x_2$ is an internal point of $-x_2 + K_1$ because for every $x \in X$ and $\lambda \in \mathbb{R}$ we have $x_1 - x_2 + \lambda x \in -x_2 + K_1$ if and only if $x_1 + \lambda x \in K_1$. Since $-x_2 + K_1 \subset K_1 - K_2$, it follows that $x_1 - x_2$ is an internal point of $K_1 - K_2$.) Put

$$K = -x_0 + (K_1 - K_2)$$

and observe that K is convex and has 0 as an internal point. Since $K_1 \cap K_2 = \emptyset$, we know that $0 \notin K_1 - K_2$. This implies that $-x_0 \notin K$ and consequently $p^K(-x_0) \geq 1$ by part (b) of Lemma 5.32. Let $Y = \text{span}(\{x_0\})$ and define the linear functional $f : Y \to \mathbb{R}$ by

$$f(\alpha x_0) = -\alpha \ \text{ for all } \alpha \in \mathbb{R}.$$

By parts (a) and (c) of Lemma 5.32 we know that p^K can be used as the comparison functional in the Hahn-Banach theorem for real linear spaces. If $\alpha \geq 0$, then

$$f(\alpha x_0) \leq 0 \leq p^K(\alpha x_0).$$

If $\alpha < 0$, then we have

$$f(\alpha x_0) = -\alpha \leq -\alpha p^K(-x_0) = p^K(\alpha x_0),$$

since $p^K(-x_0) \geq 1$ and nonnegative scalars can be factored out of p^K. It follows that

$$f(x) \leq p(x) \ \text{ for all } x \in Y.$$

By the Hahn-Banach theorem for real spaces, we may choose a linear functional $F : X \to \mathbb{R}$ such that $F(x) = f(x)$ for all $x \in Y$ and

$$F(x) \leq p^K(x) \ \text{ for all } x \in X.$$

We need to show that F separates K_1 and K_2. Let $x \in K_1$, $y \in K_2$ be given and notice that $x - y - x_0 \in K$ which tells us that $p^K(x - y - x_0) \leq 1$. Using the linearity of F, we find that

$$F(x) - F(y) - F(x_0) = F(x - y - x_0) \leq p^K(x - y - x_0) \leq 1.$$

Since $F(x_0) = -1$, we obtain

$$F(x) - F(y) \leq 0 \text{ for all } x \in K_1, y \in K_2.$$

By Lemma 5.34, F separates K_1 and K_2.

Assume now that $\mathbb{K} = \mathbb{C}$. Let X_r denote the linear space obtained by restricting X to the subfield of real scalars. Then K_1, K_2 are nonempty disjoint convex subsets of X_r and K_1 has an internal point. We may apply the separation theorem for real spaces to obtain a linear functional $F_1 : X_r \to \mathbb{R}$ satisfying

$$F_1(x) - F_1(y) \leq 0 \text{ for all } x \in K_1, y \in K_2.$$

Now F_1 satisfies

$$F_1(x + y) = F_1(x) + F_1(y), \quad F_1(ax) = aF_1(x) \text{ for all } x, y \in X, a \in \mathbb{R}.$$

Let us define $F : X \to \mathbb{C}$ by

$$F(z) = F_1(z) - iF_1(iz) \text{ for all } z \in \mathbb{C}.$$

It is straightforward to check that F is a linear functional (with respect to the linear structure over $\mathbb{C}$). Moreover, since $\text{Re}(F(z)) = F_1(z)$ for all $z \in X$, we have

$$\text{Re}(F(x)) - \text{Re}(F(y)) \leq 0 \text{ for all } x \in K_1, y \in K_2,$$

so that F separates K_1 and K_2 by Lemma 5.34. $\square$

Theorem 6.1 has numerous important applications and numerous important generalizations and variants. Several remarks pertaining to generalizations and variants are given below.

Remark 6.2: Convexity of both sets is essential in Theorem 6.1, even when X is finite dimensional. To see why, let $X = \mathbb{R}^2$,

$$K_1 = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \leq e^{-x_1}\}, \quad K_2 = \{(0, 2)\}.$$

You should convince yourself that any line which is above $(0, 1)$ and below $(0, 2)$ will have a nonempty intersection with K_2.

Remark 6.3: If X is finite dimensional the assumption that K_1 has an internal point can be dropped and the assumption that $K_1 \cap K_2 = \emptyset$ can be weakened.

Remark 6.4: The separation ensured by Theorem 6.1 need not be strict, even when X is finite dimensional. (Here by strict separation we mean that either

$$\mathrm{Re}(F(x)) - \mathrm{Re}(F(y)) < 0 \ \text{ for all } x \in M, \ y \in N$$

or

$$\mathrm{Re}(F(x)) - \mathrm{Re}(F(y)) > 0 \ \text{ for all } x \in M, \ y \in N.)$$

Strict separation or even separation with a "gap" can be ensured by additional assumptions.

Remark 6.5: There are numerous important variants of the basic separation theorem. See *Convex Analysis and Variational Problems* by I. Ekeland and R. Temam, or *Convex Analysis* by R. T. Rockafellar.

If X is normed, it is often desirable to separate two convex sets by a continuous linear functional. The assumptions made on K_1 and K_2 in Theorem 6.1 are not sufficient for this purpose. However, if the assumption that K_1 has an internal point is strengthened to $\mathrm{int}(K_1) \neq \emptyset$ then the separation of K_1 and K_2 can be achieved by an element of X^*.

Theorem 6.6: Let X be a normed linear space over $\mathbb{K}$ and let K_1, K_2 be nonempty convex subset of X. Assume that $K_1 \cap K_2 \neq \emptyset$ and that $\mathrm{int}(K_1) = \emptyset$. Then there is a nonzero continuous linear functional that separates K_1 and K_2.

The proof of Theorem 6.6 is part of Assignment 4.

Seminorms and Convexity

Let X be a linear space over $\mathbb{K}$.

Definition 6.7: A function $p : X \to \mathbb{R}$ is said to be a *seminorm* provided that

(i) $\forall x, y \in X$, we have $p(x + y) \leq p(x) + p(y)$,

(ii) $\forall x \in X, \alpha \in \mathbb{K}$, we have $p(\alpha x) = |\alpha| p(x)$.

Notice that seminorms are precisely the kind of comparison functions appearing in the Hahn-Banach theorem for real or complex spaces.

Proposition 6.8: Assume that $p : X \to \mathbb{R}$ is a seminorm. Then

(a) $p(0) = 0$,

(b) $p(x) \geq 0$ for all $x \in X$,

(c) $|p(x) - p(y)| \leq p(x - y)$,

(d) Put $Y = \{x \in X : p(x) = 0\}$. Then Y is a linear manifold,

(e) Put $K = \{x \in X : p(x) < 1\}$. Then, K is convex, balanced, and absorbing. Moreover, $p = p^K$, where p^K is the Minskowski functional for K.

Proof: To prove (a), we let $x \in X$ be given and notice that $p(0) = p(0x) = 0p(x) = 0$.

To prove (b) we let $x \in X$ be given and observe that

$$2p(x) = p(x) + p(-x) \geq p(x - x) = 0.$$

To prove (c), let $x, y \in X$ be given. Then we have

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

$$p(y) - p(x) = p(y - x + x) - p(x) \leq p(y - x) + p(x) - p(x) = p(y - x) = p(x - y).$$

It follows that $|p(x) - p(y)| \leq p(x - y)$.

To prove (d), we first observe that $0 \in Y$ by (a). Now, let $x, y \in Y, \alpha, \beta \in \mathbb{K}$ be given. Then, since $p(x) = p(y) = 0$, we have

$$0 \leq p(\alpha x + \beta y) \leq p(\alpha x) + p(\beta y) = |\alpha| p(x) + |\beta| p(y) = 0,$$

which implies $\alpha x + \beta y \in Y$.

To show that K is convex, let $x_1, x_2 \in K, t \in [0, 1]$ be given. Then, since $0 \leq p(x_1), p(x_2) < 1$ and $0 \leq t, (1 - t) \leq 1$, we have

$$p(tx_1 + (1 - t)x_2) \leq p(tx_1) + p((1 - t)x_2) = tp(x_1) + (1 - t)p(x_2) < 1.$$

To show that K is balanced, let $\alpha \in \mathbb{K}$ with $|\alpha| \leq 1$ and $x \in \alpha K$ be given. We need to show that $x \in K$. Choose $y \in K$ such that $x = \alpha y$, and note that $p(y) < 1$. Then we have

$$p(x) = p(\alpha y) = |\alpha| p(y) < 1,$$

which tells us that $x \in K$.

To show that K is absorbing, let $x \in X$ be given. If $p(x) = 0$, then $p(\lambda x) = 0$ (which implies $\lambda x \in K$) for all $\lambda \in \mathbb{K}$. If $p(x) > 0$, then $p(\lambda x) < 1$ (i.e., $\lambda x \in K$) for all $\lambda \in \mathbb{K}$ with $|\lambda| < 1/p(x)$.

To show that $p = p^K$, let $x \in X$ and $s > p(x)$ be given. Then we have $p(s^{-1}x) < 1$. This tells us that $s^{-1}x \in K$, which means that $x \in sK$. We also conclude that $p^K(x) \leq s$. Since

$$p^K(x) \leq s \text{ for all } s > p(x),$$

it follows $p^K(x) \leq p(x)$.

To establish the reverse inequality, let $x \in X$ be given and assume that $p(x) > 0$.(If $p(x) = 0$, then trivially, $p(x) \leq p^K(x)$.) Let $t \in \mathbb{R}$ with $0 < t \leq p(x)$ be given. Then $p(t^{-1}x) \geq 1$ which tells us that $t^{-1}x \notin K$. Since this holds for every $t \in (0, p(x)]$, we conclude that $p(x) \leq p^K(x)$. $\square$

Definition 6.9: A family $(p_i | i \in I)$ of seminorms is said to be *separating* provided that

$$\forall x \in X \backslash \{0\}, \ \exists i \in I \text{ such that } p_i(x) > 0.$$

Separating families of seminorms can be used to generate Hausdorff topologies on linear spaces such that addition and scalar multiplication are continuous (and having some useful additional properties).

Remark 6.10: Observe that a family $(p_i | i \in I)$ of seminorms is separating if and only if

$$\{x \in X : p_i(x) = 0 \text{ for all } i \in I\} = \{0\}.$$

Remark 6.11: If a finite family $(p_i | i = 1, 2, \cdots, N)$ of seminorms is separating then clearly $p_1 + p_2 + \cdots + p_N$ is a norm on X.

Remark 6.12: Let $(p_i | i \in \mathbb{N})$ be a countable family of seminorms that is separating and define $\rho : X \times X \to \mathbb{R}$ by

$$\rho(x, y) = \sum_{i=1}^{\infty} 2^{-i} \frac{p_i(x - y)}{1 + p_i(x - y)} \quad \text{for all } x, y \in X.$$

Then ρ is a translation invariant metric such that each p_i is continuous from (X, ρ) to $\mathbb{R}$. (Here translation invariance simply means that $\rho(x + z, y + z) = \rho(x, y)$ for all $x, y, z \in X$.) One could replace $(2^{-i} | i \in \mathbb{N})$ with another sequence $(a_i | i \in \mathbb{N})$ with $a_i > 0$ for all $i \in \mathbb{N}$ and $\sum_{i=1}^{\infty} a_i < \infty$. For some purposes, it is more convenient to use the metric given by

$$\hat{\rho}(x, y) = \max \left\{ \frac{1}{i} \min\{1, p_i(x - y)\} : i \in \mathbb{N} \right\}.$$

Remark 6.13: Any separating family $(p_i | i \in I)$ of seminorms can be used to construct a Hausdorff topology on X by taking finite intersections of sets of the form

$$\{x \in X : p_i(x - x_0) < \epsilon\}, \quad x_0 \in X, \epsilon > 0$$

as a base.

Example 6.14: Let $\mathbb{K} = \mathbb{R}$ and $X = C(\mathbb{R}; \mathbb{R})$, the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$. For each $i \in \mathbb{N}$, put

$$p_i(f) = \max\{|f(x)| : x \in [-i, i]\} \quad \text{for all } f \in X.$$

It is clear that each p_i is a seminorm (but not a norm). Now define

$$\rho(f,g) = \sum_{i=1}^{\infty} 2^{-i} \frac{p_i(f-g)}{1+p_i(f-g)} \quad \text{for all } x, y \in X.$$

Then (X,ρ) is a complete metric space. You should check as an exercise that $f_n \to f$ as $n \to \infty$ in (X,ρ) if and only if for each compact subset $K \subset \mathbb{R}$, f_n converges to f uniformly on K as $n \to \infty$. The topology induced by this metric is a very natural. Since the evaluation mappings $a \to f(a)$ are continuous for every $a \in \mathbb{R}$, it follows from Exercise 14 on Assignment 2, that this topology cannot be induced by a norm.

Remark 6.15: A linear space together with a topology induced by a complete invariant metric (such that scalar multiplication is continuous) is called a Fréchet space. Some authors require the topology to be induced by a metric for which open balls are convex. (It is important to note that two different metrics can induce the same topology and one of the metrics has the properties that all open balls are convex, while the other has the property that no open balls are convex.) We will discuss such spaces in detail later in the course.

Example 6.16: l^p with $0 < p < 1$. Given $p \in (0,1)$, let l^p denote the set of all $x \in \mathbb{K}^{\mathbb{N}}$ such that

$$\sum_{k=1}^{\infty} |x_k|^p < \infty.$$

It can be shown that l^p is a linear space and that the function $\rho : X \times X \to \mathbb{R}$ defined by

$$\rho(x,y) = \sum_{k=1}^{\infty} |x_k - y_k|^p$$

is a complete invariant metric on l^p. Balls in this metric are not convex. Moreover, it can be shown that there is no complete invariant metric having balls which are convex that induces the same topology. This lack of convexity in the topology leads to some unusual properties. In particular the only continuous linear functional on l^p is the zero functional.

Duality and Reflexivity

Let X and Y be NLS over the same field.

Definition 6.17: We say that X and Y are *isomorphic* provided that there is a bijective linear mapping $T : X \to Y$ such that T and T^{-1} are continuous. Any such mapping T is called an *isomorphism* between X and Y.

Definition 6.18: A linear mapping $T : X \to Y$ is said to be an *isometry* provided that $\|Tx\| = \|x\|$ for all $x \in X$.

Remark 6.19:

(a) If T is an isometry, then T is continuous and injective.

(b) If an isometry T is surjective (and hence bijective) then T^{-1} is an isometry.

Definition 6.20: We say that X and Y are *isometrically isomorphic* provided that there is a surjective (linear) isometry $T : X \to Y$.

Let X^* denote the (topological) dual of X and $\langle \cdot, \cdot \rangle : X^* \times X \to \mathbb{K}$ denote the duality pairing, i.e.

$$\langle x^*, x \rangle = x^*(x) \quad \text{for all } x^*, \in X^*, x \in X.$$

Recall that the canonical injection J of X into X^{**} is defined by

$$(J(x))(x^*) = \langle x^*, x \rangle \quad \text{for all } x \in X, x^* \in X^*.$$

By Proposition 5.6, we have

$$\|J(x)\|_{**} = \sup\{|\langle x^*, x \rangle| : \|x^*\|_* \leq 1\} = \|x\| \quad \text{for all } x \in X,$$

and consequently J is an isometry. (This, of course, ensures that J is indeed injective.)

Definition 6.21: We say that X is reflexive if J is surjective.

Remark 6.22: If X is reflexive then X is isometrically isomorphic to X^{**}. In 1951, R.C. James gave an example of a Banach space X such that X is isometrically isomorphic to X^{**} but X is not reflexive. (*Proc. Nat. Acad. Sci.* **37**, 174-177) Consequently, in the definition of reflexivity, it is essential that the canonical injection of X into X^{**} be an isomorphism (rather than just requiring some isometric isomorphism to exist).

Prop. 6.23: If X is reflexive, then X is complete.

Proof: Let $\{x_n\}_{n=1}^{\infty}$ be a Cauchy sequence in X. Then $\{J(x_n)\}_{n=1}^{\infty}$ is a Cauchy sequence in X^{**} since J is an isometry. X^{**} is complete since it is the dual of X^*. We may choose $x^{**} \in X^{**}$ such that $J(x_n) \to x^{**}$ as $n \to \infty$. Put $x = J^{-1}(x^{**})$. Since J is an isometry, we know that

$$\|x_n - x\| = \|J(x_n) - x^{**}\| \to 0 \quad \text{as } n \to \infty. \quad \Box$$