VI. Convexity

Convex sets

In Analysis, convex sets appear as sensible domains from which to single out solutions of certain problems by extremizing convex fl's. This chapter deals with the simple aspects of this problem. Typical examples include the search for a point at which a given blm takes on its norm, or for a ba from a convex set. In both of these examples, and in general, the convexity of the unit ball in a nls plays an important role. In this chapter, $\mathbb{F} = \mathbb{R}$.

**** convexity is local linearity ****

Let X ls. For any two $x, y \in X$, we call

$$[x \,..\, y] := \{(1-\alpha)x + \alpha y : \alpha \in [0 \,..\, 1]\}$$

the **(closed) interval** spanned by x and y. A subset K of X is called **convex** if it is closed under interval formation, i.e.,

$$x, y \in K \Longrightarrow [x \,..\, y] \subseteq K.$$

Thus convexity is a kind of local linearity. It follows that the intersection of convex sets is convex and that any linear map preserves convexity, i.e., it maps convex sets to convex sets.

The unit ball B as well as the closed unit ball B^- in a nls X are convex since

$$\|(1-\alpha)x + \alpha y\| \le (1-\alpha)\|x\| + \alpha\|y\| \text{ whenever } \alpha \in [0 \,..\, 1].$$

**** convex hull ****

The **convex hull**

$$\operatorname{conv} M$$

of the set M in the ls X is, by definition, the smallest convex set containing M, hence is the intersection of all convex sets containing M.

For example, $\operatorname{conv}\{x, y\} = [x \,..\, y]$ since the latter is convex and must be contained in every convex set containing x and y. Also,

$$\operatorname{conv}\{x, y, z\} = [x \,..\, [y \,..\, z]] = [[x \,..\, y] \,..\, z] \tag{1}$$

since both sets must be contained in any convex set containing $\{x, y, z\}$, and the first (hence, by symmetry, also the second) equals the **2-simplex**

$$\{[x, y, z]a : a \in \mathbb{R}^3_+; \|a\|_1 = 1\}.$$

Indeed, if $u = xa(1) + ya(2) + za(3)$ with a nonnegative and $a(1) + a(2) + a(3) = 1$, then either $a(2) = 0 = a(3)$, hence $u = x$, or else $u = v := x\alpha + (1-\alpha)(y\beta + (1-\beta)z)$ with $\alpha :=$

$a(1), \beta := a(2)/(a(2)+a(3)) \in [0\,..\,1]$, hence, either way, $u \in [x\,..\,[y\,..\,z]]$, while, conversely, any such v is of the form $xa(1)+ya(2)+za(3)$ with $a := (\alpha, (1-\alpha)\beta, (1-\alpha)(1-\beta)) \in \mathbb{R}^3_+$ and $\|a\|_1 = 1$. One verifies similarly that the n**-simplex**

$$\operatorname{conv}(x_0, \ldots, x_n) := \{[x_0, \ldots, x_n]a : a \in \mathbb{R}^{n+1}_+; \|a\|_1 = 1\}$$

is the convex hull of $\{x_0, .., x_n\}$. Its elements are called the **convex combinations** of the x_i's.

For an arbitrary set M with subsets M_i, the convex hull of M contains

$$[M_1\,..\,M_2] := \cup_{x_i \in M_i}[x_1\,..\,x_2],$$

hence contains, with $M^{(1)} := M$, the inductively defined sets

$$M^{(k)} := [M^{(s)}\,..\,M^{(k-s)}], \quad 0 < s < k; k = 2, 3, \ldots,$$

with the various right-hand sides here indeed equal since

$$\operatorname{conv}(x_1, \ldots, x_k) = [\operatorname{conv}(x_1, \ldots, x_s)\,..\,\operatorname{conv}(x_{s+1}, \ldots, x_k)], \quad 0 < s < k.$$

Each element of $M^{(k)}$ is a convex combination of k elements of M, i.e.,

$$M^{(k)} = \cup_{F \subset M; \#F = k} \operatorname{conv} F, \quad \text{with} \quad \operatorname{conv} F = \{[F]a : a \in \mathbb{R}^F_+; \|a\|_1 = 1\}.$$

The sequence $M^{(1)} = M, M^{(2)}, \ldots$ is increasing, and its union is convex since $[M^{(r)}\,..\,M^{(s)}] \subseteq M^{(r+s)}$. Since this union also lies in every convex set containing M, it follows that

$$\operatorname{conv} M = \cup_{n \in \mathbb{N}} M^{(n)}. \tag{2}$$

If M is finite, then $\operatorname{conv} M = M^{(\#M)}$. Whether or not M is finite,

$$\operatorname{conv} M = M^{(\dim X + 1)}$$

this is part of (8)Caratheodory's Theorem below.

H.P.(1) Here is a useful *weakening* of convexity: The set K is said to be **starlike** with respect to the set Y if, $[Y\,..\,K] \subset K$. (Thus, a set is convex iff it is starlike wrto itself.) Prove: *If K is starlike wrto Y, then K is starlike wrto* $\operatorname{conv} Y$. *(Hint: Use (1) to prove $[A\,..\,[B\,..\,C]] \subseteq [[A\,..\,B]\,..\,C]$; then use (2).)*

H.P.(2) Here is a useful *strengthening* of convexity: The set H is said to be a **flat** (or, an **affine** set) if $x, y \in H$ implies $x + \operatorname{ran}[y - x] \subseteq H$. Further, $\flat M :=$ the **affine hull of** M or **flat spanned by** M, is the smallest flat containing M. Prove: (i) $\flat M = \{[F]a : F \subset M, \#F < \infty, \sum_{f \in F} a(f) = 1\}$, *i.e., $\flat M$ is the set of all* **affine combinations** *of elements of M.* (ii) *H is a flat iff $H = x + Y$ for some lss Y (and every $x \in H$).*

** closure and interior of a convex set **

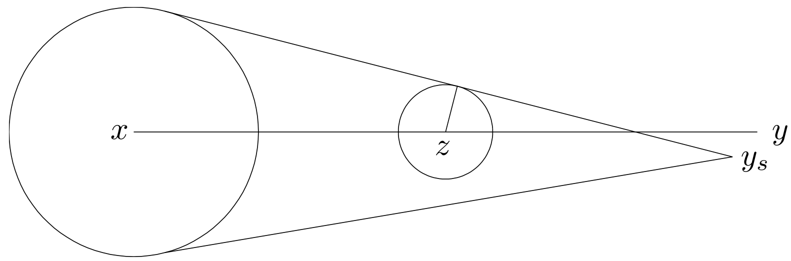
The *topological structure* of a convex set in a nls is quite simple.

(3) Lemma. *K convex, $x \in K^o$, $y \in K^- \Longrightarrow [x\,..\,y) \subseteq K^o$.*

Proof: (4)Figure tells the story, but here are the algebraic facts, to be sure. $x \in K^o \Longrightarrow \exists\{r > 0\}\ x + B_r \subseteq K$, while $y \in K^- \Longrightarrow \forall\{s > 0\}\ \exists\{y_s \in K \cap B_s(y)\}$. Hence, $\forall\{z \in [x\,..\,y)\}\ \exists\{t > 0\}\ z + B_t \subseteq K$, since, with $z =: (1-\alpha)x + \alpha y$, we have $0 \le \alpha < 1$ and

$$\begin{aligned} z + B_t &\subseteq (1-\alpha)x + \alpha y_s \ + \ \alpha(y - y_s) + B_t \\ &\subseteq (1-\alpha)x + \alpha y_s + B_{t+\alpha s} = (1-\alpha)(x + B_r) + \alpha y_s \subseteq K \end{aligned}$$

if $(t + \alpha s)/(1-\alpha) = r$, i.e., $t = (1-\alpha)r - \alpha s$, and this is positive for sufficiently small s since $1 - \alpha > 0$. $\square$



(4) Figure. Proof of (3)Lemma

(5) Corollary. *K convex $\implies K^o, K^-$ convex.*

Proof: If $x, y \in K^o$, then $[x \,..\, y] \subset K^o$ by (3)Lemma, hence K^o is convex. If $x, y \in K^-$, then $x = \lim_n x_n$, $y = \lim_n y_n$ for sequences (x_n), (y_n) in K. Therefore, for every $\alpha \in [0 \,..\, 1]$, $(1-\alpha)x + \alpha y = \lim_n((1-\alpha)x_n + \alpha y_n) \in K^-$, i.e., K^- is convex. $\square$

H.P.(3) Prove: K *convex*, $K^o \ne \{\} \implies (i) K^- = (K^o)^-$, hence $(ii) \forall\{\lambda \in X^*\}\ \sup \lambda K = \sup \lambda(K^o)$.

Remark. The lemma has an extension to any convex set consisting of more than one point, for any such convex set has **relative interior**, i.e., interior as a subset of its affine hull. It has also an extension to linear topologies more general than the norm topology.

For example, the n-simplex, $\sigma_n := \mathrm{conv}(x_0, \ldots, x_n)$, is called **proper** if the flat spanned by it is n-dimensional, i.e., if $[x_1 - x_0, \ldots, x_n - x_0]$ is 1-1. In that case, $\flat\sigma_n$ is the 1-1 affine image of $\mathbb{R}^n$ under the map

$$f : \mathbb{R}^n \to X : a \mapsto x_0 + [x_1 - x_0, \ldots, x_n - x_0]a,$$

with

$$T_n := \mathrm{conv}(e_j : j = 0, \ldots, n), \qquad e_0 := 0,$$

the pre-image of σ_n. Correspondingly, the relative interior of σ_n is the image under f of the interior of T_n. The latter equals

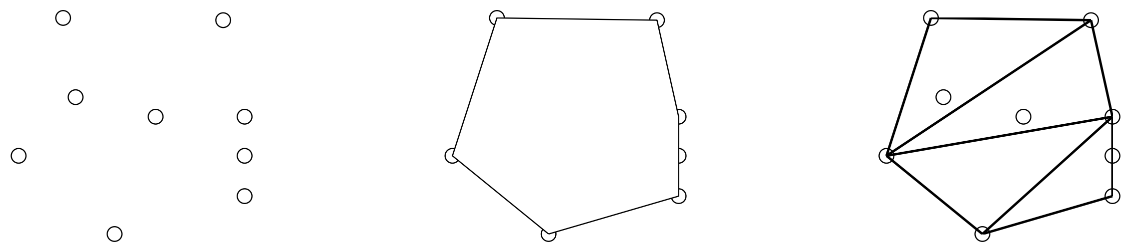
$$T_n^o = \{a \in \mathbb{R}^n : \forall\{j\}\ a(j) > 0; \sum_j a(j) < 1\}. \tag{6}$$

Indeed, the right side here is a subset of T_n and is the intersection of $n+1$ open sets, hence open. On the other hand, any point $a \in T_n$ but not in the right side here has either $a(j) = 0$ for some j or else has $\sum_j a(j) = 1$, hence has points not in T_n in every one of its neighborhoods, and so cannot be in T_n^o.

** Caratheodory **

(8) Caratheodory's Theorem. *If $M \subseteq \mathbb{R}^n$, then $\forall\{x \in \mathrm{conv}\, M\}\ \exists\{F \subseteq M\}\ \#F \le n+1$ and $x \in \mathrm{conv}\, F$. Further, if $x \in \partial\, \mathrm{conv}\, M$, then there is such an F with $\#F \le n$.*

Proof: To write $x \in \mathbb{R}^n$ as a convex combination of F is to write $\hat{x} := (x, 1) \in \mathbb{R}^{n+1}$ as a *nonnegative* combination of $\hat{F} := \{\hat{f} := (f, 1) \in \mathbb{R}^{n+1} : f \in F\}$, i.e., to write $\hat{x} = [\hat{F}]a = \sum_{f \in F}(f, 1)a(f)$ for some $a \in \mathbb{R}_+^F$.



(7) Figure. Caratheodory's Theorem in the plane.

Since $x \in \operatorname{conv} M$, it is in $\operatorname{conv} F$ for some finite subset F of M. Hence there is a *smallest* subset F of M with $x \in \operatorname{conv} F$. We prove that, for such a minimal F, $[\hat{F}]$ *is necessarily 1-1*, hence $\#F = \#\hat{F} \le \dim \mathbb{R}^{n+1} = n+1$.

For this, write $\hat{x} =: [\hat{F}]a$ for some $a \in \mathbb{R}^F$ with $a(f) > 0$ for all $f \in F$, and assume that $0 = [\hat{F}]b$ for some $b \in \mathbb{R}^F\backslash 0$. Then, without loss of generality, $b(f) > 0$ for some $f \in F$, hence $\gamma := \min\{a(f)/b(f) : b(f) > 0\}$ is well defined, and produces $c := a - \gamma b$ that vanishes on some f, yet $c \ge 0$, while

$$[\hat{F}]c = [\hat{F}]a - \gamma[\hat{F}]b = \hat{x} + 0 = \hat{x},$$

hence F is not minimal.

It remains to show that $\#F \le n$ in case x is a boundary point for $\operatorname{conv} M$. Equivalently, if the minimal F for x has $n+1$ elements, then, for any particular $f_0 \in F$, x is in $f_0 + [f - f_0 : f \in F\backslash f_0]T_n^o$ (see (6)) and this is an open set in $\operatorname{conv} F$ (hence in $\operatorname{conv} M$) since $[\hat{F}]$ is 1-1, hence onto, hence so is the linear map $[f - f_0 : f \in F\backslash f_0]$ which is therefore open, by (V.18)OMT. □

The Minkowski functional

The Separation Theorem is a version of HB that is more convenient in the consideration of maximizing convex fl's over convex sets than the original (equivalent) statement. Its proof below uses the Minkowski fl of a convex (absorbing) set as the bounding sublinear fl.

**** Minkowski fl ****

$K \subseteq X$ (ls) is **absorbing** $:\iff$ $X = \bigcup_{r>0} rK$. If K is absorbing, then the rule

$$\mu_K : x \mapsto \inf\{r > 0 : x \in rK\}$$

defines a nonnegative functional on X, the **Minkowski** functional for K. The model example is the Minkowski functional of the unit ball in a nls, which is just the norm:

$$\mu_B = \|\cdot\|.$$

Note that $\mu_B = \mu_K$ with K the *non*convex absorbing set $S \cup 0$. (Recall that $S = \partial B$.)

The Minkowski fl is positive homogeneous. Further, since $L \subseteq K \implies \mu_L \ge \mu_K$ and $\alpha > 0 \implies \mu_{\alpha K} = \mu_K/\alpha$, the Minkowski fl μ_K is bounded (i.e., bounded on (norm)-bounded sets) in case 0 is an interior point of K, and is definite (i.e., $\mu_K(x) = 0 \implies x = 0$) in case K is bounded.

H.P.(4) Verify that $K \subseteq {\mu_K}^{-1}[0\,..\,1]$, that $K^o \subseteq {\mu_K}^{-1}[0\,..\,1)$, and that μ_K is positive homogeneous.

H.P.(5)
(i) Prove: *A sublinear and bounded fl is (Lipschitz-)continuous.*
(ii) Give an example of a discontinuous bounded Minkowski fl.

Now assume that, in addition to being absorbing, K is convex, which is the context in which the Minkowski fl usually arises, for then it is sublinear.

(9) Lemma. *K (absorbing and) convex $\Longrightarrow \mu_K$ is subadditive (hence sublinear).*

Proof: If $x = \alpha x', y = \beta y'$ for $x', y' \in K$ and $\alpha, \beta > 0$, then

$$x + y = \alpha x' + \beta y' = (\alpha+\beta)\Big(\frac{\alpha}{\alpha+\beta}x' + \frac{\beta}{\alpha+\beta}y'\Big) \in (\alpha+\beta)K,$$

therefore $\mu_K(x+y) \le \inf\{\alpha+\beta : \alpha, \beta > 0; x' := x/\alpha, y' := y/\beta \in K\} = \mu_K(x) + \mu_K(y)$.
□

A converse of sorts holds in that ${\mu_K}^{-1}[0\,..\,1]$ is convex in case μ_K is subadditive.

Now assume that, in addition to being convex, K has 0 as an interior point. Then K is not only absorbing, it is even **uniformly absorbing**, i.e., $\forall\{s > 0\}\ \exists\{r > 0\}\ B_s \subseteq rK$.

(10) Lemma. *K convex, $0 \in K^o \Longrightarrow {\mu_K}^{-1}[0\,..\,1) = K^o$.*

Proof: By H.P.(4), it is sufficient to show that ${\mu_K}^{-1}[0\,..\,1) \subseteq K^o$. For this, note that $x \in {\mu_K}^{-1}[0\,..\,1) \Longrightarrow \exists\{r < 1, y \in K\}\ x = ry$, i.e., $x \in [0\,..\,y)$ with $0 \in K^o$ and $y \in K$. Now apply (3)Lemma. □

H.P.(6) Prove: *If K is* **balanced** (i.e., $-K = K$) *and convex with $0 \in K^o$, then μ_K is a continuous seminorm. If K also bounded, then μ_K is a (continuous) norm.* For such K, $K^- = {\mu_K}^{-1}[0\,..\,1]$.

The Separation Theorem

(11) Separation Theorem. *$x \in X$ nls, K convex, $x \notin K^o \ne \{\}$ $\Longrightarrow$ $\exists\{\lambda \in X^*\backslash 0\} \sup \lambda K \le \lambda x$.*

Proof: After a shift, may assume that $0 \in K^o$. Then the Minkowski functional μ_K for K is defined and sublinear. Set $Y := \mathrm{ran}[x]$, define $\lambda_0 : Y \to \mathbb{R} : \alpha x \mapsto \alpha\mu_K(x)$. Then $\lambda_0 \in Y'$ and $\lambda_0 \le \mu_K$ (on Y, of course, where else?). By the general HB Theorem (IV.29), $\exists\{\lambda \in X'\}\ \lambda x = \lambda_0 x$ and $\lambda \le \mu_K$. Since $x \notin K^o$, we have $\mu_K(x) \ge 1$ by (10)Lemma, therefore $\sup \lambda K \le \sup \mu_K K = 1 \le \mu_K(x) = \lambda x$. Further, since $0 \in K^o$, have $B_r \subseteq K$ for some $r > 0$, so λ is bounded on B_r, hence $\lambda \in X^*$. □

** refinements **

(12) Corollary. *K, L convex, $K \cap L^o = \{\} \ne L^o$ $\quad\Longrightarrow\quad$ $\exists\{\lambda \in X^*\backslash 0\}\ \sup \lambda K \le \inf \lambda L$.*

Proof: L^o convex (by (5)Proposition), hence $M := K - L^o$ convex and $0 \notin M = M^o \ne \{\}$. Therefore, by (11), we can separate 0 from M, i.e., $\exists\{\lambda \in X^*\backslash 0\}\ \sup\lambda K - \inf \lambda L^o = \sup \lambda M \le \lambda 0 = 0$. Since $\inf \lambda L^o = \inf \lambda L$ by H.P.(3) it follows that $\sup \lambda K \le \inf \lambda L$. □

H.P.(7) Prove the **Theorem of the Alternative**: *Let $A \in \mathbb{R}^{m\times n}$. Then the following two conditions are mutually exclusive: (i)* $\operatorname{ran} A \cap \{b \in \mathbb{R}^m : b < 0\} \neq \{\}$*; (ii)* $\exists\{p \neq 0\}\ 0 \le p \perp \operatorname{ran} A$.

(13) Corollary. *K closed, convex, $x \notin K \implies \exists\{\lambda \in X^*\}\ \sup \lambda K < \lambda x$.*

Proof: $x \notin K = K^- \implies r := d(x, K) > 0 \implies K \cap B_r(x) = \{\}$ and $B_r(x)^o = B_r(x) \neq \{\}$. By (12)Corollary, $\exists\{\lambda \in X^*\backslash 0\}\ \sup \lambda K \le \inf \lambda B_r(x) = \lambda x + \inf \lambda B_r = \lambda x - \|\lambda\| r < \lambda x$, the last inequality since $\|\lambda\| r \neq 0$. □

H.P.(8) Prove **Farkas' Lemma**: *Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^n$. Then the following are equivalent: (i) $Ay \ge 0 \implies b^t y \ge 0$; (ii) $\exists\{p \ge 0\}\ p^t A = b^t$.*

This lemma is often stated as a Theorem of the Alternative, i.e., not $\exists\{y\}\ Ay \ge 0, b^t y < 0$ if and only if (ii).

H.P.(9) Prove the **Dubovitskii-Milyutin Separation Theorem**: Let $K_0, \ldots, K_n$ be convex sets in the nls X with $0 \in K_i^-$ for all i and K_i open for $i > 0$. Then $\cap_i K_i = \{\}$ iff $(K_i : i = 0, \ldots, n)$ is **separated at** 0 in the sense that there exist $\lambda_0, \ldots, \lambda_n$ in X^* not all zero, with $\sum_i \lambda_i = 0$ and $\inf \lambda_i(K_i) \ge 0$, all i. (Hint: For the 'only if', apply (12)Corollary to the sets $K := \{(x, \ldots, x) \in X^n : x \in K_0\}$ and $L := K_1 \times \cdots \times K_n$ in the nls X^n (with continuous dual $(X^n)^* = (X^*)^n$).)

(14) Corollary. *If, in addition, k is a ba to x from K (i.e., $k \in K$ and $d(x, K) = \|x-k\|$), then $\exists\{\lambda || x - k\}\ \sup \lambda K = \lambda k$.*

Proof: Going through the proof of (13)Corollary a bit more carefully, we find that

$$\lambda k \le \sup \lambda K \le \lambda x - \|\lambda\| r = \lambda x - \|\lambda\| \|x - k\| \le \lambda x - \lambda(x - k) = \lambda k,$$

hence equality must hold throughout. In particular, $\|\lambda\| \|x - k\| = \lambda(x - k)$, i.e., $\lambda || x - k$.

□

H.P.(10) Prove: *A convex set in a nls is closed if and only if it is weakly closed.*

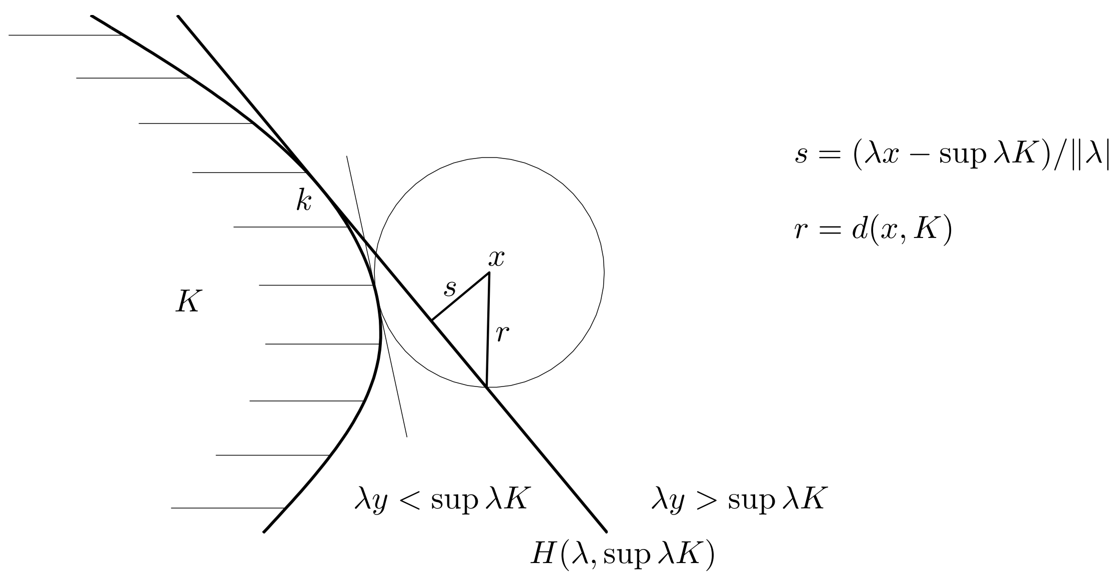
H.P.(11) Draw $K - L \subset \mathbb{R}^2$, for $K = B_2^-(-2, 0)$, $L = B_1(1, 0)$.

Application: Characterization of best approximation from a convex set

(15) Characterization Theorem. *X nls, K convex subset, $x \in X\backslash K^-$, $k \in K$. Then k is a ba to x from $K \iff \exists\{\lambda || x - k\}\ \sup \lambda K \le \lambda k$.*

Proof: '$\implies$' by (14)Corollary.

'$\impliedby$': The assumption $\lambda || x - k$ implies that, for any y with $\lambda y \le \lambda k$, $\|\lambda\| \|x - k\| = \lambda(x - k) \le \lambda(x - y) \le \|\lambda\| \|x - y\|$, i.e., that k is even a ba to x from the halfspace $\{y \in X : \lambda y \le \lambda k\}$ which, by assumption, contains K. □



(16) Figure. Characterization of best approximation: If $k \in K$ and $\lambda k = \sup \lambda K$, then $s \le r$, with equality iff k is a ba to x from K.

(16)Figure illustrates the

(17) Duality formula for distance from convex set. *For x not in the convex set K,*

$$d(x, K) = \max_{\lambda \in X^*} (\lambda x - \sup \lambda K)/\|\lambda\|.$$

Indeed, for any $\lambda \in X^*$, $\lambda x - \sup \lambda K = \inf \lambda(x - K) \le \inf \|\lambda\| \|x - K\| = \|\lambda\| d(x, K)$, while, as in the proof for (13)Corollary of Separation Theorem, $\exists\{\lambda \in X^* \backslash 0\}$ $\sup \lambda K \le \lambda x - \|\lambda\| d(x, K)$.

H.P.(12) Prove: *If Z is a lss of the nls X, L a finite-dimensional lss of X^* such that the linear map $L \to Z^* : \lambda \to \lambda\big|_Z$ is bounded below, then, for every $x \in X$, $d(x, Z \cap (x + L_\perp)) \le (1 + \beta) d(x, Z)$, with* $\beta := \sup_{\lambda \in L} \|\lambda\| / \|\lambda\big|_Z\| < \infty$. In other words, in the situation described, we wouldn't spoil the approximation error of $z \in Z$ too much if we also imposed the interpolation constraints that $z = x$ 'on' L. (Hint: you may take for granted that $(Z \cap L_\perp)^\perp = Z^\perp + L$.)

**** support fl ****

The functional $h : S_{X^*} \to \mathbb{R} : \lambda \mapsto \sup \lambda K$ is called the **support functional** for the convex set K. By the Separation Theorem, $K^- = \cap_\lambda \{y : \lambda y \le h(\lambda)\}$, *as you should verify.* This shows that every closed convex set is the intersection of **half spaces**.

A constructive instance of Hahn-Banach (of practical interest)

(18) HB Theorem for C(T). $X = C(T)$, *T compact metric, Y lss of X,* $\dim Y = n \Longrightarrow \forall\{\lambda \in Y^*\}\ \exists\{U \subseteq T$ *with* $\#U \le n, a \in \mathbb{R}^U\}$ $\lambda = (\sum_u a(u)\delta_u)|_Y$ *and* $\|\lambda\| = \|a\|_1$.

In words: Every continuous linear functional on an n-dimensional lss of $C(T)$ has a norm preserving extension to all of $C(T)$ that is also a linear combination of no more than

n point evaluations. This result is of practical interest since point values constitute the most readily available information about a function.

Proof: It is sufficient to prove that

$$(19)\qquad B^-_{Y^*} = \operatorname{conv} E, \qquad \text{with } E := \{\pm\delta^Y_t : t \in T\}, \quad \lambda^Y := \lambda|_Y.$$

For, if we know this, then we know that $\forall\{\lambda \in Y^*\backslash 0\}\ \lambda/\|\lambda\| \in S_{Y^*} = \partial B^-_{Y^*} = \partial \operatorname{conv} E$, hence, by Caratheodory, $\lambda/\|\lambda\|$ is the convex combination of $\dim Y^* = n$ point evaluations or their negative. This means that $\lambda/\|\lambda\| = \sum_U w(u)\alpha_u \delta^Y_u = \sum_U b(u)\delta^Y_u$ for some $U \subseteq T$ with $\#U \le n$, $\alpha_u = \pm 1$, and some $b \in \mathbb{R}^U$ with $\|b\|_1 = 1$. Consequently, $\lambda = (\sum_U a(u)\delta_u)|_Y$ with $a := b\|\lambda\|$, hence with $\|a\|_1 = \|\lambda\|$.

It remains to prove (19), and here, only the inequality $B^-_{Y^*} \subseteq \operatorname{conv} E$ needs proof (since $\|\delta^Y_t\| \le \|\delta_t\| = 1$, hence E is contained in the convex set $B^-_{Y^*}$, hence so is $\operatorname{conv} E$). For it, let $\lambda \notin \operatorname{conv} E$ and assume for the moment that $\operatorname{conv} E$ is closed. Then, by (13)Corollary to Separation Theorem, $\exists\{f \in Y^{**} = Y\}\ \lambda f > \sup f(\operatorname{conv} E) \ge \sup |f(T)| = \|f\|$, hence $\|\lambda\| > 1$, i.e., $\lambda \notin B^-_{Y^*}$. This proves that $B^-_{Y^*} \subseteq \operatorname{conv} E$.

Thus, to finish the proof, we need to show that $\operatorname{conv} E$ is closed. We prove this by showing that $\operatorname{conv} E$ *is compact*: Suppose (λ_r) is in $\operatorname{conv} E$. By Caratheodory, $\lambda_r = \sum_1^{n+1} a_r(i)\delta^Y_{u_{i,r}}$ and $\sum |a_r(i)| = 1$. Since the sequences $(u_{i,r})_r$ are in the compact set T, and the sequence (a_r) lies in the compact set $B^-_{\ell_1(n+1)}$, and there are just finitely many sequences involved, **AGTASMAT** (:= after going to a subsequence may assume that) $\lim a_r = a,\ \lim u_{i,r} = u_i,\ i = 1,\ldots,n+1$. This implies that $\lambda := \sum a(i)\delta^Y_{u_i} \in \operatorname{conv} E$. Further, λ is the pointwise limit of the (sub)sequence of (λ_r), hence the norm limit since this pointwise convergence is necessarily uniform on the unit ball of the *finite*-dimensional ls Y (cf. H.P.(V.3)). □

H.P.(13) How would the HB Theorem for $C(T)$ have to be changed to be valid in case $\mathbb{F} = \mathbb{C}$? (Hint: Any ls X over $\mathbb{C}$ is also a ls over $\mathbb{R}$, but, as a ls over $\mathbb{R}$, its dimension doubles since now, e.g., $[x, ix]$ is 1-1 for any $x \ne 0$.)

** application to quadrature **

(20) Proposition. *$X = C(T), T$ compact in $\mathbb{R}^m$, $\lambda : f \mapsto \int_T f(t)\,\mathrm{d}t$, Y lss of X, $n := \dim Y < \infty$. Then: $1 \in Y \Longrightarrow \exists\{U \subseteq T, w \in \mathbb{R}^U_+\}\ \#U \le n$ and*

$$\lambda|_Y = (\sum_{u\in U} w(u)\delta_u)|_Y.$$

In words: Regardless of what the finite dimensional linear space Y may be, if it contains the constant function 1, then it is possible to construct a quadrature rule (based on point values alone) that is exact on Y and uses only positive weights. This last fact is important for stability and convergence of a sequence of such rules.

Proof: $|\lambda f| = |\int_T f(t)\,\mathrm{d}t| \le \|1\|_1\|f\|$ with equality if $f = 1$. Since $1 \in Y$, we have $\|1\|_1 = \|\lambda\| = \|\lambda|_Y\|$. By (18)HB for $C(T)$, $\exists\{U \subseteq T,\ w \in \mathbb{R}^U\}\ \#U \le n$ and $\lambda_U := \sum_U w(u)\delta_u$ is a norm preserving extension of $\lambda|_Y$ to all of X. Thus $\lambda_U = \lambda$ on Y and $\|w\|_1 = \|\lambda_U\| = \|\lambda|_Y\| = \lambda 1 = \sum_U w(u)$, i.e., $\sum |w(u)| = \sum w(u)$, and this implies that $w(u) \ge 0$, all $u \in U$. □

H.P.(14) Show the existence of such a quadrature rule (perhaps using as many as $n+1$ points) even if Y does not contain the function 1.

Remark. If Y contains a nontrivial positive function f, then one gets the existence of $U \subseteq T, w \in \mathbb{R}_+^U$ with $\#U \le n := \dim Y$ and $\sum_U w(u)\delta_u = \lambda := \int \cdot$ on Y as follows: Let $Z := \ker \lambda|_Y$. Then $0 \in \operatorname{conv}\{(\delta_t)|_Z : t \in T\}$ (since we could otherwise strictly separate 0 from it, i.e., with $Z^{**} = Z$, $\exists\{z \in Z\backslash 0\}\ 0 = z^t 0 > \sup z^t(T)$, hence $\int z \ne 0$, a contradiction), hence, by Caratheodory, $\lambda|_Z = 0 = \sum_U c(u)(\delta_u)|_Z$ for some c, U with $c(u) > 0$, $\#U \le \dim Z + 1 = n$. Now choose the coefficient α so that $\alpha \sum_U c(u)\delta_u f = \int f$; then $\alpha > 0$ since $f \ge 0$.

Characterization of ba from a lss

If K is a *linear subspace* of X, then $\lambda k \ge \sup \lambda K$ implies that $\lambda \perp K$ (i.e., $K \subseteq \ker \lambda$) since $\lambda k' \ne 0$ for some $k' \in K$ implies that $\sup \lambda K \ge \sup\{\alpha\lambda k' : \alpha \in \mathbb{R}\} = \infty$. This gives the

(21) Characterization Theorem. *Y lss of nls $X, x \in X\backslash\, Y^-$, $y \in Y$. Then y is a ba to x from $Y \iff \exists\{\lambda \| x - y\}\ \lambda \perp Y$.*

**** special case ****

One uses these characterization theorems in conjunction with representation theorems for the elements of X^* to milk the equality $\lambda(x-y) = \|\lambda\|\|x-y\|$ for information useful for constructing a ba.

The classical (and motivating) *example* is m-dimensional Euclidean space, $X = \ell_2(m)$, for which $X^* \simeq X$, with each $\lambda \in X^*$ representable uniquely as scalar product $x \mapsto {x_\lambda}^t x$ with some element x_λ of X. By Hölder's inequality, $\lambda \| x - y$ implies that λ is a scalar multiple of $x - y$, hence the requirement $\lambda \perp Y$ is, in effect, the requirement that the error $x - y$ be orthogonal to Y.

It is a testimony to the power of functional analysis that the seemingly quite different characterization (in terms of alternations of the error) of a best approximation from some finite dimensional lss of $C(T)$ also derives directly from (21)Theorem, as follows.

Consider $X = C(T)$, with T compact metric, and $n := \dim Y < \infty$. A ba y to $x \in X$ is characterized by the fact that, for some $\lambda \in X^*\backslash 0$, $\lambda \| y - x$ while $\lambda \perp Y$. But, since X is quite specific, we can be more specific about λ: The two requirements on λ depend only on the action of λ on $Y_1 := \operatorname{ran}[x] + Y$. This means that any norm preserving extension μ of $\lambda|_{Y_1}$ to a lfl on all of X also satisfies $\mu \| y - x$ and $\mu \perp Y$. In particular, the extension of the form $\sum_U a(u)\delta_u$ supplied by (18)HB for $C(T)$ must satisfy these conditions. Conclusion:

(22) Characterization of ba from n-dim.lss of C(T). *Y lss of $X := C(T)$, T compact metric, $n := \dim Y < \infty$, $x \in X\backslash Y$, $y \in Y$. Then y is ba to x from $Y \iff \exists\{\lambda := \sum_U a(u)\delta_u$ with $U \subseteq T$, $\#U \le n+1$, $a \in \mathbb{R}^U\backslash 0\}\ \lambda \| x - y$ and $\lambda \perp Y$.*

One would make use of this characterization theorem in the following way: (See H.P.(IV.7)) Since $\lambda \| x - y$, we have

$$\|\lambda\|\|x-y\| = \sum a(u)(x-y)(u) \le \sum |a(u)||(x-y)(u)| \le \|a\|_1 \max_{u\in U} |(x-y)(u)| \le \|\lambda\|\|x-y\|.$$

Hence, equality must hold throughout. Assuming that none of the u's is wasting its (and our) time here, i.e., assuming that $\forall\{u \in U\}\ a(u) \neq 0$, we conclude that

$$(x-y)(u) = (\operatorname{signum} a(u))\|x-y\|, \quad \text{all } u. \tag{23}$$

This says that the points u must all be extreme points of the error function $x-y$, i.e., points at which the error $x-y$ takes on its sup norm and that, moreover, u must be a minimum (maximum) in case $a(u)$ is negative (positive).

**** special special case ****

Consider the specific choice $X = C([a\,..\,b])$, $Y = \Pi_{<n} :=$ polynomials of degree $< n$. Since the LIP($\Pi_{<n}, \operatorname{ran}[\delta_u : u \in U]$) is correct for any n-set U, a nontrivial $\lambda \in \operatorname{ran}[\delta_u : u \in U] \cap \Pi_{<n}^{\perp}$ requires that $\#U > n$. Since (22)Theorem requires $\#U \le n+1$, we must have $\#U = n+1$ and the λ is unique, up to multiplication by a scalar.

In fact, we even know an explicit form for λ. For, we know from Chapter I that, for any $x \in \mathbb{F}^{[a..b]}$, the divided difference $\delta_U x$ of x at the points in U provides the coefficient of $()^n$ in the power form of the unique polynomial $P_n x \in \Pi_n$ that agrees with x at U. Since, by uniqueness, $P_n x = x$ for any $x \in \Pi_n$, it follows that $\delta_U \perp \Pi_{<n}$. Further, since we know how to write $P_n x$ in Lagrange form:

$$P_n x = \sum_{u\in U} x(u) \prod_{v\neq u} \frac{\cdot - v}{u-v},$$

we know that $\delta_U x = \sum_{u\in U} x(u)/\prod_{v\neq u}(u-v)$. Hence, if $u_0 < \cdots < u_n$ are the points in U ordered, then

$$\delta_U = \sum_{j=0}^{n} \delta_{u_j} a(j)$$

with $a(j)a(j+1) < 0$ for all j. Thus, (22)Theorem in conjunction with (23), gives the famous

(24) Chebyshev Alternation Theorem. *$y \in \Pi_{<n}$ is a ba to $x \in C([a\,..\,b])$ iff $\exists\{u_0 < \cdots < u_n$ in $[a\,..\,b]$, $\sigma \in \{-1,1\}\}$* s.t. $(x-y)(u_j) = (-)^j \sigma \|x-y\|_\infty$, $j = 0, \ldots, n$.

H.P.(15) Why doesn't the Alternation Theorem hold for an arbitrary n-dimensional lss of $C([a\,..\,b])$?

H.P.(16) Prove Haar's result: If the n-dimensional lss Y of $C(T)$ (with T compact metric) is not a **Haar space**, i.e., has $Y \to \mathbb{R}^n : f \mapsto f\big|_U$ fail to be onto for some n-set U in T, then some $f \in C(T)$ has several ba's from Y. (Hint: Show that, for such a U, there is $\lambda = \sum_{u\in U} w(u)\delta_u \perp Y$ of norm 1 and $f \in C(T)$ of norm 1 parallel to λ. Then show that, for any $y \in Y$ of norm ≤ 1 and vanishing on U, $\lambda \parallel (1-|y|)f - \alpha y$ for any $\alpha \in [0\,..\,1]$.)

These considerations also help in the construction of such a ba, by the **Remez algorithm**, as follows. Suppose we have in hand an approximation $y \in \Pi_{<n}$ to x that, while perhaps not a ba, at least has an error $e := x - y$ that changes sign at least n times. This means that we can find $u_0 < \cdots < u_n$ so that $e(u_j)e(u_{j+1}) < 0$ for all j. Consequently,

$$\|\delta_U\| d(x, \Pi_{<n}) \ge \|\delta_U\| d(x, \ker \delta_U) = |\delta_U x| = |\delta_U e| = \sum_j |e(u_j)||a(j)| \ge \min_j |e(u_j)| \|\delta_U\|, \tag{25}$$

hence

$$\|e\|_\infty \geq d(x, \Pi_{<n}) \geq \min_j |e(u_j)|.$$

By choosing the u_j to be the sites of local extrema for e, we can make the lower bound as large as possible for the given e, and, by comparing it with the (computable) $\|e\|_\infty$, can gauge how good the error in our current approximation y is as compared with the best possible error.

If we are not satisfied, we can guarantee strict improvement of the lower bound by constructing a new approximation, z, as the ba to x from $\Pi_{<n}$ *with respect to the norm* $\|f\|_U := \max_{u\in U} |f(u)|$. Indeed, on the space $C(U)$, δ_U is, up to scalar multiples, the only lfl perpendicular to $\Pi_{<n}$, hence, by (22), any ba z to x with respect to this norm must take on its maximum error (in absolute value) at every $u \in U$ and with alternating sign. This means that we can compute (the coefficients of) z and the number $d = \pm\|x - z\|_U$ from the *linear* system $x(u_j) - z(u_j) = (-1)^j d$, $j = 0, \ldots, n$. It also means that $\|\delta_U\|\|x-z\|_U = \|\delta_U\| d_U(x, \Pi_{<n}) = |\delta_U x| = \sum_j |(x-y)(u_j)||a(j)|$, the last equality from (25), hence, for any j, $|(x-z)(u_j)| = \|x - z\|_U \geq \min_j |(x-y)(u_j)|$ with equality only if $|(x-y)(u_j)| = \text{const}$, hence (by the uniqueness of polynomial interpolation) $z = y$. In particular, the absolutely smallest of the $(n + 1)$ local extrema in the error $x - z$ 'near' those of $x - y$ is guaranteed to be at least $\|x - z\|_U$, hence strictly greater than that for $x - y$. This strict monotone increase implies the convergence of the Remez algorithm.

Convex functionals

The fl f defined on some subset of the ls X is called **convex** in case

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y), \quad \text{all } \alpha \in [0\,..\,1], \quad x, y \in \operatorname{dom} f.$$

For *example*, every sublinear fl is convex. In particular, if $A \in bL(X,Y)$, then $x \mapsto \|Ax\|$ is a convex fl. Note that the domain of a convex fl is necessarily convex, and that f is convex iff its **epigraph**

$$\operatorname{epi} f := \{(x,t) \in X \times \mathbb{R} : f(x) \leq t\}$$

is convex.

H.P.(17) An alternative but nonstandard definition would state that a fl is 'convex' if it carries convex sets to convex sets. What is the relationship (if any) between being convex and being 'convex'? Is the inverse of a convex map 'convex'?

(26) Proposition. *If $K = \operatorname{conv} E$ for some finite set E, and f is convex, then*

$$\sup f(K) = \max f(E).$$

Proof: The convexity of f implies that

$$f\Big(\sum_E a(e)e\Big) \leq \sum_E a(e)f(e) \leq \Big(\sum_E a(e)\Big) \max f(E) = \max f(E)$$

in case $a \in [0\,..\,1]^E$ with $\sum_E a(e) = 1$. Therefore, $\sup f(K) = \sup\{f\big(\sum_E a(e)e\big) : a \in [0\,..\,1]^E,\ \sum_E a(e) = 1\} \leq \max f(E) \leq \sup f(K)$. □

The argument generalizes. *If f is convex and $K = \operatorname{conv} E$, then* $\sup f(K) = \sup f(E)$. Thus, in maximizing a convex fl over a convex set K, it pays to come up with as small a set E as possible for which $\operatorname{conv} E = K$. How small can one make E? At the very least, E must contain all extreme points of K. These are all the points $x \in K$ for which $K\backslash x$ is still convex, i.e., that are not *proper* convex combinations of other points in K. More conventionally, $x \in K$ is an **extreme point** of the convex set K iff

$$x \in [y \,..\, z] \subseteq K \Longrightarrow x \in \{y, z\}.$$

Note that an extreme point is necessarily a boundary point since any interior point sits in the middle of some ball entirely contained in K, hence is the midpoint of many intervals of positive length in K. A closed convex set for which the converse holds, i.e., for which every boundary point is an extreme point is called **strictly convex**. Correspondingly, a norm is called **strictly convex** if its closed unit ball is strictly convex.

H.P.(18) Prove: *A nls X is strictly convex if and only if, for any $x, y \in X$, $\|x\| = \|(x+y)/2\| = \|y\|$ implies $x = y$.*

H.P.(19) Prove that, for $1 < p < \infty$, the closed unit ball in ℓ_p is strictly convex, hence has infinitely many extreme points. (Hint: Use equality in Hölder's inequality.) Conclude that it is usually not possible to find the (exact) norm of a lm on this space.

The closed unit ball in $\ell_1(m)$ or $\ell_\infty(m)$ is not strictly convex since it is the convex hull of finitely many points. Specifically, (as you should *verify*)

$$B_X^- = \begin{cases} \operatorname{conv}(\pm e_i : i = 1, \ldots, m) & \text{if } X = \ell_1(m); \\ \operatorname{conv}\{\sum_i e_i \varepsilon(i) : \varepsilon \in \{-1, 1\}^m\} & \text{if } X = \ell_\infty(m). \end{cases}$$

Correspondingly, we can actually compute the norm of any lm on these spaces, since this only involves maximization over the *finite* set of extreme points.

H.P.(20) Use Hölder's inequality and duality to verify this claim, i.e., to prove the formulas for $\|A\|_1$, $\|A\|_\infty$ given in (III.12)Example.

An extreme *example* of a closed convex set that is not strictly convex is provided by the closed unit ball in $X := \mathbf{L}_1[0 \,..\, 1]$, for it has *no* extreme points: If $f \in B^-\backslash 0$, then $\exists\{t \in [0 \,..\, 1]\}\ \int_0^t |f(s)|\, \mathrm{d}s = (1/2)\|f\|_1$. Choose

$$g(s) := \begin{cases} 2f(s), & s < t \\ 0, & s > t \end{cases}, \qquad h(s) := \begin{cases} 0, & s < t \\ 2f(s), & s > t \end{cases}.$$

Then $g \neq f \neq h$, yet $f = (g+h)/2$, i.e., $f \in [g \,..\, h]$ while $\|g\| = \|h\| = \|f\| \leq 1$.

H.P.(21) Prove that the closed unit ball in c_0 has no extreme points.

Remark. The closed unit ball of any dual space is compact in the topology of pointwise convergence (by (IV.4) Alaoglu's Theorem), hence has extreme points (by the Krein-Milman Theorem below). The two examples prove that neither $\mathbf{L}_1$ nor c_0 are (or, more precisely, can be identified with) the dual of some nls.

** Krein-Milman **

(27) Krein-Milman Theorem. *Any weakly compact non-empty set in the nls X is contained in the closed convex hull of its extreme points.*

Proof: Let K be a nonempty w-compact subset of X. The basic observation is the following. If $x, y \in X$, $\lambda \in X'$ and

$$z \in (x \mathinner{.\,.} y) = \{(1-\alpha)x + \alpha y : 0 < \alpha < 1\}, \tag{28}$$

then $\lambda z \le \max(\lambda x, \lambda y)$ with equality iff $\lambda x = \lambda y$. Hence, for any $\lambda \in X^*\backslash 0$, the set

$$E := K_\lambda := \{z \in K : \lambda z = \sup \lambda(K)\}$$

has the following properties: it is w-closed (as the intersection of w-closed sets) and in K, hence w-compact, and is an **extreme set for** K in the sense that $\{\} \ne E \subset K$ and, for all $x, y \in K$, $(x \mathinner{.\,.} y) \cap E \ne \{\}$ implies that $x, y \in E$ (since, if $z \in (x \mathinner{.\,.} y) \cap E$, then $\sup \lambda(K) = \lambda z \le \max(\lambda x, \lambda y) \le \sup \lambda(K)$, hence $\lambda x = \lambda y = \sup \lambda(K)$). This shows that the collection $\mathbb{E}$ of all weakly compact extreme sets for K is not empty. More than that, for any $E \in \mathbb{E}$ and any $\lambda \in X^*\backslash 0$, $E_\lambda \in \mathbb{E}$ since the same argument shows E_λ to be an extreme set for E, hence also for K (since E is an extreme set for K).

It follows that every $\lambda \in X^*$ is necessarily constant on any *minimal* $E \in \mathbb{E}$, hence any minimal element of $\mathbb{E}$ consists of just one point, and that makes that sole point an extreme point of K.

The next claim to prove is that every $E \in \mathbb{E}$ contains an extreme point of K. For this, pick $E \in \mathbb{E}$. Then $\{E' \in \mathbb{E} : E' \subseteq E\}$ is partially ordered by inclusion, hence contains a maximal totally ordered subset, $\mathbb{E}_E$. Since $\mathbb{E}_E$ is a collection of w-closed subsets of the w-compact set E having the finite intersection property, its intersection, F, is nonempty, hence in $\mathbb{E}$ (since all the other properties that characterize a set as belonging to $\mathbb{E}$ are closed under arbitrary intersections). On the other hand, by the maximality of $\mathbb{E}_E$, no proper subset of F lies in $\mathbb{E}$, i.e., F is minimal, hence consists of just one point, necessarily an extreme point for K.

It follows that the collection $\operatorname{extr}(K)$ of extreme points of K intersects every extreme set for K, hence so does $C := \operatorname{conv}(\operatorname{extr}(K))$. In particular, if there were $k \in K\backslash C^-$, then, by (13)Corollary to Separation Theorem, we could find $\lambda \in X^*$ with $\sup \lambda(C^-) < \lambda k \le \sup \lambda(K) = \lambda(K_\lambda)$, and that would be a contradiction since K is in $\mathbb{E}$, hence so is K_λ, hence the latter has points in C. □

Remark. The proof made use only of the following particulars. (i) X is a ls that is also a ts with respect to a translation-invariant topology (i.e., $\forall\{x \in X\}\ \mathbf{B}(x) = \mathbf{B}(0) + x$). (ii) The collection X^c of lfl's on X continuous in this topology is rich enough to **separate points**, i.e., to give $X^c\perp = \{0\}$. Examples of such a setup include any nls in the norm topology, but also (as used explicitly in the version of the theorem stated above) in the weak topology (recall from H.P.(10) that any weakly closed convex set is norm-closed), and even every continuous dual of a nls in the w^*-topology. It is this last setup that provides proof of the earlier claim that the closed unit ball of any X^* has extreme points; in fact, it equals the closed convex hull of its extreme points.

Unfinished business: $\operatorname{ran} A^*$ **closed** $\implies$ $\operatorname{ran} A^* = (\ker A)^\perp$

(29) Proposition. *X, Y nls, $A \in bL(X, Y)$. If $\operatorname{ran} A^*$ is closed, then $\operatorname{ran} A^* = (\ker A)^\perp$.*

Proof: We may assume that X, Y are complete, since we can always complete them, and any bounded lm on X or Y (into some Bs) has a unique extension by continuity to the completion. In particular, their duals are (essentially) unchanged.

With this assumption, it is sufficient to prove that

$$\operatorname{ran} A^* \text{ closed } \implies \operatorname{ran} A \text{ closed} \tag{30}$$

since then (V.17)Corollary does the rest.

To prove (30), we set $Z := (\operatorname{ran} A)^-$, and show that

$$C : X \to Z : x \mapsto Ax$$

is onto (hence $\operatorname{ran} A = \operatorname{ran} C = Z = (\operatorname{ran} A)^-$) by showing that C is open. For this, since $Y^* \to Z^* : \lambda \mapsto \lambda|_Z$ is onto by HB, $\operatorname{ran} C^* = \operatorname{ran} A^*$, i.e., $\operatorname{ran} C^*$ is closed while $\ker C^* = (\operatorname{ran} C)^\perp = \{0\}$ (since $(\operatorname{ran} C)^- = Z$), hence C^* is bounded below, by H.P.(V.20), i.e., $r := \inf \|\lambda C\|/\|\lambda\| > 0$. This implies that

$$B_r \subseteq (CB)^- \tag{31}$$

since $x \notin (CB)^-$ implies by (13)Corollary that, for some $\lambda \in Z^*$,

$$r\|\lambda\| \le \|\lambda C\| = \sup \lambda(CB) < \lambda x \le \|\lambda\|\|x\|,$$

i.e., $x \notin B_r^-$. From (31) and (V.13)Lemma, we conclude that C is open, hence onto, i.e., $\operatorname{ran} A = \operatorname{ran} C = Z = (\operatorname{ran} A)^-$. $\square$