EL3370 Mathematical Methods in Signals, Systems and Control

Topic 7: Dual Spaces

Cristian R. Rojas

Division of Decision and Control Systems
KTH Royal Institute of Technology

Outline

Outline

Linear Functionals

Definition

In a real vector space V, $f : V \to \mathbb{R}$ is a *linear functional* if for all $x, y \in V$, and $\alpha, \beta \in \mathbb{R}$,

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y).$$

Examples

1. In $\mathbb{R}^n$, every linear functional is of the form $f(x) = \sum_{k=1}^{n} \alpha_k x_k$, where $x = (x_1, \dots, x_n)$.
2. In $C[0,1]$, $f(x) = x(1/2)$ is a linear functional.
3. In $L_2[a,b]$, $f(x) = \displaystyle\int_a^b y(t)x(t)dt$, where $y \in L_2[a,b]$, is a linear functional.

The *kernel* of f, Ker f, is defined as Ker $f := \{x \in V : f(x) = 0\}$.

Linear Functionals (cont.)

Definition. A linear functional f in a normed space is *bounded* if there is an $M > 0$ s.t. $|f(x)| \leqslant M\|x\|$ for all $x \in V$. The smallest such M is the *norm* of f, $\|f\|$.

Theorem. Let f be a linear functional on a normed space V. The following are equivalent:

(1) f is continuous.

(2) f is continuous at 0.

(3) f is bounded.

Proof

(1) $\Rightarrow$ (2): By definition.

(2) $\Rightarrow$ (3): Let f be continuous at 0. There is a $\delta > 0$ s.t. $|f(x)| < 1$ for all $x \in V$ s.t. $\|x\| \leqslant \delta$. Then, by linearity, $|f(x)| \leqslant (1/\delta)\|x\|$ for all $x \in V$.

(3) $\Rightarrow$ (1): Pick some $x \in V$. If f is bounded, with norm M, and (x_n) is a sequence convergent to x, then $|f(x_n) - f(x)| = |f(x_n - x)| \leqslant M\|x_n - x\| \to 0$. Since x was arbitrary, f is thus continuous. $\square$

Linear Functionals (cont.)

Example
In ℓ_0 (the space of sequences with finitely non-zero entries), with norm equal to the maximum of the absolute value of its components, define for $x = (x_1, \ldots, x_n, 0, 0, \ldots)$, $f(x) = \sum_{k=1}^{n} k x_k$. This linear functional is unbounded (*why?*).

Linear functionals on V form a vector space, called *algebraic dual* of V, by defining

$$(f_1 + f_2)(x) := f_1(x) + f_2(x),$$
$$(\lambda f)(x) := \lambda f(x), \qquad x \in V,\ \lambda \in \mathbb{R}.$$

Definition

The space of bounded linear functionals on a normed space V is the *normed dual* of V, denoted as V^*. The norm of an $f \in V^*$ is $\|f\| = \sup_{\|x\| \leqslant 1} |f(x)|$.

Theorem. V^* is a Banach space.

Proof (for real normed spaces)

We only need to show that V^* is complete. Take a Cauchy sequence (x_n^*) in V^*. For every $x \in V$, $(x_n^*(x))$ is Cauchy in $\mathbb{R}$, since $|x_n^*(x) - x_m^*(x)| \leqslant \|x_n^* - x_m^*\| \|x\|$, so there is an $x^*(x) \in \mathbb{R}$ s.t. $x_n^*(x) \to x^*(x)$. The functional x^* defined in this way is linear, because $x^*(\alpha x + \beta y) = \lim x_n^*(\alpha x + \beta y) = \alpha \lim x_n^*(x) + \beta \lim x_n^*(y) = \alpha x^*(x) + \beta x^*(y)$, for all $x, y \in V$ and $\alpha, \beta \in \mathbb{R}$. In addition, for a given $\varepsilon > 0$, since (x_n^*) is Cauchy, there is an $N \in \mathbb{N}$ s.t. $\|x_n^* - x_m^*\| < \varepsilon$ for all $n, m \geqslant N$, or $|x_n^*(x) - x_m^*(x)| < \varepsilon \|x\|$ for all $x \in V$, hence taking $n \to \infty$ gives $|x^*(x) - x_m^*(x)| < \varepsilon \|x\|$. Thus, $|x^*(x)| \leqslant |x^*(x) - x_m^*(x)| + |x_m^*(x)| < \left(\varepsilon + \|x_m^*\|\right) \|x\|$, so x^* is bounded, *i.e.*, $x^* \in V^*$. Finally, as $|x^*(x) - x_m^*(x)| < \varepsilon \|x\|$, we have $x_m^* \to x^*$. Therefore, V^* is complete. $\square$

Outline

Duals of Some Common Banach Spaces

- $\mathbb{R}^n$ (with Euclidean norm): itself!

  Take the linear functional

  $$f(x) = \sum_{k=1}^{n} \alpha_k x_k, \qquad x = (x_1, \ldots, x_n).$$

  From Cauchy-Schwarz, $|f(x)| = \left|\sum_{k=1}^{n} \alpha_k x_k\right| \leqslant \sqrt{\sum_{k=1}^{n} \alpha_k^2}\,\|x\|$, so every such f is bounded, and since equality is achieved for $x = (\alpha_1, \ldots, \alpha_n)$, we have $\|f\| = \sqrt{\sum_{k=1}^{n} \alpha_k^2}$.

  Conversely, if $f \in (\mathbb{R}^n)^*$, take the standard basis vectors e_k. Every $x = (x_1, \ldots, x_n)$ can be written as $x = \sum_{k=1}^{n} x_k e_k$, so $f(x) = \sum_{k=1}^{n} x_k f(e_k) = \sum_{k=1}^{n} \alpha_k x_k$, with $\alpha_k = f(e_k)$.

  Hence the dual space of $\mathbb{R}^n$ is itself, in the sense that all bounded functionals are of the form $f(x) = \sum_{k=1}^{n} \alpha_k x_k$, with $\|f\| = \sqrt{\sum_{k=1}^{n} \alpha_k^2}$.

Duals of Some Common Banach Spaces (cont.)

- ℓ_p $(1 \leqslant p < \infty)$: ℓ_q!

  The dual of ℓ_p is ℓ_q, where $1/p + 1/q = 1$, in the sense that every bounded linear functional in ℓ_p can be written as $f(x) = \sum_{k=1}^{\infty} \alpha_k x_k$, with $\alpha = (\alpha_k) \in \ell_q$, and $\|f\| = \|\alpha\|_q$.

  The dual of ℓ_∞ is *not* ℓ_1: *see bonus slides for proof.*

- $L_p[a,b]$ $(1 \leqslant p < \infty)$: $L_q[a,b]$!

  Similarly to ℓ_p, the dual of $L_p[a,b]$ is $L_q[a,b]$, with $1/p + 1/q = 1$, since every bounded linear functional is of the form $f(x) = \displaystyle\int_a^b x(t)y(t)dt$, with $y \in L_q[a,b]$, and $\|f\| = \|y\|_q$.

Duals of Some Common Banach Spaces (cont.)

- **Dual of a Hilbert space** H
  A particular linear functional in H is $f(x) = (x, y)$ for $y \in H$. By Cauchy-Schwarz, $|f(x)| \leqslant \|x\| \|y\|$, and taking $x = y$ shows that $\|f\| = \|y\|$. Conversely,

  Theorem (Riesz-Fréchet). If $f \in H^*$, there is a unique $y \in H$ s.t. $f(x) = (x, y)$ for all $x \in H$, and $\|f\| = \|y\|$.

  Proof. If $f = 0$, the theorem is trivial. Otherwise, the set $N = \{x \in H : f(x) = 0\}$ is closed (*why?*) and not equal to H. Since $H = N \oplus N^\perp$, take a $z \in N^\perp \setminus \{0\}$, scaled so that $f(z) = 1$. We will show that y is a multiple of z. Given $x \in H$, we have $x - f(x)z \in N$, since $f(x - f(x)z) = 0$. As $z \perp N$, we have $(x - f(x)z, z) = 0$, or $(x, z) = f(x)(z, z)$, hence $y = z/\|z\|^2$. If $(x, y_1) = (x, y_2)$ for all $x \in H$ then $y_1 = y_2$, which proves the uniqueness of y. □

Outline

Hahn-Banach Theorem (Extension of Linear Functionals)

Definitions

1. Let f be a linear functional defined on a subspace M of a vector space V. An *extension* F of f (from M to N) is a linear functional defined on $N \supsetneq M$, s.t. $F\big|_M = f$.
2. A real valued function p defined on a vector space V is a *sublinear functional* if
   (a) $p(x_1 + x_2) \leqslant p(x_1) + p(x_2)$, for all $x_1, x_2 \in V$, (*sub-additivity*)
   (b) $p(\alpha x) = \alpha p(x)$, for all $\alpha \geqslant 0$ and $x \in V$. (*positive homogeneity*)

Theorem (Hahn-Banach; extension form)
Let V be a real vector space, and p a sublinear functional on V. Let f be a linear functional defined on a subspace $M \subseteq V$ satisfying $f(x) \leqslant p(x)$ for all $x \in M$. Then there is an extension F of f from M to V s.t. $F(x) \leqslant p(x)$ for all $x \in V$.

Proof

Let us define a partial order $\preceq$ on the set E of extensions of f satisfying the conditions of the theorem: say that $F \preceq G$ if G is an extension of F. For every chain $C \subseteq E$, according to this partial order, define $\hat{F}$ as $\hat{F}(x) = F(x)$ if $x \in \mathscr{D}(F)$ for some $F \in C$. $\hat{F}$ is a linear functional with domain $\bigcup_{F \in C} \mathscr{D}(F)$, and it is an upper bound for the chain C. Thus, by Zorn's lemma, E has a maximal element, say, $\bar{F}$.

The linear functional $\bar{F}$ should have domain equal to V, since otherwise there is an element $y \in V \setminus \mathscr{D}(\bar{F})$. All $x \in \operatorname{lin}(\mathscr{D}(\bar{F}) + y)$ are of the form $x = m + \alpha y$ with $m \in \mathscr{D}(\bar{F})$ and $\alpha \in \mathbb{R}$. An extension of $\bar{F}$ from $\mathscr{D}(\bar{F})$ to $\operatorname{lin}(\mathscr{D}(\bar{F}) + y)$ has the form $g(x) = \bar{F}(m) + \alpha g(y)$, so we need to show that $g(y)$ can be chosen s.t. $g(x) \leqslant p(x)$ on $\operatorname{lin}(\mathscr{D}(\bar{F}) + y)$, or, equivalently, that $\bar{F}(m) + \alpha g(y) \leqslant p(m + \alpha y)$ for all $m \in \mathscr{D}(\bar{F})$ and $\alpha \in \mathbb{R}$.

By positive homogeneity, this is equivalent to requiring that $\bar{F}(m) + g(y) \leqslant p(m + y)$ and $\bar{F}(m) - g(y) \leqslant p(m - y)$ for all $m \in \mathscr{D}(\bar{F})$, or that $\bar{F}(m) - p(m - y) \leqslant g(y) \leqslant p(m + y) - \bar{F}(m)$. Therefore, for the existence of an extension of $\bar{F}$, we need to show that $\bar{F}(m) - p(m - y) \leqslant p(m' + y) - \bar{F}(m')$ for all $m, m' \in \mathscr{D}(\bar{F})$, or $\bar{F}(m) + \bar{F}(m') \leqslant p(m - y) + p(m' + y)$. This condition holds, due to the sub-additivity of p (and that $m + m' \in \mathscr{D}(\bar{F})$): $\bar{F}(m) + \bar{F}(m') = \bar{F}(m + m') \leqslant p(m + m') = p(m - y + y + m') \leqslant p(m' + y) + p(m - y)$.

Therefore, we can choose $g(y)$ so that $g(x) \leqslant p(x)$ on $\operatorname{lin}(\mathscr{D}(\bar{F}) + y)$, thus contradicting the maximality of $\bar{F}$. This contradiction shows that $\bar{F}$ is the sought extension of f to V. □

Remark

It is not possible to avoid Zorn's lemma (or the axiom of choice), since it can be shown that the Hahn-Banach theorem is equivalent to the axiom of choice.

Hahn-Banach Theorem (Extension of Linear Functionals) (cont.)

Corollary 1

Let f be a bounded linear functional defined on a subspace M of a normed space V. Then there is an extension F of f to V s.t. $\|F\| = \|f\|$.

Proof. Take $p(x) = \|F\|\|x\|$ in the Hahn-Banach theorem. □

Corollary 2

Let $x \in V$. Then there is a nonzero $F \in V^*$ on a normed space V s.t. $F(x) = \|F\|\|x\|$.

Proof. Assume $x \neq 0$ (otherwise, take any $F \in V^*$). On $\text{lin}\{x\}$, define the bounded functional $f(\alpha x) = \alpha\|x\|$, which has norm 1. By Corollary 1, extend f to an $F \in V^*$ with norm 1. □

The Hahn-Banach theorem, particularly as Corollary 1, can be used as an existence theorem for minimization problems.

Outline

The Dual of $C[a,b]$

Definitions

- $x\colon [a,b] \to \mathbb{R}$ is of *bounded variation* if there is a $K > 0$ s.t. for every *partition* of $[a,b]$ (*i.e.*, a finite set $\{t_1,\dots,t_n\}$ s.t. $a = t_0 < t_1 < \cdots < t_n = b$), $\sum_{k=1}^n |x(t_k) - x(t_{k-1})| \leqslant K$.
- $\mathrm{BV}[a,b]$ is the normed space of all $x\colon [a,b] \to \mathbb{R}$ of bounded variation, with norm $\|x\| = |x(a)| + \mathrm{TV}(x)$, with $\mathrm{TV}(x) := \sup \sum_{k=1}^n |x(t_k) - x(t_{k-1})|$, where the sup is over all partitions of $[a,b]$.

A consequence of Hahn-Banach is that the dual of $C[a,b]$ is (almost) $\mathrm{BV}[a,b]$:

Theorem (Riesz Representation Theorem)
Let $f \in C[a,b]^*$. Then there is $v \in \mathrm{BV}[a,b]$ s.t.

$$f(x) = \int_a^b x(t)\,dv(t), \qquad x \in C[a,b] \qquad \text{(“Riemann-Stieltjes integral”)}$$

and $\|f\| = \mathrm{TV}(v)$. Conversely, every $v \in \mathrm{BV}[a,b]$ defines an $f \in C[a,b]^*$ in this way.

Digression: Riemann-Stieltjes Integral

A *partition* of the interval $[a,b]$ is a collection P of numbers $x_0,\dots,x_n$ s.t. $a = x_0 < x_1 < \cdots < x_n = b$, and denoted as

$$P = \{a = x_0 < x_1 < \cdots < x_n = b\}.$$

The *norm* of P is $\text{norm}(P) = \max_{i=0,\dots,n-1} |x_{i+1} - x_i|$.

Consider a function $f\colon [a,b] \to \mathbb{R}$, and $g \in \text{BV}[a,b]$. The *Riemann-Stieltjes integral of f w.r.t. g* is defined as

$$\int_a^b f(x)\,dg(x) := \lim_{\Delta\to 0} \sup_{\substack{P=\{a=x_0<x_1<\cdots<x_n=b\} \\ \text{norm}(P)\leqslant\Delta}} \sup_{\substack{c_i\in[x_i,x_{i+1}] \\ (i=0,\dots,n-1)}} \sum_{i=0}^{n-1} f(c_i)[g(x_{i+1}) - g(x_i)],$$

in case the limit exists. Here the first supremum is over all partitions of $[a,b]$ (of any size n) of norm at most Δ.

Proof of Riesz representation theorem

Given an $f \in C[a,b]^*$, we need to find a $\nu \in \mathrm{BV}[a,b]$ s.t. $f(x) = \int_a^b x(t) d\nu(t)$.
A natural idea is to set $x = u_s$ ($s \in [a,b]$), where $u_s(t) = 1$ for $t \in [a,s]$ and 0 otherwise, so

$$f(u_s) = \int_a^b u_s(t) d\nu(t) = \int_a^s d\nu(t) = \nu(s) - \nu(a),$$

hence we can set $\nu(a) := 0$ and $\nu(s) := f(u_s)$.

Unfortunately, $u_s \notin C[a,b]$. However, u_s belongs to the space B of bounded functions on $[a,b]$, and $C[a,b] \subseteq B$, so by Hahn-Banach it is possible to obtain an extension F of f from $C[a,b]$ to B, preserving its norm. This will allow us to define $\nu(s) := F(u_s)$!

The rest of the proof is to show that:

- $\nu \in \mathrm{BV}[a,b]$
- $f(x) = \int_a^b x(t) d\nu(t)$
- $\|f\| = \mathrm{TV}(\nu)$ (*see bonus slides for details*)

□

The Dual of $C[a, b]$ (cont.)

Riesz Representation Theorem does not provide a unique $v \in \mathrm{BV}[a, b]$, since, *e.g.*, $f(x) = x(1/2)$ can be described by

$$v(t) = \begin{cases} 0, & 0 \leqslant t < 1/2, \\ \alpha, & t = 1/2, \\ 1, & 1/2 < t \leqslant 1, \end{cases}$$

where α can be any number between 0 and 1 (*exercise!*). To avoid this, we define

Definition
$\mathrm{NBV}[a, b] \subseteq \mathrm{BV}[a, b]$, *normalized space of functions of bounded variation*, consists of those $v \in \mathrm{BV}[a, b]$ s.t. $v(a) = 0$ and are right continuous on (a, b). Here $\|v\| := \mathrm{TV}(v)$.

Then, $\mathrm{NBV}[a, b]$ is the dual of $C[a, b]$.

Outline

Let $x \in V^*$. For every $x \in V$, denote $\langle x, x^* \rangle := x^*(x)$ (resembling an inner product).
This notation also allows to define a functional f on V^* (for a given $x \in V$):

$$f(x^*) = \langle x, x^* \rangle, \qquad x^* \in V^*.$$

This f is linear, and since $|f(x^*)| = |\langle x, x^* \rangle| \leqslant \|x\| \|x^*\|$, so $\|f\| \leqslant \|x\|$, and by Hahn-Banach, there is an $x^* \in V^*$ s.t. $|\langle x, x^* \rangle| = \|x\| \|x^*\|$, so $\|f\| = \|x\|$.

Definition
$V^{**} := (V^*)^*$ is the *second dual* of a normed space V.

As we saw, there is a *natural mapping* $\varphi \colon V \to V^{**}$ given by $\varphi(x) = x^{**}$, where $\langle x^*, x^{**} \rangle = \langle x, x^* \rangle$. This mapping is linear and *norm-preserving*: $\|\varphi(x)\| = \|x\|$. However, φ is not always surjective, *i.e.*, there may be an $x^{**} \in V^{**}$ not represented by elements of V.

Definition. A normed space V is *reflexive* if φ is surjective. If so, we write $V = V^{**}$.

Second Dual Space (cont.)

Examples

1. ℓ_p and L_p $(1 < p < \infty)$ are reflexive, since $(\ell_p)^{**} = (\ell_q)^* = \ell_p$, where $1/p + 1/q = 1$.

2. ℓ_1 and L_1 are *not* reflexive.

3. Every Hilbert space is reflexive.

Outline

Alignment and Orthogonal Complements

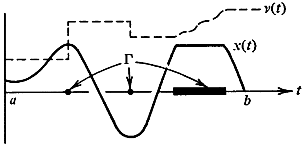
For $x \in V$ and $x^* \in V^*$, $\langle x, x^* \rangle \leqslant \|x\| \|x^*\|$. In Hilbert spaces, we have equality iff x^* is represented by a nonnegative multiple of x, *i.e.*, $\langle x, x^* \rangle = (x, \alpha x)$ for some $\alpha \geqslant 0$. In a normed space V, we define

Definition. $x^* \in V^*$ is *aligned* with $x \in V$ if $\langle x, x^* \rangle = \|x\| \|x^*\|$.

Example

Let $x \in C[a,b]$, and $\Gamma := \{t \in [a,b] \colon |x(t)| = \|x\|\} \neq \emptyset$.

An $x^*(x) = \displaystyle\int_a^b x(t) d\nu(t)$ is aligned with x iff ν only varies on Γ so that ν is nondecreasing at t if $x(t) > 0$ and nonincreasing if $x(t) < 0$ (*exercise!*).



Alignment and Orthogonal Complements (cont.)

Definition. $x \in V$ and $x^* \in V^*$ are *orthogonal* if $\langle x, x^* \rangle = 0$.

In Hilbert spaces, this coincides with the original definition of orthogonality.

Definition. The *orthogonal complement* of $S \subseteq V$, $S^\perp$, is the set of all $x^* \in V^*$ s.t. $\langle x, x^* \rangle = 0$ for all $x \in S$.

For subsets of V^*, we have a "dual" definition:

Definition. The *orthogonal complement* of $U \subseteq V^*$, ${}^\perp U$, is the set of all $x \in V$ s.t. $\langle x, x^* \rangle = 0$ for all $x^* \in U$.

Notice that $U^\perp \neq {}^\perp U$ ($U \subseteq V^*$) in general, unless V is reflexive, since $U^\perp \subseteq V^{**}$, while ${}^\perp U \subseteq V$.

Theorem. If M is a closed subspace of V, then ${}^\perp[M^\perp] = M$. (*See proof at the end*.)

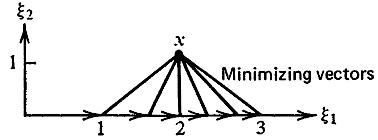
Outline

Goal: Extend the projection theorem to minimum norm problems in normed spaces.

The situation is more difficult: possibly several optima, given by nonlinear equations, ...

Example

Let $V = \mathbb{R}^2$ with norm $\|(y_1, y_2)\| = \max\{\|y_1\|, \|y_2\|\}$, and let $M = \mathbb{R} \times \{0\}$. Then $\min_{y \in M} \|(2,1) - y\| = 1$, but the optimum is not unique!



Theorem

Let $x \in V$, where V is a real normed space, and let d be the distance from x to a subspace M. Then,

$$d = \inf_{m \in M} \|x - m\| = \max_{\substack{\|x^*\| \leqslant 1 \\ x^* \in M^\perp}} \langle x, x^* \rangle,$$

where the maximum is achieved for some $x_0^* \in M^\perp$. Also, if the infimum is achieved for some $m_0 \in M$, then x_0^* is aligned with $x - m_0$.

Proof

($\geqslant$) Take $\varepsilon > 0$ and $m_\varepsilon \in M$ s.t. $\|x - m_\varepsilon\| \leqslant d + \varepsilon$. Then, for every $x^* \in M^\perp$, $\|x^*\| \leqslant 1$:

$$\langle x, x^* \rangle = \langle x - m_\varepsilon, x^* \rangle \leqslant \|x - m_\varepsilon\| \|x^*\| \leqslant d + \varepsilon,$$

and since ε was arbitrary, we have $\langle x, x^* \rangle \leqslant d$ for all $x^* \in M^\perp$, $\|x^*\| \leqslant 1$.

Proof

($\leqslant$) Let $N = \text{lin}(x+M)$. If $n \in N$, then $n = \alpha x + m$ for some $m \in M$, $\alpha \in \mathbb{R}$. Define the functional f on N as $f(n) = \alpha d$. Then

$$\|f\| = \sup_{n\in N} \frac{|f(n)|}{\|n\|} = \sup_{m\in M} \frac{|\alpha| d}{\|\alpha x + m\|} = \sup_{m\in M} \frac{d}{\|x + m/\alpha\|} = \frac{d}{\inf_{m\in M} \|x+m/\alpha\|} = 1.$$

By Hahn-Banach, let x_0^* be a norm-preserving extension of f to V. Then $\|x_0^*\| = 1$ and $x_0^* = f$ on N. By construction, $x_0^* \in M^\perp$ and $\langle x, x_0^* \rangle = d$.

Now, let $m_0 \in M$ be s.t. $\|x - m_0\| = d$, and take $x_0^* \in M^\perp$ s.t. $\|x_0^*\| = 1$ and $\langle x, x_0^*\rangle = d$. Then, $\langle x - m_0, x_0^* \rangle = \langle x, x_0^* \rangle = d = \|x - m_0\| \|x_0^*\|$, so x_0^* is aligned with $x - m_0$. $\square$

Minimum Norm Problems (cont.)

Corollary (generalized projection theorem)
Let $x \in V$ and M be a subspace of V. Then $m_0 \in M$ satisfies $\|x - m_0\| \leqslant \|x - m\|$ for all $m \in M$ iff there is an $x_0^* \in M^\perp$ aligned with $x - m_0$.

The following is a "dual" version of the previous theorem:

Theorem
Let M be a subspace of V, and $x^* \in V^*$, which is at a distance d from $M^\perp$. Then

$$d = \min_{m^* \in M^\perp} \|x^* - m^*\| = \sup_{\substack{x \in x \\ \|x\| \leqslant 1}} \langle x, x^* \rangle,$$

where the minimum is achieved for $m_0^* \in M^\perp$. If the infimum is achieved for some $x \in M$, then $x^* - m_0^*$ is aligned with x_0.

Example (control problem, again)
Consider the DC motor problem again, where we want to find the current $u : [0,1] \to \mathbb{R}$ to drive a motor governed by

$$\ddot{\theta}(t) + \dot{\theta}(t) = u(t)$$

from $\theta(0) = \dot{\theta}(0) = 0$ to $\theta(1) = 1$, $\dot{\theta}(1) = 0$, so as to minimize $\max_{t \in [0,1]} |u(t)|$.

We can formulate this problem in a dual space, by forcing $u \in (L_1[0,1])^* = L_\infty[0,1]$, so that we seek an input of minimum L_∞ norm.

As in the previous example, the constraints are

$$\int_0^1 e^{t-1} u(t) dt = \langle y_1, u \rangle = 0, \qquad y_1(t) = e^{t-1},$$
$$\int_0^1 (1 - e^{t-1}) u(t) dt = \langle y_2, u \rangle = 1, \qquad y_2(t) = 1 - e^{t-1}.$$

Example (cont.)

Therefore, the optimization problem is

$$L = \min_{\substack{u \in L_\infty \\ \langle y_1, u \rangle = 0 \\ \langle y_2, u \rangle = 1}} \|u\| = \min_{\substack{\tilde{u} \in L_\infty \\ \langle y_1, \tilde{u} \rangle = 0 \\ \langle y_2, \tilde{u} \rangle = 0}} \|\tilde{u} - \bar{u}\| = \min_{\tilde{u} \in M^\perp} \|\tilde{u} - \bar{u}\|,$$

where $\bar{u}$ is any function in $L_\infty[0,1]$ satisfying the constraints $\langle y_1, \bar{u} \rangle = 0$ and $\langle y_2, \bar{u} \rangle = 1$, and $M = \text{lin}\{y_1, y_2\}$.

From the theorems:

$$L = \min_{\substack{u \in L_\infty \\ \langle y_1, u \rangle = 0 \\ \langle y_2, u \rangle = 1}} \|u\| = \sup_{\substack{x \in M \\ \|x\| \leqslant 1}} \langle x, \bar{u} \rangle = \sup_{\|ay_1 + by_2\| \leqslant 1} \langle ay_1 + by_2, \bar{u} \rangle = \sup_{\|ay_1 + by_2\| \leqslant 1} b$$

and the constraint $\|ay_1 + by_2\| \leqslant 1$, in L_1, corresponds to $\displaystyle\int_0^1 \left| (a-b)e^{t-1} + b \right| dt \leqslant 1$.

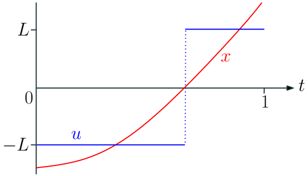
Example (cont.)
This means that the optimal value is equal to

$$\sup_{\int_0^1 |(a-b)e^{t-1}+b| dt \leqslant 1} b,$$

which is attained and can be solved explicitly, since it is simple and finite dimensional!

Qualitatively, the optimal $x(t) = (a-b)e^{t-1} + b$ is aligned with the optimal u, *i.e.*, $\langle x, u \rangle = \|x\| \|u\|$.

Since the right side is the maximum of the left side over all $u \in L_\infty$ with $\|u\| = L$, has to take only values $\pm L$, depending on $\mathrm{sgn}(x(t))$: "Bang-bang" control!

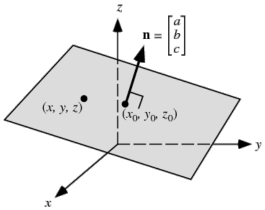


As x is the sum of an exponential and a constant, it can change sign at most once, so u has to change sign at most once.

Outline

Definition. A *hyperplane* H in a vector space V is a maximal proper linear variety, *i.e.*, a set of the form $x + M$ ($x \in V$ and M is a subspace of V) s.t. $H \neq V$ and for every $y \in V \setminus H$, $\mathrm{lin}(H + y) = V$.

Hyperplanes are the "contour surfaces" of linear functionals.



Theorem. H is a hyperplane in V iff there is a linear functional $f \neq 0$ in V and $c \in \mathbb{R}$ s.t. $H = \{x \in V : f(x) = c\}$. Also, if $0 \notin H$, there is a unique f s.t. $H = \{x \in V : f(x) = 1\}$.

Proof.

($\Rightarrow$) Pick a hyperplane $H = x_0 + M$ ($x_0 \in V$ and M is a subspace). If $x_0 \notin M$, then $\mathrm{lin}(x_0 + M) = V$, so every $x \in V$ fulfills $x = \alpha x_0 + m$ ($\alpha \in \mathbb{R}$, $m \in M$); letting $f(x) = \alpha$ yields $H = \{x \in V : f(x) = 1\}$. If $x_0 \in M$, let $x_1 \in V \setminus H$; every $x \in V$ satisfies $x = \alpha x_1 + m$, so $f(x) = \alpha$ gives $H = \{x \in V : f(x) = 0\}$.

($\Leftarrow$) If $f \neq 0$, define the subspace $M = \{x \in V : f(x) = 0\}$. Take an $x_0 \in V$ with $f(x_0) = 1$; then, for every $x \in V$, $f(x - f(x)x_0) = 0$, so $x = f(x)x_0 + m$ for some $m \in M$, *i.e.*, M is maximally proper. Now, for $c \in \mathbb{R}$, let $x_1 \in X$ s.t. $f(x_1) = c$; then, $\{x : f(x) = c\} = \{x : f(x - x_1) = 0\} = x_1 + M$, which is a hyperplane If $0 \notin H$, one can scale f by $c \neq 0$ so that $H = \{x \in V : f(x) = 1\}$. If also $H = \{x \in V : g(x) = 1\}$, then $H \subseteq \{x \in V : f(x) - g(x) = 0\}$, but the smallest subspace containing H is V, so $f = g$. $\square$

By definition, a hyperplane H is either closed or dense in V, since $H \subseteq \overline{H} \subseteq V$, and $\overline{H}$ is also a linear variety (*why?*), so either $\overline{H} = H$ or $\overline{H} = V$, because H is maximally proper (*i.e.*, it cannot be inside a larger but proper linear variety).

Theorem. Let $f \neq 0$ be a linear functional on V. Then, for every c, $H = \{x : f(x) = c\}$ is closed iff f is continuous.

Proof. If f is continuous, then $H = f^{-1}(\{c\})$ is closed since $\{c\}$ is closed for every $c \in \mathbb{R}$. Conversely, as $H = \{x : f(x) = c\}$ is a hyperplane, there is a $y \in V \setminus H$, and every $x \in V$ fulfills $x = \alpha y + h$, with $\alpha \in \mathbb{R}$ and $h \in H$. Then, consider a sequence (x_n) in V, with $x_n = \alpha_n y + h_n$ ($\alpha_n \in \mathbb{R}$, $h_n \in H$), s.t. $x_n \to x = \alpha y + h$ ($\alpha \in \mathbb{R}$, $h \in H$). Since H is closed, $d = \inf_{h \in H} \|y - h\| > 0$ (*from Homework 1*), so $\|x_n - x\| = \|(\alpha_n - \alpha)y + h_n - h\| \geqslant |\alpha_n - \alpha| d$, which implies that $\alpha_n \to \alpha$, hence $f(x_n) = \alpha_n f(y) \to \alpha f(y) = f(x)$, so f is continuous. □

See bonus slides for an example of a linear functional for which H is dense.

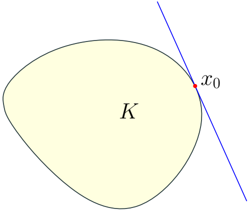
Hahn-Banach Theorem (Geometric Form) (cont.)

A hyperplane $H = \{x\colon f(x) = c\}$ generates the *half-spaces*

$$\{x\colon f(x) \leqslant c\}, \quad \{x\colon f(x) \geqslant c\}, \quad \{x\colon f(x) < c\}, \quad \{x\colon f(x) > c\}.$$

If f is continuous, the first 2 are closed and the last two are open.

Goal: establish a geometric Hahn-Banach theorem:
Given a convex set K s.t. $\operatorname{int} K \neq \emptyset$, and $x_0 \notin \operatorname{int} K$, there is a closed hyperplane containing x_0 but disjoint from $\operatorname{int} K$.



If K were $\overline{B(0,1)}$, the theorem is easy, since Hahn-Banach assures the existence of an x_0^* aligned with x_0: if $x \in \operatorname{int} \overline{B(0,1)}$,

$$\langle x, x_0^* \rangle \leqslant \|x\| \|x_0^*\| < \|x_0\| \|x_0^*\| = \langle x_0, x_0^* \rangle, \qquad \text{so } H = \{x\colon \langle x, x_0^* \rangle = \langle x_0, x_0^* \rangle\} \text{ is enough.}$$

We thus need to generalize the unit ball $\overline{B(0,1)}$ to arbitrary convex sets!

Hahn-Banach Theorem (Geometric Form) (cont.)

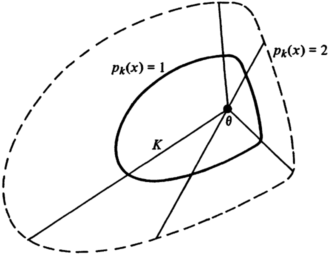
Definition (Minkowski functional)

Let K be a convex set in a normed space V, and assume $0 \in \operatorname{int} K$.
The *Minkowski functional* $p\colon V \to \mathbb{R}_0^+$ of K on V is

$$p(x) := \inf\{r > 0 : x \in rK\}.$$

Properties (assuming $0 \in$ int K)

(1) $0 \leqslant p(x) < \infty$ for all $x \in V$.
(2) $p(\alpha x) = \alpha p(x)$ for $\alpha > 0$.
(3) $p(x_1 + x_2) \leqslant p(x_1) + p(x_2)$.
(4) p is continuous.
(5) $\overline{K} = \{x : p(x) \leqslant 1\}$, $\operatorname{int} K = \{x : p(x) < 1\}$.



p is then a continuous, sublinear functional.

(*See bonus slides for proofs of these properties*.)

Hahn-Banach Theorem (Geometric Form) (cont.)

Theorem (Mazur's Theorem, Geometric Hahn-Banach)
Let K be a convex set in a real normed space V s.t. $\operatorname{int} K \neq \emptyset$. Let W be a linear variety s.t. $W \cap \operatorname{int} K = \emptyset$. Then there is a closed hyperplane containing W but no interior points of K, *i.e.*, there is an $x^* \in V^* \setminus \{0\}$ and $c \in \mathbb{R}$ s.t. $\langle w, x^* \rangle = c$ for $w \in W$, $\langle k, x^* \rangle \leqslant c$ for $k \in K$, and $\langle k, x^* \rangle < c$ for $k \in \operatorname{int} K$.

Proof. *Idea: use the Minkowski functional p in the Hahn-Banach theorem!*
Assume w.l.o.g. that $0 \in \operatorname{int} K$. Let $M = \operatorname{lin} W$. Then, W is a hyperplane in M which does not contain 0, so there is a linear functional f on M s.t. $W = \{x : f(x) = 1\}$.

Let p be the Minkowski functional of K. Since W has no interior points of K, $f(x) = 1 \leqslant p(x)$ for $x \in W$. Every point in M can be written as αx, where $\alpha \in \mathbb{R}$ and $x \in W$, so if $\alpha > 0$, $f(\alpha x) = \alpha \leqslant p(\alpha x)$ (by positive homogeneity), and if $\alpha < 0$, $f(\alpha x) \leqslant 0 \leqslant p(\alpha x)$, so $f(x) \leqslant p(x)$ in M.

By Hahn-Banach, there is an extension F of f to V s.t. $F(x) \leqslant p(x)$. Let $H = \{x : F(x) = 1\}$. Since $F(x) \leqslant p(x)$ on V, and p is continuous, so is F (because $-p(-x_n) \leqslant F(x_n) \leqslant p(x_n)$, so if $x_n \to 0$, then $F(x_n) \to 0$), and $F(x) < 1$ on $\operatorname{int} K$, while $F(x) \leqslant 1$ on K. Therefore, H is the desired hyperplane. $\square$

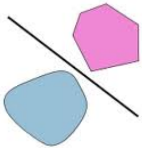
See bonus slides for example of why int $K \neq \emptyset$ is needed in Mazur's theorem.

Some corollaries

1. (*Supporting Hyperplane Theorem*). If $x \notin \operatorname{int} K \neq \emptyset$, there is a closed hyperplane H containing x s.t. K lies on one side of H (If $x \in \partial K$, K is a *supporting hyperplane*).

   Proof. A special case of Mazur's theorem, where $W = \{x\}$. □

2. (*Eidelheit Separating Hyperplane Theorem*). If K_1 and K_2 are convex, s.t. $\operatorname{int} K_1 \neq \emptyset$ and $K_2 \cap \operatorname{int} K_1 = \emptyset$. Then there is a closed hyperplane separating K_1 and K_2 (*i.e.*, an $x^* \in V^*$ s.t. $\sup_{x \in K_1} \langle x, x^* \rangle \leqslant \inf_{x \in K_2} \langle x, x^* \rangle$).

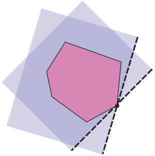
   

   Proof. Let $K = K_1 + (-K_2) = \{x_1 - x_2 : x_1 \in K_1 \text{ and } x_2 \in K_2\}$. $\operatorname{int} K \neq \emptyset$, and $0 \notin \operatorname{int} K$, so apply Corollary 1 with $x = 0$. □

3. If K is closed and convex, and $x \notin K$, there is a closed halfspace that contains K but not x.

   Proof. Let $d = \inf_{y \in K} \|x - y\| > 0$, so $B(x, d/2)$ does not intersect K. Apply Corollary 2 with $K_1 = K$ and $K_2 = \overline{B(x, d/2)}$. □

4. (*Dual formulation of convex sets*). A closed convex set is equal to the intersection of all the closed half-spaces that contain it.

   Proof. Follows directly from Corollary 3. □

Next Topic

Linear Operators

Outline

Bonus: Proof that ℓ_1 is Not the Dual of ℓ_∞

Let c_0 be the subspace of ℓ_∞ consisting of sequences $x = (x_n)$ s.t. $x_n \to 0$. We will first show that $c_0^* = \ell_1$. First, let $z \in \ell_1$ and define $\varphi_z : c_0 \to \mathbb{R}$ so that $\varphi_z(x) = \sum_{n=1}^\infty z_n x_n$. φ_z is linear, and bounded, since $|\varphi_z(x)| = \left|\sum_{n=1}^\infty z_n x_n\right| \leqslant \|z\|_1 \|x\|_\infty$, and if $x_n = \operatorname{sgn}(z_n)$, then $|\varphi_z(x)| = \|z\|_1 \|x\|_\infty$, so $\|\varphi_z\| = \|z\|_1$. Conversely, if $f : c_0 \to \mathbb{R}$ is a bounded linear functional, define $z = (z_n)$ by $z_n = f(e_n)$; then, if $v = \sum_{n=1}^N \operatorname{sgn}(z_n) e_n$, $\sum_{n=1}^N |z_n| = |f(v)| \leqslant \|f\| \|v\|_\infty = \|f\|$, so taking $N \to \infty$ we conclude that $z \in \ell_1$. Furthermore, for every $x \in c_0$, $f(x) = f\left(\sum_{n=1}^\infty x_n e_n\right) = f\left(\lim_{N\to\infty} \sum_{n=1}^N x_n e_n\right) = \lim_{N\to\infty} \sum_{n=1}^N x_n f(e_n)$ $= \sum_{n=1}^\infty z_n x_n = \varphi_z(x)$ (by continuity of f), so there is a one-to-one correspondence between c_0^* and ℓ_1.

Now we show that there are elements of ℓ_∞^* which are not of the form φ_z with $z \in \ell_1$. Since $c_0 \subseteq \ell_\infty$, and $x \in \ell_\infty$, where $x_n = 1$ for all n, does not belong to c_0, Mazur's theorem with $K = \overline{B(x, 1/2)}$ and $W = c_0$ implies the existence of a $f \in \ell_\infty^*$ s.t. $f(y) = 0$ for all $y \in c_0$, and $f(x) = 1$; this f cannot be written in the form φ_z, since otherwise $z = 0$ and $f = 0$, which is not true. $\square$

Bonus: Details of Derivation of Dual of $C[a,b]$

(1) $v \in \mathrm{BV}[a,b]$: Let $a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$, and $\varepsilon_k = \mathrm{sgn}[v(t_k) - v(t_{k-1})]$. Then,

$$\sum_{k=1}^{n} |v(t_k) - v(t_{k-1})| = \sum_{k=1}^{n} \varepsilon_k [v(t_k) - v(t_{k-1})] = F\left(\sum_{k=1}^{n} \varepsilon_k [u_{t_k} - u_{t_{k-1}}]\right) \leqslant \|F\| \underbrace{\left\| \sum_{k=1}^{n} \varepsilon_k [u_{t_k} - u_{t_{k-1}}] \right\|}_{=1} = \|F\| = \|f\|,$$

so $v \in \mathrm{BV}[a,b]$, with $\mathrm{TV}(v) \leqslant \|f\|$.

(2) $f(x) = \int_a^b x(t) dv(t)$: For $x \in C[a,b]$, let $z = \sum_{k=1}^n x(t_{k-1})[u_{t_k} - u_{t_{k-1}}] \in B$. Then, $\|z - x\|_B = \max_k \max_{t_{k-1} \leqslant t \leqslant t_k} |x(t_{k-1}) - x(t)| \to 0$ as $\max_k |t_k - t_{k-1}| \to 0$, since x is uniformly continuous. Also, continuity of F yields $F(z) \to F(x) = f(x)$, but $F(z) = \sum_{k=1}^n x(t_{k-1})[v(t_k) - v(t_{k-1})] \to \int_a^b x(t) dv(t)$, by definition of the Riemann-Stieltjes integral, so $f(x) = \int_a^b x(t) dv(t)$.

(3) $\|f\| = \mathrm{TV}(v)$: Since $\|F(z)\| = \left\| \sum_{k=1}^n x(t_{k-1})[v(t_k) - v(t_{k-1})] \right\| \leqslant \|x\| \ \mathrm{TV}(v)$, since $F(z) \to F(x) = f(x)$, we obtain $\|f(x)\| \leqslant \|x\| \ \mathrm{TV}(v)$, *i.e.*, $\|f\| \leqslant \mathrm{TV}(v)$. Combining this with (1) yields $\|f\| = \mathrm{TV}(v)$. $\square$

Bonus: Proof that ${}^{\perp}[M^{\perp}] = M$

It is clear that $M \subseteq {}^{\perp}[M^{\perp}]$.

Let $x \notin M$. On $\operatorname{lin}(x+M)$, define the linear functional $f(\alpha x + m) = \alpha$ for all $m \in M$. Then,

$$\|f\| = \sup_{m \in M} \frac{f(x+m)}{\|x+m\|} = \frac{1}{\inf_{m \in M} \|x+m\|},$$

and since M is closed, $\|f\| < \infty$. Thus, by Hahn-Banach, we can extend f to an $x^* \in V^*$. Since $f(x) = 0$ on M, $x^* \in M^{\perp}$, but also $\langle x, x^* \rangle = 1$, so $x \notin {}^{\perp}[M^{\perp}]$. $\square$

Bonus: Example of a Dense Hyperplane

It is easy to build examples of closed hyperplanes, since these are the contour surfaces of bounded functionals. Finding dense hyperplanes is a bit harder. Here is an example:

Consider the space $c_0 \subseteq \ell_\infty$ of sequences that converge to 0, and let $(e_n)_{n\in\mathbb{N}}$ be the standard basis of these spaces. Let $v_0 = (1, 1/2, 1/3, \dots) \in c_0$. The set $\{v_0, e_1, e_2, \dots\}$ is l.i. (recall that l.i. refers to *finite* linear combinations); this set can be complemented with a set $\{b_k\}_{k\in I} \subseteq c_0$ to obtain a *Hamel basis*, *i.e.*, an l.i. set $\{v_0, e_1, e_2, \dots\} \cup \{b_k\}_{k\in I}$ s.t. every $x \in c_0$ can be written as a finite linear combination of elements of this set (*recall the application of Zorn's Lemma from Topic 1*).

Define a functional $f : c_0 \to \mathbb{R}$ by

$$f\left(\alpha_0 v_0 + \textstyle\sum_{n=1}^{\infty} \alpha_n e_n + \sum_{k\in I} \beta_k b_k\right) = \alpha_0.$$

Note that $f \neq 0$ (because $f(v_0) = 1 \neq 0$), and since $e_n \in \text{Ker } f := \{x \in c_0 : f(x) = 0\}$ for every $n \in \mathbb{N}$, Ker f is a dense hyperplane in c_0. *Exercise:* prove that f is unbounded.

Bonus: Proof of Properties of Minkowski Functional

(1) $0 \leqslant p(x) < \infty$ *for all* $x \in V$:
By definition, $p(x) \geqslant 0$. Also, since $0 \in \operatorname{int} K$, $B(0,\delta) \subseteq K$ for some $\delta > 0$, so for every $\varepsilon > 0$ and $x \in V$, $x \in B(0, \|x\| + \varepsilon\delta) = (\|x\|/\delta + \varepsilon)B(0,\delta) \subseteq (\|x\|/\delta + \varepsilon)K$, thus $p(x) \leqslant \|x\|/\delta + \varepsilon < \infty$.

(2) $p(\alpha x) = \alpha p(x)$ *for* $\alpha > 0$:
$p(\alpha x) = \inf\{r > 0 \colon \alpha x \in rK\} = \inf\{\alpha \tilde{r} > 0 \colon x \in \tilde{r}K\} = \alpha p(x)$.

(3) $p(x_1 + x_2) \leqslant p(x_1) + p(x_2)$:
Given $\varepsilon > 0$, let $r_k \in [p(x_k), p(x_k) + \varepsilon]$, so $x_k/r_k \in K$ $(k = 1,2)$. Then, $(x_1 + x_2)/(r_1 + r_2) = [r_1/(r_1 + r_2)](x_1/r_1) + [r_2/(r_1 + r_2)](x_2/r_2) \in K$, since this is a convex combination of x_1/r_1 and x_2/r_2, and K is convex. Hence, $p(x_1 + x_2) \leqslant r_1 + r_2 \leqslant p(x_1) + p(x_2) + 2\varepsilon$, and since $\varepsilon > 0$ was arbitrary, letting $\varepsilon \to 0$ shows that p is sublinear.

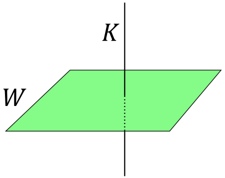
(4) *p is continuous*:
By (1), if $x_n \to 0$, $0 \leqslant p(x_n) \leqslant \|x_n\|/\delta \to 0$, so p is continuous at 0. Also, if $x_n \to x$, $p(x) \leqslant p(x - x_n) +$ $p(x_n)$ and $p(x_n) \leqslant p(x_n - x) + p(x)$, so $-p(x_n - x) \leqslant p(x) - p(x_n) \leqslant p(x - x_n)$. Thus, since $x_n - x \to 0$ and $x - x_n \to 0$, $p(x_n) \to p(x)$ and p is continuous.

(5) $\overline{K} = \{x \colon p(x) \leqslant 1\}$, $\operatorname{int} K = \{x \colon p(x) < 1\}$:
If $x \in K$, $p(x) \leqslant 1$, so if $x_n \in K \to x \in V$, by (4), $p(x) \leqslant 1$, thus $\overline{K} \subseteq \{x \colon p(x) \leqslant 1\}$, and if $x \in \operatorname{int} K$, there is an $\varepsilon > 0$ s.t. $(1 + \varepsilon)x \in K$, so $\operatorname{int} K \subseteq \{x \colon p(x) < 1\}$. Conversely, if $x \notin \overline{K}$, there is an $\varepsilon \in (0,1)$ s.t. $(1 - \delta)x \notin K$ for all $0 \leqslant \delta \leqslant \varepsilon$, hence $p(x) > 1$ and $\overline{K} = \{x \colon p(x) \leqslant 1\}$, and if $p(x) < 1$, $x \in K$, so $p^{-1}((0,1))$ is a nbd of x, thus $\operatorname{int} K = \{x \colon p(x) < 1\}$. □

Bonus: Why Do We Need int $K \neq \emptyset$ in Mazur's Theorem?

If int $K = \emptyset$, it is not possible in general to find a hyperplane H containing W s.t. K lies on one side of H.

Counter-example
In $V = \mathbb{R}^3$, let $W = \mathbb{R}^2 \times \{0\}$ and $K = \{(0,0)\} \times \mathbb{R}$, as shown in the figure below. Note that int $K = \emptyset$. While $W \cap \text{int}\, K = \emptyset$, there is only one hyperplane containing the plane W (itself: $H = W$!), and K lies on both sides of it.



Bonus: Weak* Convergence

A sequence (x_n) in a normed space V is said to (*strongly*) *converge* to $x \in V$ if $\|x - x_n\| \to 0$.

Another notion of convergence in V is: (x_n) *converges weakly* to $x \in V$ if $\langle x_n, x^* \rangle \to \langle x, x^* \rangle$ as $n \to \infty$ for all $x^* \in V^*$. If $x_n \to x$ strongly, then $\left|\langle x_n, x^* \rangle - \langle x, x^* \rangle\right| \leqslant \|x^*\| \|x_n - x\| \to 0$, so $x_n \to x$ weakly too.

Example. In ℓ_2, consider the sequence (e_n) with $e_n = (0, \dots, 0, 1, 0, \dots,$ with 1 in the n-th place. For every $x^* = (\eta_1, \eta_2, \dots)$, $\langle e_n, x^* \rangle = \eta_n \to 0$ as $n \to \infty$, so $e_n \to 0$ weakly. However, $\|e_n - 0\| = 1$ for every n, so e_n does *not* converge to 0 strongly.

In the dual space V^*, we can similarly define strong and weak convergence (since it is a normed space), but an even weaker notion is available:

Definition

(x_n^*) *converges in weak*-sense* to $x^* \in V^*$ if $\langle x, x_n^* \rangle \to \langle x, x^* \rangle$ as $n \to \infty$ for all $x \in V$.

Weak convergence in V^* implies weak* convergence, since in the former one requires that $\langle x_n^*, x^{**} \rangle \to \langle x^*, x^{**} \rangle$ for all $x^{**} \in V^{**}$, where $V^{**} \supseteq V$.

Bonus: Weak* Convergence (cont.)

Example. For $V = c_0 \subseteq \ell_\infty$, the set of sequences that converge to 0, $V^* = \ell_1$ and $V^{**} = \ell_\infty$ (see Bonus Slide 43). In $V^* = \ell_1$, $e_n \to 0$ weakly*, but not weakly, since $\langle e_n, x^{**} \rangle \not\to 0$ for $x^{**} = (1, 1, \dots) \in V^{**}$.

Weak* convergence corresponds to point-wise convergence for linear functionals, so it has the relative topology on V^* of the product topology of $\mathbb{R}^V = \prod_{x \in V} \mathbb{R}$; in particular, the weak* topology is Hausdorff but not first-countable in general. Also,

Theorem (Banach-Alaoglu). The closed unit ball in V^*, K, is weak*-compact.

Proof. Since $K = \{x^* \in V^* : |x^*(x)| \leqslant \|x\| \text{ for all } x \in V\} = V^* \cap \prod_{x \in V} f_x^{-1}([-\|x\|, \|x\|]) =: V^* \cap D$, where $f_x : V^* \to \mathbb{C}$ given by $f_x(x^*) = x^*(x)$ is weak*-continuous (see Slide 52), and D is compact by Tychonoff, it suffices to show that K is closed. If $f \in \overline{K} \cap D$, pick $x, y \in V$, $\alpha, \beta \in \mathbb{R}$, and $\varepsilon > 0$, and let $g \in K$ s.t. $|f(x) - g(x)| < \varepsilon$, $|f(y) - g(y)| < \varepsilon$ and $|f(\alpha x + \beta y) - g(\alpha x + \beta y)| < \varepsilon$. Then, $|f(\alpha x + \beta y) - \alpha f(x) - \beta f(y)| \leqslant |f(\alpha x + \beta y) - g(\alpha x + \beta y)| + |\alpha||g(x) - f(x)| + |\beta||g(y) - f(y)| < (1 + |\alpha| + |\beta|)\varepsilon$ and $|f(x)| \leqslant |f(x) - g(x)| + \|g\| < \varepsilon + 1$, and since $\varepsilon, x, y, \alpha, \beta$ were arbitrary, f is linear and $\|f\| \leqslant 1$, so $f \in K$ and K is closed. $\square$

Bonus: Weak* Convergence (cont.)

Weak*-compactness is useful for establishing existence of solutions of optimization problems involving a weak*-continuous functional on a dual space.

For example, if the variables of a convex optimization problem are linear functionals on a normed space X, and the second dual X^{**} does not have a simple characterization, one cannot write the Lagrangian dual in an explicit form (*see Topic 9!*). However, the following lemma states that one can still approximate, in a weak* sense, the dual by an optimization problem in X!

Lemma. Let X be a normed space. Then, the embedding of X into its second dual via its natural mapping, $\varphi(X)$, is weak*-dense in X^{**}.

Proof. Take a $y \in X^{**}$, $f_1, \dots, f_n \in X^*$, and some $\varepsilon > 0$. To establish the weak*-denseness of $\varphi(X)$ in X^{**}, we need to show that there is an $x \in X$ s.t.

$$|\langle f_i, y\rangle - \langle f_i, \varphi(x)\rangle| = |\langle f_i, y\rangle - \langle x, f_i\rangle| < \varepsilon, \quad \text{for all } i = 1, \dots, n.$$

We shall prove a stronger statement, namely, that $f_i(x) = y(f_i)$ for all $i = 1, \dots, n$. By a problem of Homework 4, this system of equations has a solution $x \in X$ iff for every $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, $\sum_{i=1}^n \lambda_i f_i = 0$ implies that $\sum_{i=1}^n \lambda_i y(f_i) = 0$, which holds due to the linearity of y: $\sum_{i=1}^n \lambda_i y(f_i) = y\left(\sum_{i=1}^n \lambda_i f_i\right) = y(0)$ $= 0$. This concludes the proof. $\square$

Bonus: Weak* Convergence (cont.)

The weak* topology of V^* is the weakest in which functionals $f(x^*) = \langle x, x^* \rangle$ on V^*, for some $x \in V$, are continuous (*i.e.*, every other such topology contains the open sets of the weak* topology). This is because for $|f(x^*)| = |\langle x, x^* \rangle| < \varepsilon$ to hold, the topology should contain open sets of the form $\{x^* \in V^* : |\langle x, x^* \rangle| < \varepsilon\}$ for all $x \in V$, as well as finite intersections of these sets, which generate the weak* topology. Conversely,

Theorem. Every weak*-continuous linear functional on V^* has the form $f(x^*) = \langle x, x^* \rangle$ for some $x \in V$.

Proof. Let f be a weak*-continuous linear functional on V^*. Then, for every $\varepsilon > 0$, there is a set $U = \{x^* : |\langle x_k, x^* \rangle| < \delta \text{ for } k = 1, \dots, n\}$ where $x_1, \dots, x_n \in V$, s.t. if $x^* \in U$, $|f(x^*)| < \varepsilon$. Thus, by linearity, $|f(x^*)| \leqslant (\varepsilon/\delta) \max_{k \in \{1,\dots,n\}} |\langle x_k, x^* \rangle|$, so if $\langle x_k, x^* \rangle = 0$ for all k, then $f(x^*) = 0$; assume now w.l.o.g. that $\{x_1, \dots, x_n\}$ is l.i. Let $F(x^*) = (\langle x_1, x^* \rangle, \dots, \langle x_n, x^* \rangle, f(x^*))$. Since there is a nbd of $(0, \dots, 0, 1)$ not intersecting $\mathscr{R}(F)$, by Mazur's theorem there is a $\lambda \in \mathbb{R}^{n+1} \setminus \{0\}$ s.t. $\lambda^T F(x^*) = 0$ for all x^*, *i.e.*, $\left\langle \sum_{k=1}^n \lambda_k x_k, x^* \right\rangle + \lambda_{n+1} f(x^*) = 0$ for all x^*, and since $\{x_1, \dots, x_n\}$ are l.i., $\lambda_{n+1} \neq 0$, so $f(x^*) = \left\langle \sum_{k=1}^n (-\lambda_k/\lambda_{n+1}) x_k, x^* \right\rangle$. $\square$

Let B, B^* be the closed unit balls in V, V^* respectively. The following is a converse of the Banach-Alaoglu theorem:

Theorem (Krein-Smulian). If $E \subseteq V^*$ is convex and s.t. $E \cap (rB^*)$ is weak*-compact for every $r > 0$, then E is weak*-closed.

Proof

Firstly note that E is norm-closed, since if (x_n^*) is a sequence in E that converges to $x^* \in V^*$, then (x_n^*) is bounded, *i.e.*, $\|x_n^*\| \leqslant M$ for some $M > 0$, thus $x_n^* \in E \cap (MB^*)$ for all n, so $x^* \in E \cap (MB^*) \subseteq E$. This means that if $x^* \in V^* \setminus E$, there is a ball centered at x^* which does not intersect E; thus, by translation and scaling, the theorem will be proven by showing that "*if $E \cap B^* = \emptyset$, then there exists an $x \in V$ s.t. $\langle x, x^* \rangle \geqslant 1$ for every $x^* \in E$*": since $\{x^* : \langle x, x^* \rangle \geqslant 1\}$ is a weak*-closed half-space, E is the intersection of closed half-spaces, and hence it is weak*-closed.

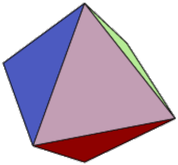
Next, given a subset $F \subseteq X$, define its *polar* $P(F) := \{x^* : |\langle x, x^* \rangle| \leqslant 1 \text{ for all } x \in F\}$; notice that the intersection of all sets $P(F)$ as F ranges over all finite subsets of $r^{-1}B$ is exactly rB^*.

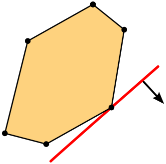
Proof (cont.)
To establish the proposition, let $F_0 = \{0\}$, and assume that finite sets $F_0, \dots, F_{n-1}$ have been chosen s.t. $F_k \subseteq k^{-1}B$ and $P(F_0) \cap \cdots \cap P(F_{n-1}) \cap E \cap nB^* = \emptyset$; note that this holds for $n = 1$. Let $Q = P(F_0) \cap \cdots \cap P(F_{n-1}) \cap E \cap (n+1)B^*$; if $P(F) \cap Q \neq \emptyset$ for every finite set $F \subseteq n^{-1}B$, the weak*-compactness of Q implies (via the finite intersection property) that $nB^* \cap Q \neq \emptyset$, which contradicts the properties of $F_0, \dots, F_{n-1}$. Hence, there is a finite set $F_n \subseteq n^{-1}B$ s.t. $P(F_n) \cap Q = \emptyset$, which yields the sequence (F_n).

By construction, (F_n) satisfies $E \cap \bigcap_{n=1}^{\infty} P(F_n) = E \cap P(\cup F_n) = \emptyset$. Now, arrange the elements of $\cup F_n$ in a sequence (x_n); note that $\|x_n\| \to 0$. Define $T \colon V^* \to c_0$ (the space of sequences converging to 0) by $Tx^* = (\langle x_n, x^* \rangle)$. $T(E)$ is a convex subset of c_0, and, due to $E \cap P(\cup F_n) = \emptyset$, $\|Tx^*\| = \sup_n |\langle x_n, x^* \rangle| \geqslant 1$ for every $x^* \in E$. Thus, since $c_0^* = \ell_1$ (*see previous bonus slides*), by the separating hyperplane theorem there is a $y \in \ell_1$ (*i.e.*, $\sum_{n=1}^{\infty} |y_n| < \infty$) s.t. $\langle Tx^*, y \rangle = \sum_{n=1}^{\infty} y_n \langle x_n, x^* \rangle \geqslant 1$ for all $x^* \in E$. The vector $x = \sum_{n=1}^{\infty} y_n x_n$ is well defined (*why?*) and it satisfies the condition $\langle x, x^* \rangle \geqslant 1$ for every $x^* \in E$. $\square$

Motivation. Consider a convex set in $\mathbb{R}^3$, such as a polytope.

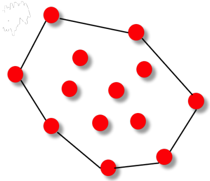
A polytope contains important subsets of points, such as *vertices*, *edges* and *faces*, which are special cases of *extreme sets*, that include an entire line segment if one of their interior points is a part of them. *Extreme points* (or *vertices*) are singleton extreme sets.





Extreme sets can be found by maximizing a linear function over the polytope: the set of maximizers is an extreme set.
Thus, one can find extreme points by iteratively maximizing linear functionals over smaller and smaller extreme sets.

Given a set of points, its *convex hull* is the set of all convex combinations of these points.
The convex hull can be found graphically (in 2D) by enclosing the points with a rubber band.
Note that a polytope is the convex hull of its vertices (or extreme points).



The Krein-Milman theorem is one of the cornerstones of functional analysis, as it extends to very general vector spaces some of the intuitive results of Euclidean geometry shown in the previous slide. To state it, we first need some definitions:

Definitions

- *Topological vector space*: a vector space V over F with a Hausdorff topology s.t. the addition $+ : V \times V \to V$ and scalar multiplication $\cdot : F \times V \to V$ operations are continuous. Every normed space induces a topological vector space.
- *Locally convex topological (LCT) vector space*: a real topological vector space V s.t. for every $x \in V$ and every nbd U of x, there is a convex nbd $W \subseteq U$ of x.
- V^* (*dual* of a topological vector space V): set of continuous linear functionals on V.
- *Extreme point* of a convex set K: a point $x \in K$ s.t. if $x = \lambda y + (1-\lambda)z$ with $y, z \in K$ and $\lambda \in [0,1]$, then either $\lambda = 0$ or $\lambda = 1$.
- $\mathrm{conv}(X)$ (*convex hull of a subset X of a real vector space*): set of all *convex combinations* of points in X (*i.e.*, expressions $\alpha_1 x_1 + \cdots + \alpha_n x_n$ where $\alpha_1, \ldots, \alpha_n \geqslant 0$, $\sum_k \alpha_k = 1$, $x_1, \ldots, x_n \in X$, and $n \in \mathbb{N}$).

Bonus: Krein-Milman and Carathéodory Theorems (cont.)

In locally convex topological vector spaces, linear functionals can "separate points":

Theorem (separation of points)
If V is an LCT vector space, V^* *separates points* in V (*i.e.*, if $x, y \in V$, $x \neq y$, there is an $f \in V^*$ s.t. $f(x) \neq f(y)$).

Proof. Since V is an LCT vector space, all properties of the Minkowski functional of a convex set hold. In particular, if U is a convex set and $0 \in \operatorname{int} U$, then p is finite, because scalar multiplication is continuous, so for every $z \in V$ there is an $\alpha \in \mathbb{R}$ s.t. $\alpha z \in U$.

Therefore, Mazur's theorem applies to LCT vector spaces, as well as its 4 corollaries. Corollary 3, in particular implies this result, by taking $K = \{y\}$. □

Corollary (separation of a convex set and a point)
If K is a closed, convex set in an LCT vector space V, and $y \in V \setminus K$, then there is an $f \in V^*$ s.t. $f(x) \leqslant c$ for all $x \in K$, and $f(y) > c$.

Theorem (Krein-Milman)
Let V be an LCT vector space, and $K \subseteq V$ a convex, compact and non-empty set. Then:

(i) K has at least one extreme point, and
(ii) K is the closure of the convex hull of its extreme points.

Proof
(i) Call a set $E \subseteq K$ *extreme* if it is convex, compact, non-empty and s.t. if $x \in E$ is a convex combination of $y, z \in K$, then both $y, z \in E$. Since K is extreme, the collection F of all extreme subsets of K is not empty. Partially order F by set inclusion; we will show that every subchain $\{E_\alpha\}$ of F has a lower bound in F, $\bigcap_\alpha E_\alpha$. $\bigcap_\alpha E_\alpha$ is clearly convex, compact (as it is the intersection of compact sets), and if $x \in \bigcap_\alpha E_\alpha$ is a convex combination of $y, z \in K$, then for every α, $x \in E_\alpha$, so both $y, z \in E_\alpha$, and hence $y, z \in \bigcap_\alpha E_\alpha$. It remains to show that $\bigcap_\alpha E_\alpha$ is non-empty: if $\bigcap_\alpha E_\alpha = \emptyset$, then $\bigcup_\alpha E_\alpha^c = K$, but since K is compact, there is a finite subcollection $\{E_{\alpha_k}^c\}_k$ of $\{E_\alpha^c\}$ whose union is K; however, since $\{E_{\alpha_k}^c\}_k$ is totally ordered, there is a largest extreme set $E_{\alpha_N}^c = K$, *i.e.*, $E_{\alpha_N} = \emptyset$, which is a contradiction.

By Zorn's lemma, K has a minimal extreme subset E. If E had at least two points, by the theorem on separation of points, there is an $f \in V^*$ that separates them.

Bonus: Krein-Milman and Carathéodory Theorems (cont.)

Proof (cont.)
Since E is compact and f is continuous and non-constant, it achieves its maximum, say, $f_{\max}$, on some proper subset M of E, which is closed (and hence compact) due to the continuity of f. M is convex, and if $x \in M$ satisfies $x = \lambda y + (1-\lambda)z$ for some $y, z \in K$ and $\lambda \in [0,1]$, then both $y, z \in E$ and furthermore $f_{\max} = f(x) = \lambda f(y) + (1-\lambda) f(z)$, so $f(y) = f(z) = f_{\max}$ and $y, z \in M$; thus, M is a smaller extreme subset of K than E, a contradiction, so E has only one point, which is an extreme point of K.

(ii) Let K_e be the set of all extreme points of K, and $C_e = \mathrm{conv}(K_e) \subseteq K$. If $x \notin \overline{C}_e$, by the corollary on the separation of a convex set and a point, there is an $f \in V^*$ s.t. $f(y) \leqslant c$ for all $y \in C_e$ and $f(x) > c$. As before, f achieves its maximum over K, $f_{\max}$, on some closed non-empty set $E \subseteq K$. If $y \in E$ satisfies $y = \lambda z + (1-\lambda)w$ for some $z, w \in K$ and $\lambda \in [0,1]$, then $f_{\max} = f(y) = \lambda f(z) + (1-\lambda) f(w)$, so $f(z) = f(w) = f_{\max}$ and $z, w \in E$, hence E is extreme. By part (i), E should contain at least one extreme point, p, which belongs to $K_e \subseteq C_e$, hence $f_{\max} = f(p) \leqslant c$. Therefore, since $f(x) > c \geqslant f_{\max}$, $x \notin K$. Thus, $\overline{C}_e \supseteq K$, which implies that $\overline{C}_e = K$. □

Example: c_0 is not the dual space of a normed space

By Banach-Alaoglu, the closed unit ball of a dual space is weak*-compact, as well as convex and non-empty, hence by Krein-Milman, it should contain at least one extreme point.

However, the closed unit ball of c_0 does not have extreme points: indeed, let $x \in c_0$, $\|x\| \leqslant 1$. By definition, $x = (x_n)$, with $x_n \to 0$, so there is an $N \in \mathbb{N}$ s.t. $|x_n| < 1/2$ for all $n \geqslant N$. Let $y^1, y^2 \in c_0$ be s.t. $y^1_n = y^2_n = x_n$ for $n \leqslant N$ and $y^1_n = x_n + 2^{-n}$, $y^2_n = x_n - 2^{-n}$ for $n > N$. Then, $\|y^1\| \leqslant 1$, $\|y^2\| \leqslant 1$, $x = (1/2)(y^1 + y^2)$, but $y^1 \neq x$, $y^2 \neq x$, so x is not an extreme point.

Bonus: Krein-Milman and Carathéodory Theorems (cont.)

To apply Krein-Milman, the following lemma is often useful:

Lemma. If V is an LCT vector space, $K \subseteq V$ a compact and convex set, and $F \subseteq K$ a compact set s.t. $\overline{\text{conv}(F)} = K$, then the extreme points of K are contained in F.

Proof. Let x be an extreme point of K not in F. Since F is compact, there is a nbd U_0 of 0 s.t. $(x + U_0) \cap F = \emptyset$, and a convex nbd U of 0 s.t. $U - U \subseteq U_0$, so $(x + U) \cap (F + U) = \emptyset$, hence $x \notin \overline{F + U}$. The family $\{y + U\}_{y \in F}$ is an open cover of F, so by compactness, $\{y_k + U\}_{k=1,\dots,n}$ is a finite subcover. Let $Q_k = \overline{\text{conv}(\overline{[y_k + U]} \cap K)} \subseteq \overline{y_k + U}$; note that Q_k is closed, and hence a compact subset of K, therefore

$$K = \overline{\text{conv}(Q_1 \cup \cdots \cup Q_n)} = \text{conv}(Q_1 \cup \cdots \cup Q_n).$$

(This result is proven by induction on n: for $n = 2$, the mapping $\varphi\colon [0,1] \times Q_1 \times Q_2 \to \text{conv}(Q_1 \cup Q_2)$ given by $\varphi(\alpha, y_1, y_2) = \alpha y_1 + (1-\alpha) y_2$ is continuous, and $[0,1] \times Q_1 \times Q_2$ is compact, hence $\varphi([0,1] \times Q_1 \times Q_2) = \text{conv}(Q_1 \cup Q_2)$ is compact, so $\overline{\text{conv}(Q_1 \cup Q_2)} = \text{conv}(Q_1 \cup Q_2)$.)

Since $x \in K$, $x = \sum_{k=1}^n \alpha_k x_k$ for $x_k \in Q_k$, $\alpha_k \geqslant 0$, $\alpha_1 + \cdots + \alpha_n = 1$. But x is an extreme point, so $x = x_k$ for some k, which implies that $x \in Q_k \subseteq y_k + \overline{U} \subseteq \overline{F + U}$, a contradiction. □

In finite dimensions, the Krein-Milman theorem can be made more explicit:

Theorem (Carathéodory) Let $X \subseteq \mathbb{R}^N$. Every point $x \in \text{conv}(X)$ can be written as a convex combination of at most $N+1$ points from X.

Proof. Let $x \in \text{conv}(X)$. Then, $x = \lambda_1 x_1 + \cdots + \lambda_n x_n$, where $\lambda_1, \dots, \lambda_n > 0$, $\lambda_1 + \cdots + \lambda_n = 1$ and $x_1, \dots, x_n \in X$. If $n > N+1$, the points $x_2 - x_1, \dots, x_n - x_1$ are l.d., *i.e.*, $\mu_2(x_2 - x_1) + \cdots + \mu_n(x_n - x_1) = 0$ for some $\mu_2, \dots, \mu_n \in \mathbb{R}$, not all zero. Defining $\mu_1 = -\mu_2 - \cdots - \mu_n$, we have that $\mu_1 x_1 + \cdots + \mu_n x_n = 0$ and $\mu_1 + \cdots + \mu_n = 0$. Since not all μ's are zero, there is at least one $\mu_k > 0$. Then,

$$x = \lambda_1 x_1 + \cdots + \lambda_n x_n - \alpha(\mu_1 x_1 + \cdots + \mu_n x_n) = (\lambda_1 - \alpha\mu_1)x_1 + \cdots + (\lambda_n - \alpha\mu_n)x_n.$$

Pick α as the minimum of λ_j/μ_j, over the indices j for which $\mu_j > 0$; note that $\alpha > 0$ and $\lambda_j - \alpha\mu_j \geqslant 0$ for all j, but $\lambda_j - \alpha\mu_j = 0$ for at least one index. Thus, $x = (\lambda_1 - \alpha\mu_1)x_1 + \cdots + (\lambda_n - \alpha\mu_n)x_n$, where each coefficient is non-negative, their sum is one, and one of them is zero. Applying this procedure iteratively one can write x as a convex combination of at most $N+1$ points. $\square$

Application: Optimal multisine spectrum

In experimental design, one searches for an input cumulative spectrum Φ^{opt} (a non-decreasing function $\Phi\colon [0,\pi] \to \mathbb{R}_0^+$ s.t. $\Phi^{\text{opt}}(0) = 0$) s.t. the *information matrix* I_F of a parameter $\theta \in \mathbb{R}^d$ is maximized in some sense, under a power constraint $\Phi(\pi) = \int_0^\pi d\Phi(\omega) = 1$. Now,

$$I_F(\Phi) \approx \int_0^\pi F(\omega) d\Phi(\omega) \in \mathbb{R}^{d\times d}, \qquad F(\omega): \text{symmetric positive semi-definite matrix.}$$

Note that, under the power constraint, $I_F(\Phi)$ lies in the convex hull of $\{F(\omega) : \omega \in [0,\pi]\}$, which has dimension $d(d+1)/2$ (since $F(\omega)$ is a *symmetric* matrix). Thus, by Carathéodory's Theorem, $I_F(\Phi^{\text{opt}})$ is a convex combination of at most $n = d(d+1)/2+1$ matrices $F(\omega_1), \dots, F(\omega_{d(d+1)/2+1})$, *i.e.*, for some $\lambda_1, \dots, \lambda_n \geqslant 0$,

$$I_F(\Phi^{\text{opt}}) = \lambda_1 F(\omega_1) + \cdots + \lambda_n F(\omega_n) = \int_0^\pi F(\omega)[\lambda_1 \delta(\omega - \omega_1) + \cdots + \lambda_n \delta(\omega - \omega_n)]\, d\omega,$$

where δ is the Dirac distribution. Therefore Φ^{opt} can be replaced by a "multisine" spectrum with frequencies at $\omega_1, \dots, \omega_n$ to obtain the same maximal information matrix!

Bonus: Positive Linear Functionals

Let X be a topological space. A linear functional F on $C(X)$ is *positive*, denoted $F \geqslant 0$, if $F(f) \geqslant 0$ whenever $f(x) \geqslant 0$ for all $x \in X$. Positive linear functionals are called *Borel measures* on X, and those of unit norm are called *Borel probability measures* on X.

Lemma (Cauchy-Schwarz inequality for positive linear functionals)

If $F: C(X) \to \mathbb{R}$ is a positive linear functional, and $f, g \in C(X)$, then $[F(fg)]^2 \leqslant F(f^2)F(g^2)$

Proof. Since $F[(\lambda f + g)^2] = \lambda^2 F(f^2) + 2\lambda F(fg) + F(g^2) \geqslant 0$ for all $\lambda \in \mathbb{R}$, the discriminant $4[F(fg)]^2 - 4F(f^2)F(g^2)$ is non-positive. $\square$

Lemma (characterization of positive linear functionals)

A linear functional $F: C(X) \to \mathbb{R}$ is positive iff $F(\mathbb{1}) = \|F\|$, where $\mathbb{1}: x \in C(X) \mapsto 1$.

Proof.

($\Rightarrow$) If $F \geqslant 0$, then $F(\mathbb{1}) \leqslant \|F\|$, since $\|\mathbb{1}\| = 1$, and, by Cauchy-Schwarz, for every $f \in C(X)$ of unit norm, $[F(f)]^2 = [F(\mathbb{1}f)]^2 \leqslant F(\mathbb{1}^2)F(f^2) \leqslant F(\mathbb{1})\|F\|$, since $\|f^2\| \leqslant \|f\|^2 \leqslant 1$; taking the supremum over all f of unit norm, $\|F\|^2 \leqslant F(\mathbb{1})\|F\|$, so in conclusion $F(\mathbb{1}) = \|F\|$.

($\Leftarrow$) If $f \in C(X)$ satisfies $f(x) \geqslant 0$ for all x, and $\|f\| \leqslant 1$, then also $\mathbb{1}(x) - f(x) \geqslant 0$, and $\|\mathbb{1} - f\| \leqslant 1$, so $\|F\| - F(f) = F(\mathbb{1}) - F(f) = F(\mathbb{1} - f) \leqslant \|F\|$, thus $F(f) \geqslant 0$. By linearity, F is positive. $\square$

Lemma (Jordan decomposition for linear functionals)
Every $l \in C(X)^*$, where X is compact, can be written as $l = l_+ - l_-$, where $l_+, l_- \in C(X)^*$ are positive, and $\|l\| = \|l_+\| + \|l_-\|$.

Proof. Let $S = \{l \in C(X)^* : l(\mathbb{1}) = \|l\| \leqslant 1\}$ and $U = \{l \in C(X)^* : \|l\| \leqslant 1\}$; by Banach-Alaoglu, S and U are weak*-compact. We will first show that U equals $K = \mathrm{conv}(S \cup (-S))$. If $l \in K$, then $l = \alpha l_1 - (1-\alpha) l_2$, with $l_1, l_2 \in S$ and $\alpha \in [0,1]$ (since S is convex, every convex combination in $S \cup (-S)$ can be reduced to this form); hence $\|l\| \leqslant \alpha \|l_1\| + (1-\alpha)\|l_2\| \leqslant 1$, so $K \subseteq U$. Furthermore, K is weak*-compact: indeed, define $\varphi \colon S \times S \times [0,1] \to U$ as $\varphi(l_1, l_2, \alpha) = \alpha l_1 - (1-\alpha) l_2$; this map is continuous and $S \times S \times [0,1]$ is the product of compact sets, hence it is compact, and $K = \mathscr{R}(\varphi)$ is compact. Suppose now that $l \in U \setminus K$. By the corollary on the separation of a convex set and a point, there is a weak*-continuous functional f on $C(X)^*$ s.t. $f(l) > c$ and $f(m) \leqslant c$ for all $m \in K$. This f is of the form $f(m) = m(x)$ for some $x \in C(X)$, and since K is *symmetric* (*i.e.*, $m \in K$ iff $-m \in K$), $|m(x)| \leqslant c$ for all $m \in K$. Let $t \in X$ s.t. $|x(t)| = \max_{\tau \in X} |x(\tau)|$, and define $\hat{m} \in C(X)^*$ as $\hat{m}(y) = y(t)$ for all $y \in C(X)$, hence $\hat{m} \geqslant 0$ and $\|\hat{m}\| = 1$, so $\|x\| = |\hat{m}(x)| = \sup_{m \in S} |m(x)| = \sup_{m \in S} |f(m)| \leqslant c$. Thus, $\|x\| \leqslant c$ but $l(x) > c$, contradicting that $\|l\| \leqslant 1$; hence $K = U$. Now, take an $l \in C(X)^*$; assume w.l.o.g. that $\|l\| = 1$. As $l \in U = \mathrm{conv}(S \cup (-S))$, $l = \alpha l_1 - (1-\alpha) l_2$ for some $l_1, l_2 \in S$ and $\alpha \in [0,1]$. Define the positive functionals $l_+ = \alpha l_1$ and $l_- = (1-\alpha) l_2$, and note that $\|l_+\| + \|l_-\| = \alpha \|l_1\| + (1-\alpha)\|l_2\| \leqslant 1 = \|l\| = \|l_+ - l_-\| \leqslant \|l_+\| + \|l_-\|$. □

Lemma (Extreme points of Borel probability measures)
The extreme points of the set $\Delta(X)$ of Borel probability measures on X are the *delta measures* δ_x for $x \in X$, where $\delta_x(f) = f(x)$ for all $f \in C(X)$.

Proof (Barvinok, 2002, and Dunford&Schwartz, 1958)
Let $x \in X$. Assume that $\delta_x = \alpha l_1 + (1-\alpha) l_2$, where $l_1, l_2 \in \Delta(X)$ and $\alpha \in (0,1)$. If $f \in C(X)$ satisfies $f(x) = 1$ and $f(y) \leqslant 1$ for all $y \in X$, then $l_k(f) = l_k(\mathbb{1} - (\mathbb{1} - f)) = l_k(\mathbb{1}) - l_k(\mathbb{1} - f) \leqslant 1$ for $k = 1,2$, but $\delta_x(f) = f(x) = 1$, so $l_k(f) = 1$. Hence, l_1, l_2 agree with δ_x on every function with its maximum at x. Now, take any $f \in C(X)$, and write it as $f = f_1 - f_2$, where $f_1(y) = \min\{f(x), f(y)\}$ and $f_2(y) = \min\{0, f(x) - f(y)\}$. Then, f_1, f_2 attain their maxima at x, so $l_k(f) = l_k(f_1) - l_k(f_2) = f_1(x) - f_2(x) = f(x) = \delta_x(f)$, *i.e.*, δ_x is extreme.
Now, let $A = \{\delta_x : x \in X\} \subseteq \Delta(X)$. Since $\Delta(X)$ is convex and weak*-closed, $\overline{\text{conv}(A)} \subseteq \Delta(X)$ (the left side is the weak*-closure). If $l \in C(X)^* \setminus \overline{\text{conv}(A)}$, by the corollary on the separation between a convex set and a point, there is an $f \in C(X)$ s.t. $l(f) > c$ but $\delta_x(f) = f(x) \leqslant c$ for all $x \in X$, so $l \notin \Delta(X)$. Therefore, $\overline{\text{conv}(A)} = \Delta(X)$. On the other hand, A is weak*-closed in $C(X)^*$ (*why?*), and hence weak*-compact by Banach-Alaoglu, so the lemma on Slide 61 implies that all extreme points of $\Delta(X)$ are in A. □

Corollary. The extreme points of $U = \{l \in C(X)^* : \|l\| \leqslant 1\}$ are δ_x and $-\delta_x$ for $x \in X$.

Bonus: Convex Functions and Jensen's Inequality

Definitions

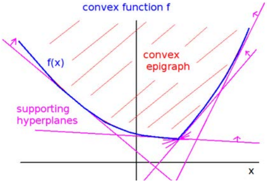
Let X be a normed space, and $f : X \to \mathbb{R}$.

- f is *convex* if $f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in X$ and $\lambda \in [0,1]$.
- $\operatorname{epi} f := \{(x,y) \in X \times \mathbb{R} \colon y \geqslant f(x)\}$: *epigraph* of $f : X \to \mathbb{R}$.

A function is *convex* iff its epigraph is a convex set, so it can be represented as the intersection of closed half-spaces that contain it (by Corollary 4 of Mazur's theorem). Now, for all $a, b \in \mathbb{R}$ and $x^* \in X^*$,

$$a \leqslant \langle x, x^* \rangle + by \quad \text{whenever } y \geqslant f(x)$$

$$\Leftrightarrow \quad a \leqslant \langle x, x^* \rangle + b f(x).$$



If $H = \{(x,y) \in X \times \mathbb{R} \colon a \leqslant \langle x, x^* \rangle + by\}$ contains $\operatorname{epi} f$, then b cannot be zero nor negative; otherwise f would be undefined when $a > \langle x, x^* \rangle$. Thus, every closed half-space containing $\operatorname{epi} f$ has the form $H = \{(x,y) \in X \times \mathbb{R} \colon c + \langle x, x^* \rangle \leqslant y\}$, and

$$f(x) = \sup_{(c, x^*) \colon c + \langle \tilde{x}, x^* \rangle \leqslant f(\tilde{x}) \text{ for all } \tilde{x} \in X} c + \langle x, x^* \rangle.$$

Bonus: Convex Functions and Jensen's Inequality (cont.)

Jensen's Inequality

Let X be a set, and E a Borel probability measure on a set of functions from X to $\mathbb{R}$. Also, let $g: X \to \mathbb{R}$ be s.t. $E(g)$ is well-defined, and let $f: \mathbb{R} \to \mathbb{R}$ be a convex function. Then

$$E(f \circ g) \geqslant f(E(g)),$$

if the left-hand side is well-defined.

Proof. Take $a, b \in \mathbb{R}$ s.t. $f(y) \geqslant ay + b$ for all $y \in \mathbb{R}$. Then $(f \circ g)(x) = f(g(x)) \geqslant ag(x) + b$ for all $x \in X$, so applying E yields $E(f \circ g) \geqslant aE(g) + b$. Taking the supremum over all $a, b \in \mathbb{R}$ s.t. $f(y) \geqslant ay + b$ for every $y \in \mathbb{R}$ finally gives

$$E(f \circ g) \geqslant \sup_{(a,b):\, a+by \leqslant f(y) \text{ for all } y \in X} aE(g) + b = f(E(g)).$$

$\square$

Remark. The Borel probability measure E can be, *e.g.*, the Riemann/Lebesgue integral, or the mathematical expectation operator with respect to a probability measure.

Weierstrass' Theorem, on the approximation of continuous functions on a real interval by polynomials, was considerably generalized by M.H. Stone in two directions:

- it applies to continuous functions on an arbitrary compact Hausdorff space, and
- instead of polynomials, it allows for more general algebras of functions.

We will provide a functional-analytic proof of this result, based on the following lemma:

Lemma (spanning criterion)
Let V be a normed space, $x \in V$, and $Y \subseteq V$. Then, $x \in \operatorname{clin} Y$ iff every $f \in V^*$ that vanishes on Y also vanishes on x (*i.e.*, if $f(y) = 0$ for all $y \in Y$, then $f(x) = 0$).

Proof

(⇒) If $f(y) = 0$ for all $y \in Y$, then by linearity $f(y) = 0$ for all $y \in \operatorname{lin} Y$, and by continuity $f(x) = 0$ since $x \in \operatorname{clin} Y$.

(⇐) Assume $x \notin \operatorname{clin} Y$, and define the functional f on $\operatorname{lin}(\operatorname{clin} Y + x)$ by $f(y + \alpha x) = \alpha$ for all $y \in \operatorname{clin} Y$ and α. Let $d = \inf_{y \in \operatorname{clin} Y} \|x - y\|$, which is strictly positive since $\operatorname{clin} Y$ is closed. Then, $\|y + \alpha x\| = |\alpha| \|\alpha^{-1} y + x\| \geqslant d|\alpha|$, so $\|f\| \leqslant d^{-1}$, and Corollary 1 of Hahn-Banach provides an extension $F \in V^*$ of f that vanishes on Y but not on x. □

Theorem (Stone-Weierstrass)

Let X be a compact Hausdorff space, and let E be a *subalgebra* of $C(X)$, *i.e.*, a linear subspace of $C(X)$ s.t. if $f,g \in E$, also $fg \in E$. In addition, assume that $\mathbb{1} \in E$, and that E *separates points* of X, *i.e.*, for every pair $x, y \in X$, $x \neq y$, there is an $f \in E$ s.t. $f(x) \neq f(y)$. Then, E is dense in $C(X)$.

Proof (adapted from version due to Louis de Branges)

By the spanning criterion, E is dense in $C(X)$ iff for every $l \in C(X)^*$, $l(x) = 0$ for all $x \in E$ implies that $l = 0$. Assume the contrary, and let U be the set of $l \in C(X)^*$ of norm $\leqslant 1$ that vanish on E; by Banach-Alaoglu, U is weak*-compact. By Krein-Milman, U has an extreme point, say, l, which should have unit norm. (*Why?*)

Let $g \in E$ be s.t. $0 < g(t) < 1$ for all $t \in X$. Since E is an algebra, the functionals $l', l'' \in C(X)^*$ defined by $l'(f) = l(gf)$ and $l''(f) = l((1-g)f)$ also vanish in E. Write l as $l = l_+ - l_-$, where $l_+, l_- \geqslant 0$, and also l' and l'' as $l'(f) = l(gf) = l_+(gf) - l_-(gf) =: l'_+(f) - l'_-(f)$ and $l''(f) = l_+((1-g)f) - l_-((1-g)f) =: l''_+(f) - l''_-(f)$, where $l'_+, l'_-, l''_+, l''_- \geqslant 0$, and note that $\|l\| = \|l_+\| + \|l_-\| = l_+(\mathbb{1}) + l_-(\mathbb{1}) = l'_+(\mathbb{1}) + l''_+(\mathbb{1}) + l'_-(\mathbb{1}) + l''_-(\mathbb{1}) = \|l'_+\| + \|l''_+\| + \|l'_-\| + \|l''_-\|$, while $\|l\| = \|l' + l''\| \leqslant \|l'\| + \|l''\| \leqslant \|l'_+\| + \|l''_+\| + \|l'_-\| + \|l''_-\|$, so combining these expressions we see that they should be all equalities, and in particular $\|l\| = \|l'\| + \|l''\|$.

Proof (cont.)
Let $a = \|l'\| \geqslant \min_{x \in X} g(x)\|l\| > 0$ and similarly $b = \|l''\| > 0$, so $a + b = \|l\| = 1$. Then, $l = a(l'/a) + b(l''/b)$ shows that l is a convex combination of l'/a and l''/b, both of which have unit norm and vanish in E. Since l is an extreme point, we have that $l = l'/a = l''/b$, *i.e.*, $l_{\pm}([1 - g/a]f) = 0$ for all $f \in C(X)$, or $|l|(g)\ |l|(f) = |l|(gf)$, where $|l|(f) = l_+(f) + l_-(f)$.

Using $|l|(g)\ |l|(f) = |l|(gf)$, we will now show that all functions in the nullspace of $|l|$ in E (which is a closed hyperplane) have a common zero. Indeed, every $g \in \mathrm{Ker}\ |l| \cap E$ should have a zero in X, because otherwise $g^{-1} \in C(X)$, so taking $f = g^{-1}$ gives $|l|(g)|l|(g^{-1}) = |l|(\mathbb{1}) > 0$, which contradicts the assumption that $|l|(g) = 0$. Now, if $g_1, \dots, g_n \in \mathrm{Ker}\ |l| \cap E$, then they share a common zero, since otherwise $g = g_1^2 + \cdots + g_n^2 \in E$ is strictly positive, while $|l|(g) = |l|(g_1^2) + \cdots + |l|(g_n^2) = [|l|(g_1)]^2 + \cdots + [|l|(g_n)]^2 = 0$, which is a contradiction. Finally, for every $g \in E$, let $X_g = \{x \in X : g(x) = 0\}$; X_g is a closed set in X, and every finite intersection $X_{g_1} \cap \cdots \cap X_{g_n}$ is non-empty, so by the compactness of X, $\bigcap_{g \in \mathrm{Ker}\ |l|} X_g \neq \emptyset$, *i.e.*, all functions in $\mathrm{Ker}\ |l| \cap E$ share a common zero.

Bonus: Stone-Weierstrass Theorem (cont.)

Proof (cont.)
The kernel of any $h \in E^*$ for which $h(fg) = h(f)h(g)$ (called a *homomorphism*, such as $|l|$) satisfies another important property: it is an *ideal* in E, *i.e.*, it is a linear subspace of E s.t. if $f \in E$ and $g \in \text{Ker}\ h$, then $fg \in \text{Ker}\ h$ (since $h(fg) = h(f)h(g) = 0$); furthermore, it is a *maximal ideal*, that is, a proper subset of E which is not contained in a larger proper ideal of E (since Ker h has co-dimension 1, so adding an extra dimension would yield E).

Other homomorphisms on E include $\delta_x|_E$ for every $x \in X$. Since E separates points on X, $\text{Ker}\ \delta_x|_E = \{g \in E : g(x) = 0\} \neq \text{Ker}\ \delta_y|_E$ for $x \neq y$ (indeed: pick a function $g \in E$ s.t. $g(x) = 0$ and $g(y) = 1$; this g belongs to $\text{Ker}\ \delta_x|_E$ but not to $\text{Ker}\ \delta_y|_E$).

Therefore, $\text{Ker}\ |l| \cap E \subseteq \text{Ker}\ \delta_x|_E$ for some $x \in X$ corresponding to the common zero of the functions in $\text{Ker}\ |l| \cap E$, but since $\text{Ker}\ |l|$ is a *maximal* ideal in E, we have that $\text{Ker}\ |l| \cap E = \text{Ker}\ \delta_x|_E$, *i.e.*, $|l|(f) = \alpha\delta_x(f) = \alpha f(x)$ for all $f \in E$, with $\alpha = 1$ due to the positivity of $|l|$.

Since δ_x is an extreme point of the unit ball of $C(X)^*$, $\delta_x = |l| = l_+ + l_-$ implies that $l_\pm = \alpha_\pm \delta_x$, so $l = c\delta_x$. On the other hand, $\mathbb{1} \in E$, so $c\delta_x(\mathbb{1}) = l(\mathbb{1}) = 0$, thus $c = 0$ and $l = 0$, which contradicts the assumption that $\|l\| = 1$, hence E is dense in $C(X)$. $\square$