IV. The (continuous) dual

We call

$$X^* := bL(X, \mathbb{F}) = \{\lambda \in X' : \lambda \text{ is bounded}\}$$

the **continuous dual** of the nls X. It is at times useful to know that X^* can be identified in a natural way with a closed lss of $b(B)$, the complete metric space of bounded functions on the unit ball B of X.

(1) Lemma. *For any nls X, the restriction map*

(2) $$r : X^* \to b(B) : \lambda \mapsto \lambda|_B$$

is linear, an **isometry** *(i.e., $\forall\{\lambda\}\ \|r(\lambda)\| = \|\lambda\|$), and its range is closed.*

Proof: Since λ is bounded, i.e., $\lambda(B)$ is a bounded subset of $\mathbb{F}$, the restriction $\lambda|_B$ of λ to B is a bounded function. Moreover,

$$\|\lambda\| = \sup|\lambda(B)| = \|\lambda|_B\|_\infty.$$

This shows that r is a linear isometry into $b(B)$. To see that its range is closed, let $g \in b(B)$ be the limit of some sequence $(\lambda_{n|B})$. Then g is necessarily linear on B. This makes it possible to extend g to a fnl λ on all of X by the recipe

$$\lambda x := \alpha g(x/\alpha), \quad \text{all } \alpha > \|x\|, \text{ all } x \in X$$

and to verify that $\lambda \in X'$, hence $\in X^*$ since $\|\lambda\| = \|g\|_\infty$. Therefore, $g = r(\lambda)$, i.e., ran r is closed. □

It follows that $\operatorname{ran} r$ is complete (as a closed subset of the complete ms $b(B)$), and, since r is an isometry, this implies that X^* is a **Banach space** (=: **Bs**), i.e., a *complete* nls.

(3) Proposition. *The continuous dual X^* of any nls X is complete.*

A second advantage that X^* has over X is that its closed unit ball is compact in a natural topology.

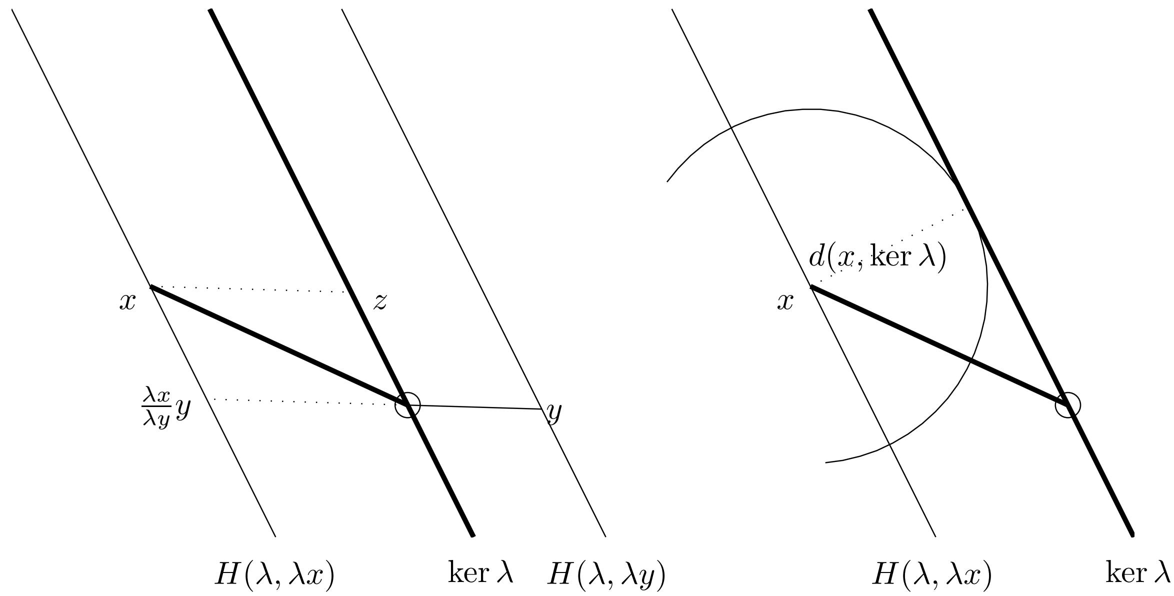
(4) Alaoglu's theorem. *The closed unit ball of X^* is compact in the topology of pointwise convergence.*

Proof: Let $Y := B^-_{X^*}$. Then Y is a subset of $\times_{x\in X}\{z \in \mathbb{F} : |z| \le \|x\|\}$, and the latter, by (II.41) Tykhonov's Theorem, is compact in the topology of pointwise convergence. Hence, by (II.25)Lemma, it is sufficient to prove that Y is closed in this topology.

For this, let f be an element in the pointwise closure of Y. Then, for arbitrary $\alpha, \beta \in \mathbb{F}$ and arbitrary $x, y \in X$, and arbitrary $\varepsilon > 0$, there is some $\lambda \in Y$ with $|f(s) - \lambda s| < \varepsilon$ for $s \in \{x, y, \alpha x + \beta y\}$. Consequently, $|f(x)| < |\lambda x| + \varepsilon \le \|x\| + \varepsilon$ and

$$|f(\alpha x+\beta y)-\alpha f(x)-\beta f(y)| = |(f-\lambda)(\alpha x+\beta y)-\alpha(f-\lambda)(x)-\beta(f-\lambda)(y)| < (1+|\alpha|+|\beta|)\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this shows that $f \in Y$. □



(6) Figure. A lfl is constant on hyperplanes parallel to its kernel. The distance of such a hyperplane from the kernel is proportional to the lfl's value on that hyperplane.

**** hyperplanes and lfl's ****

The action of a nontrivial (bounded or unbounded) lfl λ on a ls X is easy to visualize: Each element of X lies in exactly one of the **hyperplane**s

$$H(\lambda, t) := \{x \in X : \lambda x = t\},$$

and each such hyperplane is a translate of any other such hyperplane, since each is a translate of $\ker\lambda$, i.e.,

$$\forall\{x \in H(\lambda, t)\}\ H(\lambda, t) = \ker\lambda + x.$$

This reflects the fact that $H(\lambda, 0) = \ker\lambda$ has codimension 1, i.e.,

$$X = \ker\lambda \dot{+} \operatorname{ran}[y],$$

for any $y \notin \ker\lambda$; hence the term "hyperplane". To see this, observe that, for any $x \in X$ and any $y \notin \ker\lambda$, $x - (\lambda x/\lambda y)y \in \ker\lambda$.

**** elimination ****

This observation is the basic step of the well known numerical linear algebra process called

(5) Elimination. *To convert x into something in* $\ker\lambda$, *pick $y \notin \ker\lambda$ and compute*

$$z := x - (\lambda x/\lambda y)y \in \ker\lambda.$$

H.P.(1) Prove that $x \mapsto (\lambda x/\lambda y)y$ is a linear projector. What is its range, what are its interpolation functionals?

**** a useful formula ****

Elimination provides the following useful formula:

$$(7) \qquad \forall\{\lambda \in X'\}\ \forall\{x \in X\backslash \ker\lambda\} \quad \frac{d(x,\ker\lambda)}{|\lambda x|} = \begin{cases} 0 & (\ker\lambda)^- = X; \\ 1/\|\lambda\| & \text{otherwise.} \end{cases}$$

Indeed, if there is some $y \in X\backslash(\ker\lambda)^-$, then, by elimination, for any $x \in X$, $d(x,\ker\lambda) = d((\lambda x/\lambda y)y, \ker\lambda) = |\lambda x/\lambda y| d(y,\ker\lambda)$, hence $(\ker\lambda)^- = B_0^-(\ker\lambda) = \ker\lambda$ and further, for any $x \notin \ker\lambda$,

$$\|\lambda\| = \sup_{y\notin\ker\lambda} \frac{|\lambda y|}{\|y\|} = \frac{|\lambda x|}{d(x,\ker\lambda)} \sup_{y\notin\ker\lambda} \frac{d(y,\ker\lambda)}{\|y\|} = \frac{|\lambda x|}{d(x,\ker\lambda)},$$

the last equality by (III.7)Riesz' Lemma. The fact that $d(x,\ker\lambda) = 0$ for all $x \in X$ iff $\ker\lambda$ is dense in X, is trivial.

**** λ is continuous iff $\ker\lambda$ is closed ****

We infer from (7) that λ is bounded in case $(\ker\lambda)^- \neq X$. Put differently, it says that, for $\lambda \in X'\backslash X^*$, $(\ker\lambda)^- = X$, i.e., the kernel of a *discontinuous* lfl is dense. Since such a lfl is necessarily nontrivial, it says that $\ker\lambda \neq (\ker\lambda)^-$ for a discontinuous lfl. Conversely, if λ is continuous, then $\ker\lambda = \lambda^{-1}\{0\}$ is closed as the pre-image of a closed set under a continuous map. This proves:

(8) Proposition. *Let X be a ls, and $\lambda \in X'$. Then: λ is continuous iff $\ker\lambda$ is closed.*

(9) Corollary. *For $\lambda \in X'$, $\ker\lambda$ is either closed or dense (with $\ker\lambda$ both closed and dense iff $\lambda = 0$).*

**** error estimates ****

(7) proves the **useful identity**

(10) Lemma. $\forall\{\lambda \in X^*, x \in X\} \quad |\lambda x| = \|\lambda\| d(x,\ker\lambda)$.

This identity contains all basic Numerical Analysis error estimates, in the following way.

In applications of (10), λ is an error functional, and $\ker\lambda$ is not completely known. Rather, one knows some set F contained in $\ker\lambda$, and so obtains the bound

$$|\lambda x| \le \|\lambda\| d(x,F).$$

(11) Example. $\lambda := \int_a^b \cdot - h\sum_1^{n-1}\delta_{a+jh} - (h/2)(\delta_a + \delta_b)$ (with $h := (b-a)/n$) is the error in the composite trapezoidal rule, and this rule is exact for all linear polynomials. Further, on $C([a\,..\,b])$, $\|\lambda\| \le 2|b-a|$. Hence (with $d(f,Y) \le \|f\|$ for any lss Y)

$$|\lambda f| \le 2|b-a| d(f,\Pi_1) = O(|b-a|\|f\|_\infty).$$

Actually, the composite trapezoidal rule is exact for $\Pi^0_{1,h}$:= all broken lines on $[a\,..\,b]$ with breakpoints $a + jh$, all j, and $d(f,\Pi^0_{1,h}) \le \frac{1}{8}h^2\|D^2 f\|_\infty$. Hence $|\lambda f| \le \|\lambda\| d(f,\Pi^0_{1,h}) = O(|b-a|h^2\|D^2 f\|_\infty)$, a much better estimate.

**** existence of ba from a hyperplane ****

(12) Corollary. $\lambda \in X^*\backslash 0$ *takes on its norm at* x *iff 0 is a ba to* x *from* $\ker\lambda$.

Indeed, since $\|\lambda\|\|x\| \ge |\lambda x| = \|\lambda\| d(x, \ker\lambda)$, we have $\|\lambda\|\|x\| = |\lambda x|$ iff $\|x\| = d(x, \ker\lambda)$, and, since $0 \in \ker\lambda$, this last equality can only hold if 0 is a ba.

Since (III.13)Example provides an example of a bounded lfl that does not take on its norm, while its kernel is closed by (8)Proposition, this provides the illustration, promised earlier, of a closed lss that fails to provide ba's (since, if y is a ba from the lss Y to x, then 0 is a ba from Y to $x - y$).

H.P.(2) Let $x, k \in X$ nls, $\lambda \in X^*\backslash 0$. Prove: $|\lambda x - \lambda k| \le \|\lambda\|\|x-k\|$ *with equality iff* $B^-_{\|x-k\|}(x) \cap H(\lambda, \lambda k) \ne \{\} = B_{\|x-k\|}(x) \cap H(\lambda, \lambda k)$. Draw the picture.

One says that $\lambda \in X^*\backslash 0$ and $x \in X\backslash 0$ are **parallel** and writes

$$\lambda || x$$

in case $\lambda x = \|\lambda\|\|x\|$, in analogy to the situation in familiar Euclidean space; see below.

If $V = [v_1, v_2, \ldots, v_n]$ is a basis for the (necessarily finite-dimensional) nls X and $\Lambda = [\lambda_1, \lambda_2, \ldots, \lambda_n]$ is its **dual basis**, i.e., $\Lambda^t V = 1$, then, for each j, $\ker\lambda_j = \operatorname{ran}[v_i : i \ne j]$. Assume that V is **normalized**, meaning that $\|v_j\| = 1$ for all j. Since $H(\lambda_j, 1) = v_j + \ker\lambda_j$, therefore $\|\lambda_j\| = 1$ iff $\min\{\|x\| : x \in H(\lambda_j, 1)\} = 1$ iff $d(v_j, \operatorname{ran}[v_i : i \ne j]) = 1$ iff $\lambda_j || v_j$. Since, for a *matrix* V, $\det V$ is unchanged if we modify v_j by any element of $\operatorname{ran}[v_i : j \ne i]$, you can deduce from this, with a little bit of effort, the following.

(13) Auerbach's Theorem. *Any finite-dimensional nls has a normalized basis whose dual basis is also normalized.*

H.P.(3) Prove Auerbach's Theorem. (Hint: Show that it is sufficient to consider the case $X = \mathbb{F}^n$; choose a normalized $V \in \mathbb{F}^{n\times n}$ for which $|\det V|$ is maximal.)

Representation of bounded linear functionals

Practical as well as theoretical work with bounded lfl's on a specific nls relies heavily on *representation*. Typically, with X a nls of functions on some common domain T, one can often show that, for suitable f, the map

$$X \to \mathbb{F} : x \mapsto \int_T x(t) f(t)\, \mathrm{d}t$$

is a *bounded* linear functional on X. In that case, f is called the **representer** of this lfl. It is at times possible to represent every $\lambda \in X^*$ in this way as integration (or summation) against some function from a certain class. Such a representer is essential for the calculation of the norm of a linear functional. Such representation permits the translation of abstract results concerning the existence of some bounded linear functional on X into a concrete statement about the existence of some function with certain desirable properties. Precisely, a **representation** for X^* is an onto linear isometry (i.e., a linear norm-preserving invertible map) from some nls Y to X^*. One writes

$$Y \simeq X^*$$

if there is such a map (with the actual map often understood from the context) and calls the two spaces **linearly isometric** to each other.

Here are the standard representations for X^* for standard function spaces X.

(14) Examples. (i) $(\mathbb{F}^m)^*$. In order to be precise about the norm on $\mathbb{F}^m$ used, we use the notation

$$\ell_p(m) := (\mathbb{F}^m, \|\cdot\|_p).$$

Recall the 1-1 correspondence between $y \in \mathbb{F}^m$ and $y^t \in L(\mathbb{F}^m, \mathbb{F})$ given by the scalar product:

$$y^t : \mathbb{F}^m \to \mathbb{F} : x \mapsto y^t x := \sum_1^m y(i)x(i).$$

By (III.5)Proposition, $\ell_p(m)^* = L(\mathbb{F}^m, \mathbb{F})$. To determine the map norm of y^t as an element of $\ell_p(m)^*$, we use

(15) Hölder's inequality. *If* $y, x \in \mathbb{F}^m$, $1 \le p \le \infty$, $1/p + 1/p^* = 1$, *then* $|y^t x| \le \|y\|_{p^*}\|x\|_p$. *Equality:* $y^t x = \|y\|_{p^*}\|x\|_p$ *iff* $y = 0$ *or else, for some* $r \ge 0$,

$$x(i) = \begin{cases} r\,\text{signum}\, y(i)\ |y(i)|^{p^*/p}, & \text{if } 1 < p < \infty \\ \|x\|_\infty\ \text{signum}\, y(i) \ \ \text{for } y(i) \ne 0, & \text{if } p = \infty, \text{ hence } p^* = 1 \end{cases}, \quad \text{all } i.$$

Here, $\text{signum}\,\alpha$ is defined (implicitly) by

$$\alpha\ \text{signum}\,\alpha := |\alpha|,$$

i.e., $\text{signum}\,\alpha = \overline{\alpha}/|\alpha|$ for $\alpha \in \mathbb{F}\backslash 0$ (with $\overline{a}$ the complex conjugate of the possibly complex number a). For $\alpha = 0$, it is customary to define, more precisely, $\text{signum}\,\alpha := 0$.

Note that, for $p = 2$, equality in $y^t x \le \|y\|_2\|x\|_2$ occurs iff $y = 0$ or $x = r\overline{y}$ for some $r \ge 0$.

We conclude that $\|y^t\| = \sup |y^t x|/\|x\|_p = \|y\|_{p^*}$, hence that $\ell_p(m)^* \simeq \ell_{p^*}(m)$, with p^* the **conjugate** of p, i.e., $1/p + 1/p^* = 1$.

(ii) ℓ_p. The nls

$$\ell_p := \ell_p(\mathbb{N}) := \{x \in \mathbb{F}^{\mathbb{N}} : \|x\|_p := \big(\sum_n |x(n)|^p\big)^{1/p} < \infty\}$$

has ℓ_{p^*} as its continuous dual, as long as $1 \le p < \infty$. It is again Hölder's Inequality that makes it possible to establish this. In particular, for $x \in \ell_p$ and $1 \le p < \infty$, $x = \lim_{n\to\infty} \sum_{j\le n} e_j\, x(j)$, hence, for any $\lambda \in \ell_p^*$, $\lambda x = \lim_{n\to\infty} \lambda(\sum_{j\le n} e_j\, x(j)) = \sum_j \lambda(e_j)\, x(j)$, therefore $\|(\lambda e_j)_{j=1}^\infty\|_{p^*} \le \|\lambda\|$; etc. See H.P.(4) for full details in a related case.

The continuous dual of

$$\ell_\infty := \ell_\infty(\mathbb{N}) := \{x \in \mathbb{F}^{\mathbb{N}} : \|x\|_\infty := \sup |x(n)| < \infty\}$$

contains ℓ_1, but contains other things besides. (Actually, it's a mess.)

ℓ_1 is the continuous dual of some sequence space, viz. the closed subspace

$$c_0 := \{x \in \ell_\infty : \lim x(n) = 0\}$$

of ℓ_∞ consisting of all null sequences (i.e., sequences converging to 0), as you will show.

H.P.(4)
(i) Prove that c_0 is a closed lss of ℓ_∞. (Hint: $c_0 = \ker\nu$, with $\nu : \ell_\infty \to \mathbb{R}_+ : x \mapsto \limsup |x(n)|$.)
(ii) Let P_n be the **truncation projector on** $\mathbb{F}^{\mathbb{N}}$, i.e.,

$$(P_n x)(i) := \begin{cases} x(i), & i \le n \\ 0, & i > n \end{cases}.$$

Prove that, for all $x \in c_0$, $\lim P_n x = x$.
(iii) For $\lambda \in c_0^*$, let $y_\lambda : \mathbb{N} \to \mathbb{F} : n \mapsto \lambda e_n$, with $e_n := (\delta_{in} : i \in \mathbb{N})$ the nth unit-sequence, all n. Use P_n to prove that

$$\forall\{x \in c_0\}\ \lambda x = \sum_1^\infty y_\lambda(i)x(i).$$

(iv) Prove that $\|\lambda\| = \|y_\lambda\|_1$.
(v) Prove that $c_0^* \simeq \ell_1$.

H.P.(5) Prove that the closed lss

$$c := \{x \in \ell_\infty : \lim x(n) \text{ exists}\}$$

of ℓ_∞ of all convergent sequences has on it a continuous linear functional λ that *cannot* be represented as a scalar product $x \mapsto \sum_i y_\lambda(i)x(i)$ and so conclude that c^* is larger than ℓ_1 in this sense.

(iii) $\mathbf{L}_p[a\,..\,b]$. For $1 \le p < \infty$, all continuous linear functionals on

$$X := (C([a\,..\,b]), \|\cdot\|_p)$$

can be represented with the aid of a scalar product, i.e., for all $\lambda \in X^*$, there exists a function y_λ so that

$$\lambda x = \int_a^b y_\lambda(t)x(t)\,\mathrm{d}t, \text{ all } x \in X.$$

To make this precise, though, it is not sufficient to use the Riemann integral. For, while every continuous or even piecewise continuous function y on $[a..b]$ gives rise to a continuous lfl y^t on X via

$$y^t x := \int_a^b y(t)x(t)\,\mathrm{d}t,$$

not every $\lambda \in X^*$ is obtainable this way. It is necessary to admit all functions $y \in \mathbf{L}_{p^*}[a..b]$, i.e., all y on $[a\,..\,b]$ for which $|y|^{p^*}$ is **Lebesgue integrable**, i.e., for which

$$\left(\|y\|_{p^*}\right)^{p^*} := \int_a^b |y(t)|^{p^*}\,\mathrm{d}t < \infty,$$

with the integral taken in the Lebesgue sense. The salient facts concerning Lebesgue integration can be found, e.g., in Groetsch. The notion is built upon the **Lebesgue measure**, which is a particular way to assign to certain sets a nonnegative number, their **measure**, in an organized way, e.g., so that the measure of a disjoint union of sets is the sum of their measures, and so that the measure of any interval is its length or diameter, etc. There is no time to develop this in the present course nor do we really need to if you are willing to take certain facts on faith. E.g., a **set of measure zero** is any set that, for any $\varepsilon > 0$, can be covered by the union of open intervals whose lengths sum to no more than ε. The relationship to Riemann integration is as follows. In **Riemann integration**, we deal with partitions Δ of $[a \,.\,. b]$ into finitely many *intervals* I. With each such partition Δ, we associate the set

$$v_\Delta := \sum_{I\in\Delta} f(I)\,\text{meas}(I)$$

where meas(I) is the measure of the interval I, i.e., its length. If Δ' is also a partition of $[a \,.\,. b]$, then one says that Δ' **refines** Δ and writes $\Delta' \gg \Delta$ in case $\forall\{I \in \Delta\}\ \exists\{I_1, \ldots, I_r \in \Delta'\}\ I = I_1 \cup \ldots \cup I_r$. Correspondingly, $v_{\gg\Delta} := \cup_{\Delta'\gg\Delta} v_{\Delta'}$. With this, one defines

$$v = \int_a^b f(t)\,\mathrm{d}t \quad :\Longleftrightarrow \quad \big(v_{\gg\Delta}\big) \succ \mathbf{B}(v),$$

i.e., $\forall\{\varepsilon > 0\}\ \exists\{\Delta\}\ \forall\{\Delta' \gg \Delta\}\ \forall\{w \in v_{\Delta'}\}\ |w - v| < \varepsilon$. In **Lebesgue integration**, more general partitions are allowed, hence a function is more likely to be Lebesgue integrable than Riemann integrable. Now a partition can consist of any kind of measurable sets, not just intervals, but the definition of integral stays otherwise the same.

(16) Hölder's inequality. *If $y \in \mathbf{L}_{p*}[a \,.\,. b]$ and $x \in \mathbf{L}_p[a \,.\,. b]$ with $1/p + 1/p^* = 1$, then the product $yx : t \mapsto y(t)x(t)$ is Lebesgue integrable and*

$$|y^t x| = |\int_a^b y(t)x(t)\,\mathrm{d}t| \le \|y\|_{p^*}\|x\|_p.$$

Equality: $y^t x = \|y\|_{p^*}\|x\|_p$ iff $y = 0$ or else, for some $r \ge 0$,

$$x(t) = \begin{cases} r\,\text{signum}\,y(t)\ |y(t)|^{p^*/p}, & \text{if } 1 < p < \infty \\ \|x\|_\infty\,\text{signum}\,y(t), & \text{if } y(t) \ne 0, p = \infty \end{cases} \quad \text{a.e.}$$

Here, **a.e.** := **almost everywhere** is meant to indicate that the asserted relation is to hold for all $t \in [a \,.\,. b]$ excepting a set of measure zero. This exception will appear always since changing a function y on a set of measure zero will not change y^t, i.e., will not change its action as a linear functional on $C([a \,.\,. b])$. In particular, it will not change its norm $\|y\|_{p^*}$. For this reason, $\mathbf{L}_{p^*}[a \,.\,. b]$ consists, strictly speaking, not of functions, but of equivalence classes of functions, with x and y belonging to the same equivalence class iff $x = y$ a.e. Hölder's inequality allows us to talk about the continuous dual of $\mathbf{L}_p[a \,.\,. b]$ as well. If $p < \infty$, then this turns out to be representable via the scalar product by $\mathbf{L}_{p^*}[a \,.\,. b]$, with $1/p^* + 1/p = 1$. For $p = \infty$, though, the continuous dual is much messier.

(iv) The previous example can be further generalized. Instead of an interval, one might take some suitable subset of $\mathbb{R}^m$. Further, measures other than the Lebesgue measure could be used (e.g., a weighted Lebesgue integral or a discrete measure). Finally, instead of scalar-valued functions, one may consider functions into some fixed nls Y, with the norm in Y playing the role played by the absolute value in $\mathbb{F}$. For example, the obvious consequence

$$\|\sum_{y\in M} y w(y)\| \ \le\ \sum_{y\in M} w(y)\|y\|$$

of the triangle inequality in Y to finite linear combinations with nonnegative weights has the following obvious continuous analog, called **Minkowski's inequality for integrals**,

$$\|\int_a^b w(t)y(t)\,\mathrm{d}t\| \;\le\; \int_a^b w(t)\|y(t)\|\,\mathrm{d}t,$$

in which $y : [a \mathinner{\ldotp\ldotp} b] \to Y$ and w is nonnegative.

(v) $C([a \mathinner{\ldotp\ldotp} b])$. The **Riesz Representation Theorem** asserts that the continuous dual of $X := C([a \mathinner{\ldotp\ldotp} b])$ (with the max-norm) can be represented by the space $NBV[a \mathinner{\ldotp\ldotp} b]$ of functions of *Normalized Bounded Variation*. This means that, for $\lambda \in X^*$, there exists exactly one $y_\lambda \in NBV[a \mathinner{\ldotp\ldotp} b]$ so that

$$\lambda x = \int_a^b x(t)\,\mathrm{d}y_\lambda(t), \quad \text{all } x \in C([a \mathinner{\ldotp\ldotp} b]),$$

with the integral taken in the Riemann-Stieltjes sense (which means that $\int_a^b x(t)\,\mathrm{d}y(t)$ is the limit of sums $\sum_{I\in\Delta} x(I)\,\mathrm{meas}_y(I)$, with Δ an arbitrary finite partition of $[a \mathinner{\ldotp\ldotp} b]$ into intervals $I =: [I^- \mathinner{\ldotp\ldotp} I^+]$, and $\mathrm{meas}_y(I) := y(I^+) - y(I^-)$).

For example, the lfl δ_v of point evaluation at $v > a$ is represented by the **Heaviside** function

$$(\cdot - v)_+^0 = \chi_{[v..b]} : t \mapsto \begin{cases} 1, & t \ge v \\ 0, & t < v \end{cases},$$

with

$$(t)_+ := \max\{0, t\}$$

the **truncation** function. (What is δ_a represented by?)

To recall, the (total) **variation** of $y : [a \mathinner{\ldotp\ldotp} b] \to \mathbb{R}$ is, by definition, the (extended) real number

$$\mathrm{var}\, y := \sup_{a\le t_1<\cdots<t_r\le b} \sum_j |y(t_{j+1}) - y(t_j)|,$$

and y is said to be of **bounded variation** if this number is finite. For any function y on $[a \mathinner{\ldotp\ldotp} b]$ of bounded variation and any $g \in X = C([a \mathinner{\ldotp\ldotp} b])$, and any interval partition Δ, $|\sum_{I\in\Delta} x(I)\,\mathrm{meas}_y(I)| \le \|x\|_\infty \,\mathrm{var}\, y$, while, for any refinement Δ' of Δ,

$$|\sum_{I\in\Delta} x(I)\,\mathrm{meas}_y(I) - \sum_{J\in\Delta'} x(J)\,\mathrm{meas}_y(J)| \le \omega(x, |\Delta|)\,\mathrm{var}\, y,$$

with $|\Delta| := \max_{I\in\Delta} |I_+ - I_-|$. Hence

$$\lambda_y : X \to \mathbb{F} : x \mapsto \int_a^b x(t)\,\mathrm{d}y(t)$$

is well-defined, linear, and bounded, with $\|\lambda_y\| \le \mathrm{var}\, y$.

Conversely, let $\lambda \in X^*$. By the Hahn-Banach Theorem (27) later in this chapter, there exists a norm-preserving extension of λ to $X_1 := b([a\mathinner{\ldotp\ldotp} b])$; let μ be any such and define

$$y := y_\lambda : [a\mathinner{\ldotp\ldotp} b] \to \mathbb{F} : s \mapsto \mu\,\chi_{[a..s]}.$$

Then, $\operatorname{var} y \le \|\lambda\|$ since, for any interval partition Δ of $[a\mathinner{\ldotp\ldotp} b]$, and with $\varepsilon_I(y(I_+) - y(I_-)) := |y(I_+) - y(I_-)|$ for $I \in \Delta$,

$$\sum_{I\in\Delta} |y(I_+) - y(I_-)| = \sum_{I\in\Delta} \varepsilon_I(y(I_+) - y(I_-)) = \mu\left(\sum_{I\in\Delta} \varepsilon_I \chi_{(I_-..I_+]}\right) \le \|\mu\| = \|\lambda\|.$$

Further, for any $x \in X$, the piecewise constant function

$$I_\Delta x := \sum_{I\in\Delta} x(I_-)(\chi_{[a..I_+]} - \chi_{[a..I_-]})$$

converges uniformly to x as $|\Delta| \to 0$ while

$$\mu I_\Delta x = \sum_{I\in\Delta} x(I_-)(y(I_+) - y(I_-))$$

converges to $\int_a^b x(t)\,\mathrm{d}y(t)$ by the latter's definition. Hence, by the continuity of μ, the integral must equal $\mu x = \lambda x$. This implies that $\lambda_y = \lambda$, hence, in particular, $\|\lambda\| \le \operatorname{var} y$, therefore, altogether, $\operatorname{var} y_\lambda = \|\lambda\|$.

We now know that the map $y \mapsto \lambda_y \in X^*$ is norm-reducing and onto, but there is no claim that it is 1-1. For that, one selects, from the many y that represent λ, a particular one, called normalized, namely the one that vanishes at a and is continuous from the right at every $t > a$.

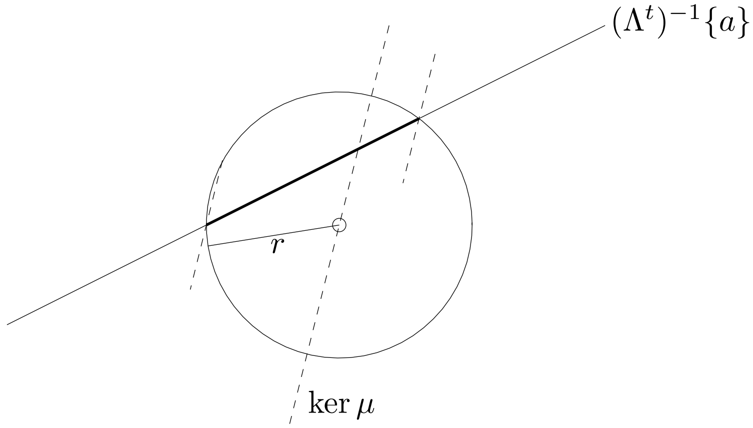
The collection of all such normalized functions of bounded variation is denoted by $NBV[a\mathinner{\ldotp\ldotp} b]$, for **Normalized Bounded Variation**. It is not hard to see that

$$NBV[a\mathinner{\ldotp\ldotp} b] \;=\; NM[a\mathinner{\ldotp\ldotp} b] - NM[a\mathinner{\ldotp\ldotp} b],$$

with $NM[a..b]$ the collection of all $f : [a..b] \to \mathbb{R}$ with (i) $f(a) = 0$, (ii) f right-continuous on $(a\mathinner{\ldotp\ldotp} b]$, (iii) $f(b) < \infty$, and (iv) $s < t \Longrightarrow f(s) \le f(t)$.

H.P.(6) Prove: *Let f and g be functions defined on the interval $[a\mathinner{\ldotp\ldotp} b]$. If $\int_a^b f(s)\,\mathrm{d}g(s)$ is defined, then so is $\int_a^b g(s)\,\mathrm{d}f(s)$, and their sum equals $fg|_a^b := (fg)(b) - (fg)(a)$.*

H.P.(7) Prove that, on $bC(T)$ (with the max-norm and with T ms) and for any finite $U \subseteq T$ and any $a \in \mathbb{F}^U$, the lfl $\sum_{u\in U} a(u)\delta_u$ has norm equal to $\|a\|_1$. (You may wish to prove first that $\forall\{s > 0, p \in T\}$ the function $T \to \mathbb{R} : t \mapsto (1 - d(t,p)/s)_+$ is continuous.) Does this still hold when $\#U \not< \infty$ (but still $\|a\|_1 := \sum_{u\in U} |a(u)| < \infty$)?



(17) Figure. The hypercircle

Application: Interpolation error and optimal recovery

We are now in a position to take up again the problem (introduced in (I.35)Problem) of estimating μg for given $\mu \in X'$ from the information that $\Lambda^t g = a$ for some $\Lambda \in L(\mathbb{R}^m, X')$. We found then that we could either give μg exactly (in case $\mu = c^t\Lambda^t$, i.e., $\mu \in \operatorname{ran}\Lambda$), or else we could say nothing.

We now consider the same problem with the additional information that

$$\mu \in X^*, \quad \Lambda \in bL(\mathbb{R}^m, X^*), \quad \|g\| \le r.$$

Throughout the discussion, we will assume that $(\Lambda^t)^{-1}\{a\} \neq \{\}$, i.e., there actually is some $g \in X$ with $\Lambda^t g = a$. This will be so regardless of what a might be in case Λ^t is onto, i.e., in case Λ is 1-1 (cf. (I.32)Proposition). With this assumption, the set $\mu(B_r^- \cap (\Lambda^t)^{-1}\{a\})$ of all possible values for μg is a bounded interval, and is contained in the intersection $\cap_{\varepsilon>0}\mu(B_{r+\varepsilon} \cap (\Lambda^t)^{-1}\{a\})$ which, as we will see, is the Golomb-Weinberger interval for μg; see (22).

**** model example ****

Throughout the discussion, you might hang on to the following example:

$X = C^{(1)}[0\,..\,1]$ with the norm

$$\|x\| := \max\{|x(0)|, \|Dx\|_\infty\},$$

$\Lambda^t g := g|_U$ with $\#U = m$, $\mu g := \int_0^1 g(t)\,\mathrm{d}t$.

**** interpolation ****

Also recall from Chapter I that the standard approach to estimating μg is by way of a **rule**, i.e., to approximate μg by $\mu P g$, with P the lprojector given by some F and L. If Λ is a basis for L, and V is the corresponding dual basis for F, then $P = V\Lambda^t = \sum_i [v_i]\lambda_i$, hence

$$\mu P = \mu V \Lambda^t = \sum_i (\mu v_i)\lambda_i.$$

**** model example (cont.) ****

If $X = C^{(1)}[0\,..\,1]$, $\mu := \int_0^1 \cdot$, and $\Lambda^t g := g|_U$ with $\#U = m$, and $F = \Pi_{<m}$, then $\ell_u : t \mapsto \prod_{u' \neq u}(t-u')/(u-u')$, $u \in U$, is the sequence in F dual to $(\delta_u : u \in U)$, and the resulting **quadrature rule** for the **nodes** U is

$$\int_0^1 g(t)\,\mathrm{d}t \sim \sum w_u g(u),$$

with the **weights**

$$w_u = \mu \ell_u = \int_0^1 \ell_u(t)\,\mathrm{d}t, \quad u \in U.$$

**** Lebesgue inequality ****

Since P is a linear projector, we have

$$\forall\{\lambda \in L\}\ \lambda(1-P) = 0 \quad \text{and} \quad \forall\{f \in F\}\ (1-P)f = 0,$$

therefore

$$\forall\{\lambda \in L,\ f \in F\}\ \mu g - \mu P g = \mu(1-P)g = (\mu - \lambda)(1-P)(g-f).$$

By minimizing over all $\lambda \in L$ and $f \in F$, this gives the

(18) Lebesgue inequality. $|\mu g - \mu P g| \le d(\mu, L)\|1-P\| d(g,F)$.

Precise estimates for $d(g,F)$ and/or $d(\mu,L)$ can be obtained in specific instances through the use of Approximation Theory. A formula for $\|P\|$ is given in (38). For the time being, I merely settle when $P \in bL(X)$, i.e., when $\|P\| < \infty$:

(19) Proposition. $P \in bL(X) \iff L \subseteq X^*$.

Proof: $P \in bL(X) \implies \Lambda^t = \Lambda^t V \Lambda^t = \Lambda^t|_F P$ is bounded, since $\dim F < \infty$. Conversely, if $L \subseteq X^*$, then Λ^t is bounded, while V is always bounded, hence so is $P = V\Lambda^t$. □

**** segue into optimal recovery ****

If we only know that $\|g\| \le r$, then the best estimate for $d(g,F)$ we can give is r, because of (III.7)Riesz' Lemma. Further, $\mu P g$ is entirely computable from $\Lambda^t g$. Hence, from Lebesgue's Inequality, this gives the *computable* interval

$$\mu P g \pm d(\mu, L)\|1-P\| r$$

within which μg must lie.

**** best rule in the sense of Sard ****

Sard pointed out that we can compute the (usually) better interval

$$\overset{\mathfrak{M}}{\lambda} g \pm d(\mu, L) r$$

by the simple device of determining the ba $\overset{\mathfrak{M}}{\lambda}$ to μ from $L = \operatorname{ran} \Lambda$. For, since we know the vector $a := \Lambda^t g$, we can calculate $(\Lambda c) g = c^t \Lambda^t g$ for any $c \in \mathbb{R}^m$. With this, we get the computable estimate

$$\forall \{c \in \mathbb{R}^m\} \; |\mu g - (\Lambda c) g| \le \|\mu - \Lambda c\| r. \tag{20}$$

In particular, we can choose for $\lambda := \Lambda c$ a **best approximation from L to μ**, i.e.

$$\overset{\mathfrak{M}}{\lambda} \in L \text{ s.t. } \|\mu - \overset{\mathfrak{M}}{\lambda}\| = d(\mu, L).$$

Such a $\overset{\mathfrak{M}}{\lambda}$ (read 'lambda crown(ed)') is called a **best rule** (for μ from L) **in the sense of Sard**, and the resulting bound

$$\overset{\mathfrak{M}}{\lambda} g - d(\mu, L) r \;\le\; \mu g \;\le\; \overset{\mathfrak{M}}{\lambda} g + d(\mu, L) r \tag{21}$$

is called the **Sard interval** for μg (based on the information $\Lambda^t g = a$ and $\|g\| \le r$).

**** GW-interval ****

Often, though, we can do even better than that by milking (20) for all the information it contains. From (20), we deduce that

$$-\|\mu - \lambda\| r \le \mu g - \lambda g \le \|\mu - \lambda\| r, \quad \text{all } \lambda \in L,$$

therefore

$$-\inf_{\lambda \in L} (\lambda g + \|\mu + \lambda\| r) \;=\; \sup_{\lambda \in L} (\lambda g - \|\mu - \lambda\| r) \le \mu g \le \inf_{\lambda \in L} (\lambda g + \|\mu - \lambda\| r). \tag{22}$$

We call this the **Golomb-Weinberger interval** for μg. This interval is nonempty when the given information is consistent, i.e., when

$$\{g \in X : \Lambda^t g = a, \; \|g\| \le r\} \ne \{\}, \tag{23}$$

but the converse need not hold. In fact, the GW-interval is continuous and monotone in r (as well as in $\|\cdot\|$) as well as closed, hence, for given $r = r_0$, is equal to the intersection of all intervals for $r > r_0$. Since (23) is satisfied for any r greater than $r_a := \inf\{\|g\| : g \in X, \Lambda^t g = a\}$, it follows that the GW-interval is not empty for $r = r_a$, even though (23) might be violated in that case.

H.P.(8) Locate r_a in (17)Figure. Give an example, in a *complete* nls, for which $r_a = r$ but (23) does not hold.

As we will see later (see the discussion preceding (40)Proposition), the GW-interval provides the best possible information about μg given that we know only that $\Lambda^t g = a$ and that $\|g\| \le r$. The interval end points are computable (in principle) since they are given as the inf (sup) of a convex (concave) function on a finite-dimensional linear space. In particular, these are minima (maxima). For this computation, it is necessary to evaluate the norm $\|\nu\|$ of $\nu := \mu - \lambda$, and that is usually not possible unless one knows how to represent the action of ν in terms of some scalar product.

**** a sharp estimate for $\int g$ ****

(24) Example. Bound $\mu g := \int_0^1 g(t)\,\mathrm{d}t$, given that $g(0) = 0, g(1) = 1$, and $\|Dg\|_\infty \le 2$ (e.g., $g = ()^2$).

Choose $X = C^{(1)}[0\,..\,1]$, with $\|x\| := \max\{|x(0)|, \|Dx\|_\infty\}$. Then $\|g\| \le r$ with $r = 2$. Also, $\Lambda^t g = (g(0), g(1))$.

To obtain a formula for $\|\nu\| = \|\mu - \lambda\|$, use the Fundamental Theorem of Calculus

$$x(t) = x(0) + \int_0^t (Dx)(s)\,\mathrm{d}s = x(0) + \int_0^1 (t-s)_+^0 Dx(s)\,\mathrm{d}s$$

to write νx in terms of $x(0)$ and Dx, i.e., as

$$\nu x = a x(0) + \int_0^1 b(s) Dx(s)\,\mathrm{d}s$$

for some coefficient a and some function b. Then

$$\|\nu\| = \sup \frac{a x(0) + \int b Dx}{\max\{|x(0)|, \|Dx\|_\infty\}} = |a| + \|b\|_1 = |a| + \int_0^1 |b(s)|\,\mathrm{d}s$$

with equality achieved by $x := \mathrm{signum}(a) + \int_0^{\cdot} \mathrm{signum}\, b$. This function is not in X in case b changes sign since then x has only a piecewise continuous first derivative, but there are $y \in X$ close to this x in the sense that $\|y\| \sim \|x\|$ and $\nu y \sim \nu x$. I could have avoided this point by including in X all functions with piecewise continuous first derivative.

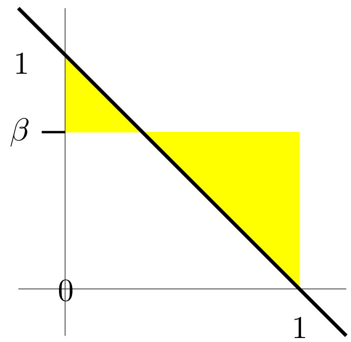
For our particular situation, we compute

$$\mu x = x(0) + \int_0^1 \int_0^1 (t-s)_+^0 Dx(s)\,\mathrm{d}s\,\mathrm{d}t = x(0) + \int_0^1 (1-s) Dx(s)\,\mathrm{d}s$$

$$\delta_0 x = x(0), \quad \delta_1 x = x(0) + \int_0^1 Dx(t)\,\mathrm{d}t.$$

So, $\nu := \mu - \Lambda c =: \mu - \alpha\delta_0 - \beta\delta_1$ has the *representation*

$$\nu x = (1 - \alpha - \beta) x(0) + \int_0^1 \{(1-s) - \beta\} Dx(s)\,\mathrm{d}s.$$



(25) Figure. $\|1-\beta-\cdot\|_1$ equals the shaded (unsigned) area; it is at a minimum when $\beta = 1/2$.

Therefore

$$\|\nu\| = \|\mu - \alpha\delta_0 - \beta\delta_1\| = |1-\alpha-\beta| + \|1-\beta-\cdot\|_1.$$

(i) Find $\overset{\Omega}{\lambda}$, i.e., minimize $\|\nu\| = \|\mu - \alpha\delta_0 - \beta\delta_1\|$ over α, β. Whatever β might be, the choice $\alpha = 1-\beta$ can only make $\|\nu\|$ smaller; so we choose α that way. This leaves the task of minimizing

$$\|1-\beta-\cdot\|_1 = \begin{cases} (1-2\beta)/2, & \beta \le 0 \\ (\beta^2 + (1-\beta)^2)/2, & 0 \le \beta \le 1 \\ (2\beta - 1)/2, & 1 \le \beta \end{cases}.$$

This is uniquely minimized when $\beta = 1/2$, hence $\|\mu - \overset{\Omega}{\lambda}\| = 1/4$ and $\overset{\Omega}{\lambda} = (\delta_0 + \delta_1)/2$, the *trapezoidal* rule. Thus $\overset{\Omega}{\lambda} g = 1/2$ and $|\mu g - 1/2| \le 2(1/4) = 1/2$, i.e., the Sard estimate is

$$\mu g \in [0\,.\,.\,1].$$

(ii) Find the GW-interval. This requires the computation

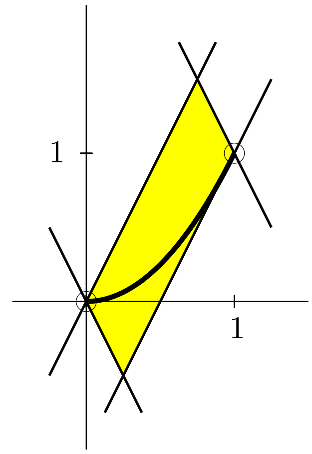
$$\begin{matrix}\sup\\ \inf\end{matrix} \bigl(\lambda g \,{}^{-}_{+}\, r\|\mu - \alpha\delta_0 - \beta\delta_1\|\bigr) = \begin{matrix}\sup\\ \inf\end{matrix} \bigl(\beta \,{}^{-}_{+}\, 2\|1-\beta-\cdot\|_1\bigr) = \begin{matrix} 1/8 \text{ (at } \beta = 3/4) \\ 7/8 \text{ (at } \beta = 1/4) \end{matrix},$$

in which, once again, the best choice of α makes the term involving it disappear, while the extremization of $\beta \mapsto \beta \mp 2\|1-\beta-\cdot\| = \beta \mp (\beta^2 + (1-\beta)^2)$ on $[0\,.\,.\,1]$ leads to the equation $1 \mp 2(\beta - (1-\beta)) = 0$, or $\beta = \frac{1}{2} \pm \frac{1}{4}$. This gives the better estimate

$$\mu g \in [1/8\,.\,.\,7/8].$$

In this simple example, we can obtain the closure of the interval of all possible values for μg immediately as follows. Since $\|Dg\|_\infty \le 2$, and $g(0) = 0, g(1) = 1$, the function g must lie inside the cones with vertex $(t, g(t))$, $t = 0, 1$, and sides having slopes ± 2. The pointwise smallest function in the resulting parallelogram is made up of its two lower sides, the pointwise biggest of its two upper sides, and these give the extreme values 1/8 and 7/8 for μg.

Note that $r < 1$ would have been inconsistent with $\Lambda^t g = (0, 1)$, i.e., there would have been no solution, while, for $r = 1$, the GW-interval would be $\{1/2\}$.



(26) Figure. All functions f that agree with $()^2$ at 0 and 1 and also have $\|Df\|_\infty \le 2$ must lie in the shaded area.

H.P.(9) Consider the example with the additional information that, also, $g(1/2) = 1/4$. (Note: This still fits the function $()^2$.)
(i) Determine again the GW-interval. (Feel free to use the abovementioned shortcut.)
(ii) Determine Sard's best rule and the resulting interval for μg.

H.P.(10) Find the best estimate for $f(b)$, given that $f(a) = c$, $\int_a^b f(t)\,dt = d$, and f is Lipschitz-continuous with Lipschitz constant $\kappa \le r$. Actually, that is a bit messy, so, find the best rule in the sense of Sard and find a formula for the smallest possible value of r in terms of a, b, c, d. (If you feel more comfortable with numbers, you may choose $(a, b, c, d, r) = (3, 4, 1, 0, 2)$.)

Hahn-Banach

The optimal recovery discussion contains the essential step in the classical proof of the

(27) Hahn-Banach Theorem. *Any $f \in Y^*$ for a lss Y of the nls X has a norm-preserving extension $g \in X^*$, i.e., there exists $g \in X^*$ such that $\|g\| \le \|f\|$ and $g|_Y = f$.*

For, the extension is made up one dimension step at a time. This means that we try to extend f to the lss $Y_1 := \mathrm{ran}[z] + Y$ for any particular $z \notin Y$. Call the extension g. Then

$$g : Y_1 \to \mathbb{R} : \alpha z + y \mapsto \alpha\zeta + fy$$

with the scalar $\zeta = gz$ to be chosen so that $\|g\| \le \|f\|$. You see the parallel? We know g on Y and wonder what gz could be, given that we know that $\|g\| \le \|f\|$. Therefore, the possible range of values ζ is given by the GW-interval

$$\sup_{y\in Y} fy - \|f\|\,\|z - y\| \le \zeta \le \inf_{y\in Y} fy + \|f\|\,\|z - y\|. \tag{28}$$

To verify this directly, recall that (28) is equivalent to

$$|\zeta - fy| \le \|f\|\,\|z - y\|, \quad \text{all } y \in Y,$$

hence to

$$|\alpha\zeta - fy| \le \|f\|\,\|\alpha z - y\|, \quad \text{all } \alpha \in \mathbb{R}\backslash 0, y \in \alpha Y = Y,$$

therefore to

$$|gy| \le \|f\|\,\|y\|, \quad \text{all } y \in Y_1,$$

i.e., to $\|g\| \le \|f\|$. It remains to show that the GW-interval (28) is not empty. With $g = f$ on Y, this follows from the fact that

$$\forall\{y, y' \in Y\}\; fy - \|f\|\,\|z - y\| \le fy' + \|f\|\,\|z - y'\|,$$

i.e.,

$$f(y - y') \le \|f\|\bigl(\|z - y\| + \|z - y'\|\bigr).$$

This inequality is obvious once one sees that

$$f(y - y') \le \|f\|\,\|y - y'\| = \|f\|\,\|y - z \;+\; z - y'\| \le \|f\|\bigl(\|y - z\| + \|y' - z\|\bigr).$$

This proves that $f \in Y^*$ can always be extended in a norm-preserving way to $Y_1 = \operatorname{ran}[z] + Y$, whatever z might be. Repetition of the argument will therefore prove existence of such a norm-preserving extension, one dimension at a time, on ever larger superspaces Y_n of Y. If the codimension of Y in X is finite, this will finish the argument. If it is countable (as it would be if X is separable), then complete induction will finish the argument. In the general case, **transfinite induction** is needed, i.e., Zorn's Lemma, or Hausdorff's Maximality Theorem. The latter asserts that any partially ordered set contains a maximal totally ordered subset. In our context, the partially ordered set in question is the collection G of all pairs (g, Y_g), with $g \in (Y_g)^*$ and $\|g\| \le \|f\|$ and $g|_Y = f$; in particular, $Y \subset Y_g$. The partial ordering is provided by

$$(g, Y_g) < (h, Y_h) := Y_g \subseteq Y_h \text{ and } h|_{Y_g} = g.$$

With Hausdorff's Maximality Theorem as our support, let H be a maximal totally ordered subset of G. Define $Z := \bigcup_H Y_h$. Then Z is a lss. Further, the definition

$$g : Z \to \mathbb{R} : x \mapsto hx \;\text{ if } x \in Y_h$$

is consistent since $hx = h'x$ whenever x also lies in $Y_{h'}$. It gives a norm-preserving extension g of f. If now $Z \ne X$, then the earlier argument would allow construction of a proper norm-preserving extension of g to a yet larger lss $Z_1 = \operatorname{ran}[z] + Z$ of X, and this would contradict the maximality of H. □

**** general HB ****

In the full generality, the Hahn-Banach Theorem is not constructive. I will give later (see (VI.18)) an instance of practical importance that indicates how one might construct such a norm-preserving extension. For that, I need to discuss convexity. In that connection I need the HB theorem in terms of a generalization of norm, i.e., in terms of a sublinear functional.

Definition. X ls. $p : X \to \mathbb{R}$ is **sublinear** :=

$$\forall\{x \in X, \alpha > 0\}\;\; p(\alpha x) = \alpha p(x) \qquad \textbf{(positive homogeneous)}$$
$$\forall\{x, y \in X\}\;\; p(x + y) \le p(x) + p(y) \qquad \textbf{(subadditive)}$$

For $f, g \in \mathbb{R}^X$, we say that f **is bounded by** g (in symbols: $f \le g$), in case $\forall\{x \in X\}\; f(x) \le g(x)$.

H.P.(11) Adapt the above proof to show the following

(29) General Hahn-Banach Theorem. *Let p be a sublinear fl on the ls X, Y a lss$\dot{T}$hen*

$$f \in Y', \ f \le p \quad \Longrightarrow \quad \exists\{g \in X'\} \ g|_Y = f \text{ and } g \le p.$$

H.P.(12) Let q be a **superlinear** functional on the nls X, i.e., q is positive homogeneous and **superadditive** (meaning that $\forall\{x, y \in X\} \ q(x+y) \ge q(x) + q(y)$).
(i) Prove that $p : x \mapsto \inf\{\|x+y\| - q(y) : y \in X\}$ is a sublinear functional, provided that $q \le \|\cdot\|$.
(ii) Where in the proof is the condition $q \le \|\cdot\|$ used?
(iii) Prove that, for any sublinear fl p, $p(0) = 0$ and $-p(-x) \le p(x)$.

** some uses of HB **

If Y is a lss of the nls X, then the HB theorem ensures that the map $X^* \to Y^* : \lambda \mapsto \lambda|_Y$ is always onto. However, this map is 1-1 iff $Y^- = X$. Hence, if $Y^- \ne X$, then Y^* and X^* are not obviously comparable.

The HB theorem ensures the existence of lfl's satisfying certain linear conditions (as given by the lss Y) and a certain consistent bound. Thus, in conjunction with representation theorems for X^*, it provides existence of solutions to certain (usually variational) problems, as illustrated in the H.P.(20). More explicitly, it says that there is $\lambda \in X^*$ satisfying $\lambda|_Y = \lambda_0$ and $\|\lambda\| \le r$ iff $\|\lambda_0\| \le r$. Here is a very simple example.

(30) Corollary. *For all x in the nls X, $\|x\| = \max_{\lambda \in X^*} |\lambda x|/\|\lambda\|$.*

Proof: Since $|\lambda x| \le \|\lambda\|\|x\|$, we have $\sup_{\lambda \in X^*} |\lambda x|/\|\lambda\| \le \|x\|$, and there is trivially equality here for every λ in case $x = 0$. So assume $x \ne 0$, and define $Y := \operatorname{ran}[x]$, and $f : Y \to \mathbb{R} : \alpha x \mapsto \alpha\|x\|$. Then $f \in Y' = Y^*$, and $fx = \|x\|$, and $1 = \|f\|$. By HB, we can find $g \in X^*$ s.t. $\|g\| = 1$ and $g|_Y = f$. Then $gx = \|x\| = \|g\| \, \|x\|$, i.e., $g \| x$. Therefore, $\sup_\lambda |\lambda x|/\|\lambda\| \ge |gx|/\|g\| = \|x\|$. □

The corollary implies that the **canonical embedding**

$$J : X \to X^{**} : x \mapsto (\lambda \mapsto \lambda x)$$

is an *isometry*, and each Jx takes on its norm. X is called **reflexive** in case J is onto. This is the case for any finite-dimensional X and also for $\mathbf{L}_p$ for $1 < p < \infty$, but not for $p = 1$ or $p = \infty$. The space c_0 is not reflexive, and neither is $C([a \,..\, b])$. Note that, by (3)Proposition, a reflexive space is necessarily complete.

It follows that, if X is reflexive, then every $\lambda \in X^{**}$ takes on its norm. But, since then also X^* is reflexive, we have

(31) Corollary. *If the nls X is reflexive, then every $\lambda \in X^*$ takes on its norm.*

The converse also holds, but is much harder to prove.

** the dual of a lm **

The (continuous) **dual** of $A \in bL(X, Y)$ is the lm

$$A^* : Y^* \to X^* : \lambda \mapsto \lambda A.$$

It is also bounded, in fact has the same norm as A:

(32) Corollary. $\|A^*\| = \|A\|$.

Proof: $\|A^*\| = \sup_{\lambda \in Y^*} \frac{\|\lambda A\|}{\|\lambda\|} = \sup_{x \in X, \lambda \in Y^*} \frac{|\lambda A x|}{\|\lambda\|\|x\|} = \sup_{x \in X} \frac{\|Ax\|}{\|x\|} = \|A\|$. □

Application: A sharp lower bound for $d(x,Y)$

In Approximation Theory and, more generally, in Optimization, linear functionals are used to characterize ba's, or minima of other functionals. These characterizations are based on the fact that there is a **dual problem** whose solution the minimum of the original problem serves, in turn, to characterize. Here is the simplest example (a dual version is given in (39)Corollary), the socalled

(33) Duality (in Approximation Theory). *For any lss Y of the nls X and any $x \in X$,*

$$d(x,Y) = \max_{0 \neq \lambda \perp Y} |\lambda x| / \|\lambda\|.$$

Proof: $\forall\{0 \neq \lambda \perp Y\}$ $Y \subseteq \ker\lambda$, so $d(x,Y) \geq d(x, \ker\lambda) = |\lambda x|/\|\lambda\|$, by (10)Lemma. So, $d(x,Y) \geq \sup_{\lambda \perp Y} |\lambda x|/\|\lambda\|$. Hence, if $d(x,Y) = 0$, then equality holds trivially.

In the contrary case, set $Y_1 := \operatorname{ran}[x] + Y$, $\lambda_0 : Y_1 \to \mathbb{R} : \alpha x + y \mapsto \alpha d(x,Y)$. Then $\|\lambda_0\| = \sup_{\alpha,y} |\alpha d(x,Y)|/\|\alpha x + y\| = \sup_y d(x,Y)/\|x - y\| = 1$. By HB, can extend λ_0 to $\lambda \in X^*$ with $\|\lambda\| = \|\lambda_0\| = 1$, thus $\lambda x/\|\lambda\| = \lambda x = \lambda_0 x = d(x,Y)$ and $\lambda \perp Y$, showing that the sup is taken on and equals $d(x,Y)$. □

H.P.(13) Show that, with $x \in X := C([a\,..\,b])$ and $t_0, \ldots, t_n$ arbitrary distinct points in $[a\,..\,b]$, $d(x, \Pi_{<n}) \geq |\delta_{t_0,\ldots,t_n} x| / \sum_i \prod_{j \neq i} |t_i - t_j|^{-1}$. (See (I.47) for the definition of the divided difference $\delta_{t_0,\ldots,t_n}$.)

For $Y \subseteq X$ and $M \subseteq X^*$, I use the abbreviations

$$Y^\perp := (\perp Y) \cap X^*, \qquad M_\perp := M\perp.$$

In these terms, (33) says that $d(x,Y)$ is the norm of x as a linear functional on $Y^\perp$, i.e., $d(x,Y) = \|(Jx)|_{Y^\perp}\|$; see also (39)Corollary.

(34) Corollary. *Let Y be a lss of the nls X, and $x \in X$. Then, $x \in Y^- \iff \forall\{\lambda \perp Y\}\ \lambda x = 0$; i.e.*

$$Y^- = \bigcap_{\lambda \perp Y} \ker\lambda \;=\; \big(Y^\perp\big)_\perp. \tag{35}$$

In particular, we get the

(36) Corollary. *The lss Y of the nls X is dense in X $\iff$ $Y^\perp = \{0\}$.*

H.P.(14)
- (i) Give an example of a complete nls X and a lss L of X^* for which $L^- \neq (L_\perp)^\perp$. (Hint: $X = C([0\,..\,1])$, $L = \operatorname{ran}[\delta_t : t > 0]$, and cf. H.P.(7).)
- (ii) Show that (39)Corollary need not hold for an arbitrary lss L of X^*.

H.P.(15) Let $A \in bL(X,Y)$. Prove:
- (i) $\ker A = (\operatorname{ran} A^*)_\perp$, but $\operatorname{ran} A \subseteq (\ker A^*)_\perp$ *with equality iff* $\operatorname{ran} A$ *is closed.*
- (ii) $\ker A^* = (\operatorname{ran} A)^\perp$, but $\operatorname{ran} A^* \subseteq (\ker A)^\perp$ *with equality only if* $\operatorname{ran} A^*$ *is closed.*

H.P.(16) Give an example of Bs's X, Y and $A \in bL(X,Y)$ for which $(\ker A)^\perp \backslash (\operatorname{ran} A^*)^- \neq \{\}$. (Hint: Consider $A : c[0\,..\,1] \to \ell_1 : f \mapsto (f(x(n))/n^2 : n \in \mathbb{N})$ with $(x(n) : n \in \mathbb{N})$ an enumeration of the rationals in $(0\,..\,1]$.)

Remark. See (VI.29) for a proof that $\operatorname{ran} A^* = (\ker A)^\perp$ in case $\operatorname{ran} A^*$ is closed.

H.P.(17) Let X be a nls of sequences containing $\mathrm{ran}[e_1, e_2, \ldots]$. (a) Prove: *If the action of the continuous dual of X can be represented as scalar product with certain sequences, then* $\mathrm{ran}[e_1, e_2, \ldots]$ *is dense.* (b) Prove that the converse holds in case the norm on X is **monotone** in the sense that $|x| \le |y|$ implies $\|x\| \le \|y\|$.

** bounded linear projectors and their norm **

(37) Corollary (to HB). *Every finite-dimensional lss Y of a real nls X is the range of some bounded linear projector.*

Proof: If $\dim Y = n$, then there exists $V \in L(\mathbb{R}^n, Y)$ invertible, hence $V^{-1} \in L(Y, \mathbb{R}^n) = bL(Y, \mathbb{R}^n)$, i.e., $V^{-1} =: [\mu_1, \ldots, \mu_n]^t$ for some (μ_j) in $Y' = Y^*$ (these are the coordinate fl's for the basis V; cf. Chapter I). By HB, we can extend μ_i to some $\lambda_i \in X^*$. Then $\Lambda^t = [\lambda_1, \ldots, \lambda_n]^t \in bL(X, \mathbb{R}^n)$ and $\Lambda^t V = 1$, hence $P := V\Lambda^t$ is a bounded lprojector with $\mathrm{ran}\, P = \mathrm{ran}\, V = Y$. □

H.P.(18) Prove: *Every n-dimensional lss of an nls X is the range of a lprojector with norm $\le n$.* (Hint: (13)Auerbach's Theorem)

Here is a formula for the norm of a linear projector.

(38) Proposition. *If P is the lprojector given by F and L, with F a finite-dimensional lss of the nls X and L a lss of X^*, then*

$$\|P\| = \max_{f \in F} \min_{\lambda \in L} \frac{\|f\| \|\lambda\|}{|\lambda f|}.$$

Proof: By H.P.(III.17),

$$\|P\| = \sup_{x \notin \ker P} \|Px\| / d(x, \ker P).$$

By duality, $d(x, \ker P) = \max_{\lambda \perp \ker P} \frac{|\lambda x|}{\|\lambda\|}$, while $\lambda \perp \ker P$ iff $\ker \lambda \supseteq L\perp$ iff $\lambda \in L$, the last by (I.31)Lemma. Thus

$$\|P\| = \sup_x \min_{\lambda \in L} \frac{\|Px\| \|\lambda\|}{|\lambda x|}.$$

But, for $\lambda \in L$, $\lambda x = \lambda P x$, i.e., it is sufficient to take the sup over $x \in \mathrm{ran}\, P = F$. □

H.P.(19) Let P be a lprojector in some nls X, let $V \in bL(\ell_\infty(n), \mathrm{ran}\, P)$ be a basis for its range, and let $\Lambda \in bL(\ell_1(n), \mathrm{ran}\, P' \subset X^*)$ be a basis for the space $\mathrm{ran}\, P'$ of its interpolation functionals. Prove that $\|P\|$ is in the interval $\|(\Lambda^t V)^{-1}\|_\infty [1/(\|V^{-1}\| \|\Lambda^{-1}\|) \,..\, \|V\| \|\Lambda\|)$.

Remark. If $X = C(T)$, then, for any $A \in L(X)$, $\|A\| = \|\ell_A\|_\infty$, with $\ell_A : t \mapsto \|\delta_t A\|$. This is often a better starting point for computing the norm of a lprojector on $C(T)$ than is (38)Proposition.

** norm of a lfl on a lss **

The HB Theorem ensures that, for any lss Y of the nls X and any $\nu \in Y^*$,

$$\|\nu\| = \min\{\|\lambda\| : \lambda \in X^*, \lambda|_Y = \nu\}.$$

This is often the only way to compute $\|\nu\|$, provided a ready representation for X^* makes it easy to compute $\|\lambda\|$ for $\lambda \in X^*$.

(39) Corollary (to HB). *If X is a nls, $\nu \in X^*$, and L a finite-dimensional lss of X^*, then $\|\nu|_{\ker L}\| = d(\nu, L)$.*

Proof: Set $Y := \ker L = L_\perp$. By HB, $\|\nu|_Y\| = \min\{\|\lambda\| : \lambda \in X^*, \lambda|_Y = \nu|_Y\}$, while $\lambda|_Y = \nu|_Y$ iff $\lambda - \nu \in Y^\perp = (L_\perp)^\perp = L$, the last equality by (I.31)Lemma since $\dim L < \infty$. □

H.P.(20) Use (39) to determine the maximum of $\int_0^1 x f(x)\,dx$ over all measurable f with $\int_0^1 f(x)\,dx = 0$ and $\int_0^1 |f(x)|^2\,dx = 1$.

** GW-interval is sharp **

We now return to the Optimal Recovery problem of providing all the information we can about μg, given that $\mu \in X^*$, $\|g\| \le r$, and $\Lambda^t g = a$ for some 1-1 $\Lambda \in L(\mathbb{R}^m, X^*)$ and $a \in \mathbb{R}^m$. We established that μg must lie in the GW-interval given in (22), i.e.,

$$\sup_{\lambda \in \operatorname{ran}\Lambda} (\lambda g - \|\mu - \lambda\| r) \le \mu g \le \inf_{\lambda \in \operatorname{ran}\Lambda} (\lambda g + \|\mu - \lambda\| r).$$

But we have yet to show that each s in this interval is actually the value of μg for some such g.

This will certainly be so in case $\inf\{\|g\| : [\Lambda, \mu]^t g = (a, s)\} < r$ since then there will even be such g with $\|g\| < r$. To investigate further, recall that Λ is 1-1, and assume that $\mu \notin \operatorname{ran}\Lambda$ (there being nothing to prove otherwise). Then $[\Lambda, \mu]$ is 1-1, hence $[\Lambda, \mu]^t$ is onto, hence we may (and do) pick g_0 such that $[\Lambda, \mu] g_0 = (a, s)$, and compute

$$\inf\{\|g\| : [\Lambda, \mu]^t g = (a, s)\} = \inf\{\|g_0 - h\| : [\Lambda, \mu]^t h = 0\} = d(g_0, \ker[\Lambda, \mu]^t),$$

while, by (33)Duality,

$$d(g_0, \ker[\Lambda, \mu]^t) = \max_{\lambda \perp \ker[\Lambda, \mu]^t} \frac{|\lambda g_0|}{\|\lambda\|} = \max_{\lambda \in \operatorname{ran}[\Lambda, \mu]} \frac{|\lambda g_0|}{\|\lambda\|},$$

the last equation because $\lambda \perp \ker[\Lambda, \mu]$ iff $\lambda \in \operatorname{ran}[\Lambda, \mu]$, by (I.31)Lemma. We conclude that

$$\inf\{\|g\| : [\Lambda, \mu]^t g = (a, s)\} = \|M_s\|,$$

with M_s the lfl

$$M_s : \operatorname{ran}[\Lambda, \mu] \to \mathbb{R} : \Lambda c + \alpha\mu \mapsto (\Lambda c + \alpha\mu) g_0 = c^t a + \alpha s.$$

Recalling the proof of (27)HB, we recognize that $\|M_s\| \le r$ since s lies in the GW-interval. We conclude that *for any $\varepsilon > 0$, $B_{r+\varepsilon}$ intersects the 'feasible set'* $([\Lambda, \mu]^t)^{-1}\{(a, s)\}$. It is in this sense that the GW-interval is sharp.

If $\|M_s\| < r$, then even B_r intersects the 'feasible set', i.e., there even exists $g \in X$ with $\|g\| < r$ and $\Lambda^t g = a$ and $\mu g = s$. But if $\|M_s\| = r$, then the 'feasible set' may not contain an element of norm $\le r$ (even if X is complete, cf. H.P.(8)), unless the situation is special. E.g., we do know by HB that X^{**} contains some M with $\|M\| = \|M_s\|$ and $M = M_s$ on $\operatorname{ran}[\Lambda, \mu]$. Therefore, if X is reflexive, then $J^{-1} M$ is an element in the 'feasible set' and of norm $\|M\| = \|M_s\| \le r$.

For the record, here is a simple version of the main point just established.

(40) Proposition. *Let X be a nls, $\Lambda \in L(\mathbb{R}^n, X^*)$, $g_0 \in (\Lambda^t)^{-1}\{a\}$ for some $a \in \mathbb{R}^n$, and $M_a : \operatorname{ran}\Lambda \to \mathbb{F} : \lambda \mapsto \lambda g_0$. Then* $\inf \|(\Lambda^t)^{-1}\{a\}\| := \inf_{g\in(\Lambda^t)^{-1}\{a\}} \|g\| = \|M_a\|$.

H.P.(21) Prove: *The image $J(B_X)$ of the unit ball of X in X^{**} is w*-dense in the unit ball $B_{X^{**}}$ of the bidual of X, i.e., dense with respect to the topology of pointwise convergence (on their common domain, i.e., on X^*).* (Hint: Recall that the typical w*-neighborhood of $h \in X^{**}$ (i.e., neighborhood in the topology of pointwise convergence on X^*) is of the form $B_{r,L}(h) = \{k \in X^{**} : \max_{\lambda\in L} |(h-k)\lambda| < r\}$ for some finite subset L of X^* and some positive r, hence is certain to contain all $k \in X^{**}$ that agree with h on $\operatorname{ran}[L]$, and apply (40)).

H.P.(22) Give an example of a complete nls X and some finite-rank data map Λ^t, so that, for some choice of a, r, and $\mu \in X^*$, the GW-interval $GW(\Lambda^t, a, r, \mu)$ is an interval of positive length, yet for none of its points s is there $g \in B_r^-$ with $\Lambda^t g = a$ and $\mu g = s$. (Hint: Perhaps a variant of (III.13)Example could work here?)

