VII. Inner Product Spaces

An **inner product space** (=: **ips**) is a nls in which the norm derives from an **inner product**, i.e., from a hermitian positive definite form. (Here, **form** is used in the sense of **functional**, i.e., scalar-valued map, particularly a scalar-valued map on the cartesian product of the space with itself.) In such a space, one has a notion of **angle** in addition to (translation- and scale-invariant) distance and so is much closer to the familiar Euclidean n-space $\ell_2(n)$ than in other nls's.

We have occasion to use the *complex* scalar field $\mathbb{C}$. Recall the agreement that $\mathbb{F}$ stands for either $\mathbb{R}$ or $\mathbb{C}$.

** definition **

X ls. $X \times X \to \mathbb{F} : (x, y) \mapsto \langle x, y\rangle$ is an **inner product** :=

$\forall\{y \in X\}$	$\langle\cdot, y\rangle$ is linear	**(linearity)**
$\forall\{x \in X\backslash 0\}$	$\langle x, x\rangle > 0$	**(positive definite)**
$\forall\{x, y \in X\}$	$\langle y, x\rangle = \overline{\langle x, y\rangle}$	**(skew-symmetric or hermitian)**

The bar denotes formation of the complex conjugate. If $\mathbb{F} = \mathbb{R}$, then the skew-symmetry becomes just symmetry, $\langle y, x\rangle = \langle x, y\rangle$.

Note that $\langle x, \cdot\rangle$ is **skew-linear**, i.e., additive but skew homogeneous, since

$$\langle x, y + z\rangle = \overline{\langle y + z, x\rangle} = \overline{\langle y, x\rangle} + \overline{\langle z, x\rangle} = \langle x, y\rangle + \langle x, z\rangle$$

but

$$\langle x, \alpha y\rangle = \overline{\langle \alpha y, x\rangle} = \bar{\alpha}\overline{\langle y, x\rangle} = \bar{\alpha}\langle x, y\rangle.$$

The model *example* is $X = \mathbb{F}^n$ with $\langle , \rangle$ the scalar product, i.e.,

$$\langle x, y\rangle := y^c x := \sum_i x(i)\overline{y(i)} \quad = x^t\overline{y} = \overline{y}^t x.$$

(Unfortunately, it is traditional to have the inner product skewlinear in the second slot rather than the first, therefore the switch in the order here.) The corresponding continuous *example* is: X some ls of functions on some $T \subset \mathbb{R}^m$ with the inner product given by the integral

$$\langle f, g\rangle := \int_T f(t)\overline{g(t)}\,\mathrm{d}t.$$

** bilinear forms **

Each lm A on the ips X into X gives rise to a (skew)bilinear form by the rule

$$\langle , \rangle_A : X \times X \to \mathbb{F} : (x, y) \mapsto \langle Ax, y\rangle,$$

and A is called hermitian or positive definite or whatever in case $\langle , \rangle_A$ is hermitian or positive definite or whatever. In particular, arguments concerning inner products are applicable to a more general bilinear form $\langle , \rangle_A$ to the extent that it shares the properties of the ip $\langle , \rangle$ used in those arguments.

H.P.(1)

(i) Prove: *If X ips over $\mathbb{F} = \mathbb{C}$, $A \in L(X)$, $|x|_A := \langle Ax, x\rangle$, then*

$$4\langle Ax, y\rangle = \sum_{i=1}^{4} \mathrm{i}^i |x + \mathrm{i}^i y|_A.$$

This means that the bilinear form $\langle , \rangle_A$ can be reconstructed from its values on the "diagonal", i.e., on $\{(x,x) : x \in X\}$. This is called **polarization**.

(ii) Prove: *If $\mathbb{F} = \mathbb{C}$ and $A \in L(X)$ is positive definite, then $\langle , \rangle_A$ is an ip.* Explain why this conclusion cannot be drawn when $\mathbb{F} = \mathbb{R}$.

** ips is nls **

The "norm" on the ips X is defined by

$$\|x\| := \langle x, x\rangle^{1/2}.$$

It is positive definite and positive homogeneous, by inspection. A proof of the triangle inequality can be obtained along the following lines. Compute

$$\|x \pm y\|^2 = \langle x \pm y, x \pm y\rangle = \langle x, x\rangle \pm \langle x, y\rangle \pm \langle y, x\rangle + \langle y, y\rangle,$$

using the bi-additivity of $\langle , \rangle$. Therefore

$$\|x \pm y\|^2 = \|x\|^2 \pm 2\,\mathrm{Re}\langle x, y\rangle + \|y\|^2. \tag{1}$$

Taking a cue from $\ell_2(n)$, one calls $x, y \in X$ **orthogonal** and writes

$$x \perp y$$

in case $\langle x, y\rangle = 0$. This gives

(2) Pythagoras. $x \perp y \Longrightarrow \|x \pm y\|^2 = \|x\|^2 + \|y\|^2$.

Perhaps the most important aspect of an ips X is the possibility of using the elements of X to represent linear functionals on X. This possibility exists since

$$\forall \{y \in X\} \quad y^t \ := \ \langle \cdot, y\rangle \ \in \ X'.$$

Note that $\ker y^t = \{x \in X : x \perp y\} =: \{y\}\perp$, the **orthogonal complement** of y. Also, for $y \neq 0$, $y^c y = \langle y, y\rangle \neq 0$. Therefore (recall (IV.5)elimination), for $y \neq 0$,

$$\forall \{x \in X\} \quad \big(x - \frac{\langle x, y\rangle}{\langle y, y\rangle} y\big) \perp y.$$

So, by Pythagoras,

$$\|x\|^2 = \|x - \frac{\langle x, y\rangle}{\langle y, y\rangle} y\|^2 + \|\frac{\langle x, y\rangle}{\langle y, y\rangle} y\|^2 \ \geq \ |\frac{\langle x, y\rangle}{\langle y, y\rangle}|^2 \|y\|^2 = |\langle x, y\rangle|^2 / \|y\|^2$$

with equality iff $x = \frac{\langle x,y\rangle}{\langle y,y\rangle} y$. This proves

(3) CBS (Cauchy-Bunyakovski-Schwarz). $|\langle x, y\rangle| \le \|x\|\|y\|$ *with equality iff* $[x, y]$ *is not 1-1.*

As a consequence of the CBS inequality, $\|y^c\| = \|y\|$, hence *the map* $y \mapsto y^c$ *embeds* X *isometrically, but only skewlinearly, in* X^*. (Since X^* is complete (being a dual space), this map cannot be onto unless X is complete. The Riesz-Fischer Theorem below says that the map is onto in that case.) Also, $y^c || y$ (for $y \ne 0$). Further, with (1),

$$\|x + y\|^2 \le \|x\|^2 + 2\|x\|\|y\| + \|y\|^2.$$

This proves the **triangle inequality**

$$\|x + y\| \le \|x\| + \|y\|,$$

and so finishes the proof of the fact that *an ips is a nls wrto the norm* $\|\cdot\| := \langle\cdot,\cdot\rangle^{1/2}$. This nls is called **Hilbert space** (=: **Hs**) in case it is complete.

H.P.(2) Prove that the CBS inequality holds (though equality is more complicated) even when X is only a **semi-inner product space**, meaning that $\langle,\rangle$ is only positive semidefinite, i.e., still $\|x\| \ge 0$ for all $x \in X$, but equality is possible even for some nonzero x. (Hint: Show that only the case $\|x\| = 0 = \|y\|$ needs to be considered; then consider it.)

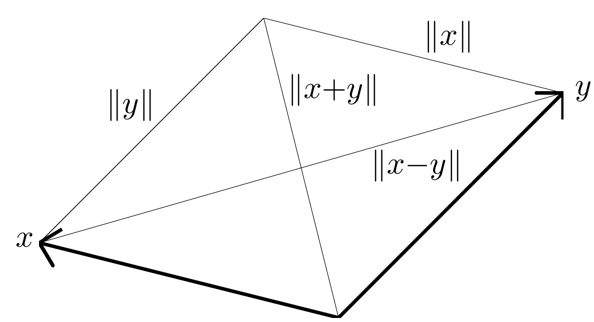
H.P.(3) Prove that, for any $A \in BL(X)$, with X ips, $\sup_{\|x\|,\|y\|\le 1} |\langle Ax, y\rangle| = \|A\|$.

H.P.(4) (a) Prove that *any weakly convergent sequence in an ips converges in norm iff its norms converge to the norm of its weak limit.* (b) Show that the sequence (e_n) in ℓ_2 converges weakly, but not strongly, i.e., not in norm, to 0.

** parallelogram law **

(1) implies the

(4) Parallelogram Law. $\|x + y\|^2 + \|x - y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big)$.



(5) Figure. Parallelogram law

Von Neumann showed that this law is *characteristic* of a norm derived from an ip, i.e., *the parallelogram law implies that*

$$(x, y) \mapsto \sum_{i=1}^{4} \mathrm{i}^i \|x + \mathrm{i}^i y\|^2/4$$

is an inner product and it gives rise to the norm from which we started. (Here, the two purely imaginary terms are omitted in case $\mathbb{F} = \mathbb{R}$.)

H.P.(5) Prove von Neumann's assertion. (Hint: additivity of $\langle x, \cdot\rangle$ follows by two applications of the Parallelogram Law, whence (with some H.P. from Chapter I) homogeneity wrto rationals; now use continuity and effect of substitution of iy for y; etc.)

(6) Corollary 1. *An ips is strictly convex.*

Proof: $\|x\| = \|y\| = 1 = \|(x+y)/2\| \implies 4 = 4\|(x+y)/2\|^2 = \|x+y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big) - \|x-y\|^2 = 4 - \|x-y\|^2$, i.e., $x = y$. □

**** ba from convex sets ****

(7) Corollary 2. *If K is a complete convex subset of the ips X, then every $x \in X$ has exactly one ba from K.*

Proof: After a shift by x, we may assume that $x = 0$. Let (k_n) be a **minimizing** sequence in K, i.e., $\lim\|k_n\| = d(0,K)$. By the Parallelogram law,

$$\begin{aligned} 0 \le \|k_n - k_m\|^2 &= 2\big(\|k_n\|^2 + \|k_m\|^2\big) - 4\|(k_n+k_m)/2\|^2 \\ &\le 2\big(\|k_n\|^2 + \|k_m\|^2\big) - 4d(0,K)^2 \xrightarrow[n,m\to\infty]{} 0 \end{aligned}$$

since $(k_n + k_m)/2 \in K$ and $\|k_n\|^2 \to d(0,K)^2$. This shows (k_n) to be Cauchy. Since K is complete, this gives $\lim k_n = k$ for some $k \in K$, while $\|k\| = \lim\|k_n\| = d(0,K)$.

This proves existence of a ba. Uniqueness follows from the strict convexity or directly from the above argument, by replacing k_m by k'_n, with (k'_n) any other minimizing sequence. □

Thus, if K is convex and complete, then the rule

$$x \mapsto \text{ ba to } x \text{ from } K$$

defines a map, called the **orthogonal projector** onto K and denoted by

$$P_K.$$

Note that any closed subset of a *finite-dimensional* lss of a nls is complete.

**** characterization of ba from lss ****

The characterization of $P_K x$ is particularly striking when K is a lss, Y. While the characterization: $y = P_Y x \iff x - y \perp Y$ can be proved directly from Pythagoras (see (10)Figure), it is instructive to derive it from general principles as follows.

(8) Proposition. *Let X ips, $z \in X\backslash 0$, $\lambda \in X^*\backslash 0$. Then:* $\lambda || z \iff \lambda = \frac{\|\lambda\|}{\|z\|} z^c$.

Proof: Indeed, if $\lambda = (\|\lambda\|/\|z\|)z^c$, then $\lambda z = \|\lambda\|\|z\|$, i.e., $\lambda||z$. Conversely, if $\lambda z = \|\lambda\|\|z\|$, then, as $|\lambda z| = \|\lambda\| d(z, \ker\lambda)$ by (IV.10)Lemma, therefore $\|z\| = d(z, \ker\lambda)$. So, for any $x \in X$, $x - (\lambda x/\lambda z)z \in \ker\lambda$, and, with this,

$$\|z\|^2 \le \|z - (x - (\lambda x/\lambda z)z)\|^2 = \|z\|^2 - 2\,\mathrm{Re}\langle x - (\lambda x/\lambda z)z, z\rangle + \|x - (\lambda x/\lambda z)z\|^2.$$

This implies that

$$2\,\mathrm{Re}\langle x - (\lambda x/\lambda z)z, z\rangle \le \|x - (\lambda x/\lambda z)z\|^2$$

for all $x \in X$. But since the left-hand side is linear in x, while the right-hand side is quadratic in x, this cannot hold unless the left-hand side vanishes identically. Precisely, by substituting $\mathrm{signum}\langle x - (\lambda x/\lambda z)z, z\rangle \alpha x$ for x, we find that

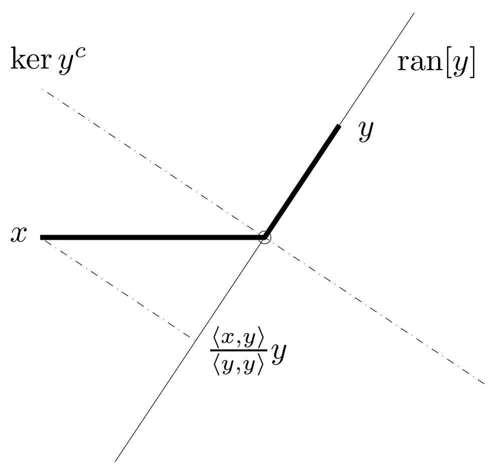
$$\alpha|2\langle x - (\lambda x/\lambda z)z, z\rangle| \le \alpha^2\|x - (\lambda x/\lambda z)z\|^2, \quad \text{all } \alpha \ge 0,$$

and this implies that $|\langle x - (\lambda x/\lambda z)z, z\rangle| = 0$, hence $\langle x, z\rangle = \lambda x\,(\langle z, z\rangle/\lambda z) = \lambda x(\|z\|/\|\lambda\|)$ for all $x \in X$. □

(9) Corollary. *If Y is a lss of the ips X and $x \in X$ and $y \in Y$, then $y = P_Y x$ iff $x - y \perp Y$.*

Proof: By (VI.21)Theorem, $y = P_Y x \iff \exists\{\lambda \in X^*\}\ Y \perp \lambda || x - y$, while by the proposition, $\lambda || x - y$ iff $\ker \lambda = \ker(x - y)^c = \{x - y\}\perp$. □

H.P.(6) Prove (9) directly from Pythagoras along the lines suggested by (10)Figure.



(10) Figure. Orthogonal projection onto $\mathrm{ran}[y]$

P_Y is defined if Y is complete, in particular, if $\dim Y < \infty$. We conclude that, in that case, P_Y *provides the solution to the* $LIP(Y, Y^c)$, with

$$Y^c := \{y^c : y \in Y\}.$$

In particular, P_Y is a *linear* projector, and

$$X = \mathrm{ran}\, P_Y \oplus \ker P_Y = Y \oplus (Y\perp)$$

with

$$Y\perp := {Y^c}_\perp = \bigcap_{y \in Y} \ker y^c = \{x \in X : \forall\{y \in Y\}\ y \perp x\}$$

the **orthogonal complement** of the set Y. Further, $1 - P_Y = P_{Y\perp}$.

H.P.(7) Let $(Y_n)_{n=-\infty}^{\infty}$ be a nested sequence of closed linear subspaces in a Hs X, and set $Y_{-\infty} := \cap_n Y_n$, $Y_\infty := (\cup_n Y_n)^-$. Prove that P_{Y_n} converges pointwise to $P_{Y_{\pm\infty}}$ as $n \to \pm\infty$. (Hints: if $Y \subset Z$, then $\|P_Y x\| \le \|P_Z x\|$ and, for any $y \in Y$, $\langle P_Z x - P_Y x, y\rangle = \langle P_Z x - x, y\rangle - \langle P_Y x - x, y\rangle = 0$. Also, it is sufficient to prove that P_{Y_n} converges pointwise to P_{Y_∞}.)

** numerics **

For $n = \dim Y < \infty$, the characterization of the ba leads to the socalled **normal equations**: With $V = [v_1, v_2, \ldots, v_n]$ any basis for Y, we have $P_Y x = Va$ iff

$$\sum_{j=1}^{n} \langle v_j, v_i\rangle a(j) = \langle x, v_i\rangle, \qquad i = 1, \ldots, n.$$

In the language developed in Chapter 1 (see (I.48)), $P_Y x$ is the unique interpolant in $Y = \operatorname{ran} V$ to x with respect to the row map

$$V^c : x \mapsto (\langle x, v_i\rangle : i = 1, \ldots, n).$$

In particular, $P_Y = V(V^cV)^{-1}V^c$ with the **Gramian** $V^cV = (\langle v_j, v_i\rangle)$ invertible since it is square and 1-1 ($V^cVa = 0$ implies $0 = a^cV^cVa = \langle Va, Va\rangle = \|Va\|^2$, hence $Va = 0$, hence $a = 0$). As discussed in Chapter 1, the basis V must be chosen with care, so as to make the Gramian well-conditioned. A particularly happy choice is to make the basis **orthogonal**, i.e., so that the Gramian is diagonal. For such a basis,

$$P_Y x = V(V^cV)^{-1}V^c x = \sum_1^n v_j \frac{\langle x, v_j\rangle}{\langle v_j, v_j\rangle}.$$

(Neat minds, in fact, require an ortho**normal** basis, i.e., one for which the Gramian is 1. This requires taking square roots, and who wants to do that?) In principle, an orthogonal basis V can be constructed from any old basis $Z =: [z_1, z_2, \ldots, z_n]$ by a bootstrap procedure called **Gram-Schmidt-orthogonalization** (see Chapter 1), in which v_i is obtained as the error in the ba to z_i from $\operatorname{ran}[z_1, z_2, \ldots, z_{i-1}] = \operatorname{ran}[v_1, v_2, \ldots, v_{i-1}]$, hence v_i is orthogonal to v_j for $j < i$:

$$v_i := z_i - \sum_{j<i} v_j \frac{\langle z_i, v_j\rangle}{\langle v_j, v_j\rangle}, \quad i = 1, 2, \ldots, n.$$

But, if the basis Z is badly conditioned, the basis V so computed may fail to be close to orthogonal.

H.P.(8) A notorious example of the limitations of Gram-Schmidt in the face of an ill-conditioned basis Z is provided by the choice $Y = \Pi_k$, $\langle f, g\rangle := \int_a^b f(t)g(t)\,dt$, for $0 \ll a < b$, with Z the power basis, i.e., $z_j := ()^{j-1}$, all j. Carry out Gram-Schmidt for this example numerically, choosing, e.g., to approximate $\langle f, g\rangle = \int_9^{10} f(t)g(t)\,dt$ by $\sum_i f(t_i)g(t_i)w(i)$ with $t_i = 9 + (i-1)/20$, $i = 1, \ldots, 21$, and choosing $k = 10$, and check just how orthogonal the resulting basis V is. The trouble will be particularly apparent if you choose to express the v_j in power form.

**** Riesz-Fischer ****

Of particular interest is ba from $\ker\lambda$ for $\lambda \in X^*\backslash 0$ to some $x \notin \ker\lambda$. If y is a ba to x from $\ker\lambda$, then, by (9)Corollary, $\ker(x-y)^c \supset \ker\lambda$, i.e. (directly, or by (I.31)Lemma), $\lambda = \alpha(x-y)^c$ for some scalar α. Hence if X is complete, and therefore $\ker\lambda$ is complete for any $\lambda \in X^*$, then, by (7)Corollary, $\ker\lambda$ provides ba's, and thus λ can be written as x^c for some $x \in X$. This completes the proof of

(11) Riesz-Fischer Theorem. *If X is Hs and $\lambda \in X'$, then*

$$\lambda \in X^* \quad \Longleftrightarrow \quad \lambda = x^c \textit{ for some } x \in X.$$

Thus, the isometry $X \to X^* : y \mapsto y^c$ is onto if and only if X is complete. In particular, we can think of X^* as the completion of X. I will use λ^{-c} for the unique $x \in X$ that represents $\lambda \in X^*$ in this way, i.e., $\lambda =: \langle\cdot, \lambda^{-c}\rangle$.

An Application: Optimal interpolation

The **optimal** interpolant to an x in the nls X with respect to the data map $\Lambda^t \in bL(X, \mathbb{F}^n)$ is, by definition, any $f \in X$ of smallest norm among all f that agree with x on Λ^t, i.e., any element in

$$\operatorname{argmin}\{\|f\| : \Lambda^t f = \Lambda^t x\}.$$

Since $\{f \in X : \Lambda^t f = \Lambda^t x\} = x + \ker\Lambda^t = x + (\operatorname{ran}\Lambda)_\perp$ does not depend on the columns of Λ but only on its range, we may assume wlog that $\Lambda =: [\lambda_1, \ldots, \lambda_n]$ is 1-1.

Now assume that X is a Hs. Then

$$\Lambda^t = V^c : z \mapsto (\langle z, v_i\rangle : i = 1, \ldots, n),$$

with $v_i := {\lambda_i}^{-c}$, all i, and $V := [v_1, \ldots, v_n]$ 1-1 since $x \mapsto x^c$ is a skewlinear isometry. Hence, $P := V(\Lambda^t V)^{-1}\Lambda^t = P_{\operatorname{ran} V}$ is the orthoprojector onto $\operatorname{ran} V$ and, in particular, $\ker\Lambda^t = \ker P \perp \operatorname{ran} P = \operatorname{ran} V$. This implies that

$$\|Px + h\|^2 = \|Px\|^2 + \|h\|^2, \quad \forall h \in \ker\Lambda^t.$$

Since $\{f \in X : \Lambda^t f = \Lambda^t x\} = x + \ker\Lambda^t = Px + \ker\Lambda^t$, this shows that Px is the sole element of $\operatorname{argmin}\{\|f\| : \Lambda^t f = \Lambda^t x\}$.

H.P.(9) Let $\Xi \in \mathbb{F}^{m\times n}$ with $m \le n$ be of full rank. Prove that $t = \Xi^*(\Xi\Xi^t)^{-1}b$ is the unique minimum 2-norm solution of the equation $\Xi ? = b$. Also prove that $t(i) = t(j)$ in case $\Xi(:,i) = \Xi(:,j)$.

For practical work with such an **optimal** interpolant for given Λ, one does need the means for finding the representers v_i for the λ_i.

(12) Example Consider the problem of finding

$$\operatorname{argmin}\{\int_0^1 |Df(s)|^2 \, \mathrm{d}s : f|_T = g|_T\}$$

for some fixed $g \in C^{(1)}[0\,..\,1]$ and some finite set $T \subset [0\,..\,1]$ that we assume to contain the point 0, for convenience. Then this is the same as finding

$$\operatorname{argmin}\{\|f\|^2 : f|_T = g|_T\}$$

with

$$\|f\|^2 := \langle f, f\rangle, \qquad \langle f, h\rangle := f(0)h(0) + \int_0^1 Df(s)Dh(s)\,\mathrm{d}s.$$

Note that $X := C^{(1)}[0\,..\,1]$ is an ips with respect to this (with only the definiteness needing some check: if $\langle f, f\rangle = 0$, then $\int (Df(s))^2\,\mathrm{d}s = 0$, hence $Df = 0$, and also $f(0) = 0$, therefore $f = 0$). Further (as in (IV.24)Example), for any $t \in [0\,..\,1]$,

$$\delta_t f = f(0) + \int_0^t Df(s)\,\mathrm{d}s = \langle f, v_t\rangle,$$

with

$$v_t(u) := 1 + \int_0^u (t-s)_+^0\,\mathrm{d}s = 1 - (t-u)_+ + t = 1 + u - (u-t)_+.$$

Note that v_t fails to be in $C^{(1)}[0\,..\,1]$, but is continuous with a piecewise continuous first derivative, hence in X if we extend X to contain all such functions. In fact, the completion of $C^{(1)}[0\,..\,1]$ with respect to the given ip norm is the space

$$\mathbf{L}_2^{(1)}[0\,..\,1] := \{u \mapsto c + \int_0^u g(s)\,\mathrm{d}s : c \in \mathbb{R}, g \in \mathbf{L}_2[0\,..\,1]\}.$$

We conclude that $\operatorname{ran} V = \Pi_{1,T}^0 :=$ the space of continuous piecewise linear functions on $[0..1]$ with breaks at the points in T (and nowhere else). Consequently, among all functions $f \in \mathbf{L}_2^{(1)}[0\,..\,1]$ that agree with a given g at T, the broken-line interpolant to g uniquely minimizes $\int_0^1 |Df(s)|^2\,\mathrm{d}s$.

H.P.(10) Straighten out a final detail in the above example: What is the nature of functions in $\operatorname{ran} V$ on the interval $[\max T\,..\,1]$ in case that interval has interior?

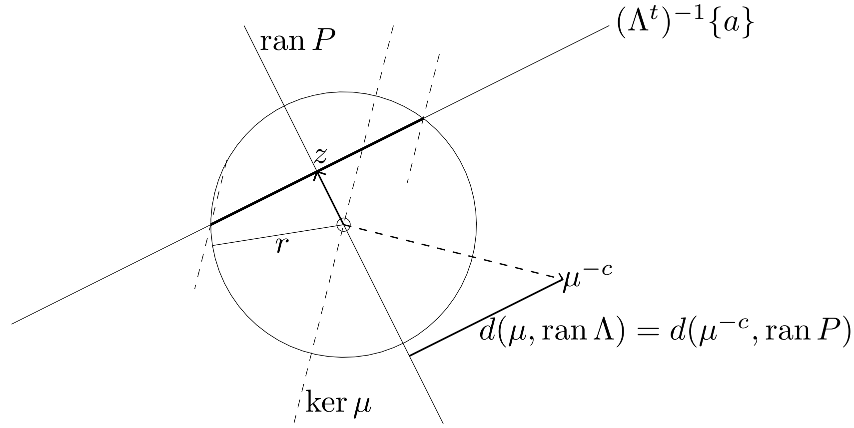
Application: Synge's hypercircle

In a Hs, the problem of optimal recovery has a particularly nice and simple answer. To recall from Chapter IV, this problem concerns sharp bounds for μg, given that $\mu \in X^*$ and that

$$g \in G := G(\Lambda, a, r) := \{g \in X : \Lambda^t g = a, \|g\| \le r\}$$

for given $\Lambda^t \in bL(X, \mathbb{F}^m)$, $a \in \mathbb{F}^m$, and r. Synge called $G(\Lambda, a, r)$ a **hypercircle** for obvious reasons (see (13)Figure).

Since X is a Hs, we can find $V \in bL(\mathbb{F}^n, X)$ so that $\Lambda^t = V^c$ and, from the preceding discussion on optimal interpolation, we know that the orthogonal projector $P := P_{\operatorname{ran} V}$ has the interpolation functionals $\operatorname{ran}\Lambda$. In particular, assuming that Λ^t is onto, hence $(\Lambda^t)^{-1}\{a\}$ is not empty, we know that P maps the entire flat $(\Lambda^t)^{-1}\{a\}$ to one point, viz.



(13) Figure. The hypercircle and the GW-interval

the point in $(\Lambda^t)^{-1}\{a\}$ closest to the origin. Call this point z. In particular the hypercircle $G(\Lambda, a, r)$ is not empty iff $\|z\| \le r$.

Further, $(\Lambda^t)^{-1}\{a\} = z + \ker P = z + \ker \Lambda^t$. Since $z \in \operatorname{ran} P \perp \ker \Lambda^t$, this implies that $G = B_r^- \cap (\Lambda^t)^{-1}\{a\} = z + B_s^- \cap \ker \Lambda^t$, with $s := \sqrt{r^2 - \|z\|^2}$. Further, $\mu(B_s^- \cap \ker \Lambda^t) = [-1\,..\,1]s\|\mu|_{\ker\Lambda^t}\| = [-1\,..\,1]sd(\mu, \operatorname{ran}\Lambda)$, the last equality from (IV.39)Corollary to HB. Therefore,

$$\mu G = \mu z + [-1\,..\,1]d(\mu, \operatorname{ran}\Lambda)\sqrt{r^2 - \|z\|^2}.$$

Note that $\mu z = \mu P g$ for any $g \in (\Lambda^t)^{-1}\{a\}$, and that μP is the ba to μ from $\operatorname{ran} P' = \operatorname{ran}\Lambda$. In particular, μP is Sard's best rule $\overset{\mathfrak{M}}{\lambda}$ from $\operatorname{ran}\Lambda$ for μ. Thus, in a Hs, Sard's best rule provides the center of the GW-interval, and the radius of the GW-interval is the number $d(\mu, \operatorname{ran}\Lambda)\sqrt{r^2 - \|z\|^2}$, and it is entirely computable from the data Λ, a, and r, and is usually smaller than the radius $d(\mu, \operatorname{ran}\Lambda)r$ provided by the Sard estimate. Note that the actual computation requires the construction of P, hence of the representers of the rows of Λ^t.

H.P.(11) Show that the number $\max\{f(t) : \|Df\|_2 = \gamma,\ f\big|_S = a\}$ can be obtained as the unique choice of α for which $\|Df_0\|_2^2 + \|Df_1\|_2^2 = \gamma^2$, with f_0 the broken-line interpolant matching the given data a, and f_1 the broken-line interpolant to zero data at S *and* the additional datum $f_1(t) = \alpha$.

H.P.(12) Compute the GW-interval for the setup in (IV.24)Example, but with the bound $\|Dg\|_2 \le 2$.

.

H.P.(13) **Cubic spline interpolation.** Consider the ips $X := C^{(2)}[0\,..\,1]$ with ip

$$\langle f, g\rangle := f(0)g(0) + (Df)(0)(Dg)(0) + \int_0^1 (D^2 f)(t)(D^2 g)(t)\,\mathrm{d}t. \tag{14}$$

(a) Prove that, for any strictly increasing sequence $0 = \xi_1 < \cdots < \xi_n = 1$, the representers of the columns of $\Lambda := [\delta_0 D, \delta_{\xi_1}, \ldots, \delta_{\xi_n}, \delta_1 D]$ span the space $\$:= \Pi_{3,\xi}^2$ of twice continuously differentiable piecewise cubic polynomials on $[0\,..\,1]$ with breakpoints $\xi_1, \xi_2, \ldots, \xi_n$. (Hint: Prove that the representer of any column of Λ lies in $\$$, that Λ is 1-1 (e.g., show that Λ^t is onto by showing that Λ^t is 1-1 on Π_{n+1}), and that $\dim \$ \le n+2 (= \#\Lambda)$.)

(b) Prove that, for every $f \in X$, there exists a unique $s \in \$$ for which $\Lambda^t s = \Lambda^t f$ and that this s uniquely minimizes $\int_0^1 (D^2 g)(t)^2\,\mathrm{d}t$ over all $g \in X$ with $\Lambda^t g = \Lambda^t f$.

**** reproducing kernels ****

If X is a Hs of scalar-valued functions, all on the same domain T, and if $\delta_t \in X^*$ for all $t \in T$, then there exists $k : T \times T \to \mathbb{F}$ so that $\delta_t f = \langle f, k(\cdot, t)\rangle$ for all $f \in X$. The function k is called the **reproducing kernel** for X, and X is called a **reproducing kernel Hs**. In such a Hs, the representer for $\lambda \in X^*$ can be computed as the function

$$\lambda^{-c} : T \to \mathbb{F} : t \mapsto \overline{\lambda k(\cdot, t)}.$$

Note that $k(\cdot, t) \in X$, hence

$$k(s,t) = \delta_s k(\cdot, t) = \langle k(\cdot, t), k(\cdot, s)\rangle = \overline{\langle k(\cdot, s), k(\cdot, t)\rangle} = \overline{k(t,s)},$$

showing that *any reproducing kernel k is hermitian.*

For *example*, the function space

$$\mathbf{L}_2^{(2)}[0\,..\,1] := \{t \mapsto a + bt + \int_0^1 (t-s)_+ h(s)\,\mathrm{d}s : a, b \in \mathbb{R}, h \in \mathbf{L}_2[0\,..\,1]\}$$

is a reproducing kernel Hs with respect to the ip (14). In fact, by Taylor's formula,

$$f(t) = f(0) + (Df)(0)t + \int_0^1 (t-s)_+ (D^2 f)(s)\,\mathrm{d}s = \langle f, k(\cdot, t)\rangle,$$

with

$$k(s,t) := 1 - t^3/6 + s(t + t^2/2) + (t-s)_+^3/3!.$$

**** computing the representer ****

The only question remaining is how to compute the *representer* λ^{-c} for given λ, particularly when we are not in a reproducing kernel Hs. As the earlier conclusion indicates, this can be done in principle by constructing a ba from $\ker\lambda$. A more practical and widely used scheme constructs the representer as the

$$\mathbf{argmin}\Phi_\lambda(X),$$

i.e., as the point at which $\inf \Phi_\lambda(X)$ is taken on, where, for $\lambda \in X'$, we define

$$\Phi_\lambda : X \to \mathbb{R} : x \mapsto \|x\|^2 - 2\,\mathrm{Re}\,\lambda x.$$

The motivation is simple: if λ^{-c} exists, then $\Phi_\lambda(x) = \|x\|^2 - 2\,\mathrm{Re}\langle x, \lambda^{-c}\rangle + \|\lambda^{-c}\|^2 - \|\lambda\|^2 = \|x - \lambda^{-c}\|^2 - \|\lambda\|^2$, and this has the unique minimizer $x = \lambda^{-c}$. Since, by (11)Theorem, every $\lambda \in X^*$ has a representer, if not in X, then in its completion, this proves:

(15) Proposition. *$y \in X$ represents λ* $\iff$ $y = \mathrm{argmin}\,\Phi_\lambda(X)$.

For completeness, we now show that Φ_λ fails to have a lower bound in case $\lambda \in X' \backslash X^*$.

(16) Proposition. *Φ_λ is bounded away from $-\infty$* $\iff$ $\lambda \in X^*$.

Proof: $\lambda \in X^* \implies \Phi_\lambda(x) = \|x\|^2 - 2\,\mathrm{Re}\,\lambda x \ge \|x\|^2 - 2\|\lambda\|\|x\| + \|\lambda\|^2 - \|\lambda\|^2 = (\|x\| - \|\lambda\|)^2 - \|\lambda\|^2 \ge -\|\lambda\|^2$. Conversely, if not $\lambda \in X^*$, then, for some (x_n) in S_X, $\lambda x_n > n$, all n, hence $\Phi_\lambda(x_n) = 1 - 2\,\mathrm{Re}\,\lambda x_n < 1 - 2n \xrightarrow[n\to\infty]{} -\infty$. □

H.P.(14) Prove directly, i.e., without recourse to (11), that $y \in X$ represents $\lambda \in X'$ iff $y = \operatorname{argmin} \Phi_\lambda(X)$. (Hint: Work out that $\Phi_\lambda(x+y) = \|x\|^2 - 2\operatorname{Re}\mu x + \Phi_\lambda(y)$, with $\mu := \lambda - y^c \in X'$, then argue that y cannot be a minimizer for Φ_λ unless $\mu = 0$ (cf. proof of (8)).)

(15)Proposition suggests finding the representer λ^{-c} for $\lambda \in X^*$ by minimizing Φ_λ over a finite-dimensional lss Y_n of X, getting y_n as the minimizer, and then letting $\dim Y_n \to \infty$ appropriately, i.e., so that

$$X = \liminf{}_{n\to\infty} Y_n := \{x \in X : \lim_{n\to\infty} d(x, Y_n) = 0\}.$$

This is the essence of the Rayleigh-Ritz-Galerkin method.

H.P.(15) Let X ips, let $\Phi : X \to \mathbb{F} : x \mapsto \|\Lambda^t x - a\|^2 + \rho\|x\|^2$, with $\Lambda^t : X \to \mathbb{F}^n : x \mapsto (\langle x, \lambda_i\rangle)_1^n$ and $\lambda_i \in X$, $a \in \mathbb{F}^n$ and $\rho > 0$ given, and $\|\Lambda^t x - a\|^2 := (\Lambda^t x - a)^c(\Lambda^t x - a)$. Prove that there is a unique $x = \operatorname{argmin} \Phi$ and that such $x \in Y := \operatorname{ran}[\lambda_1, \lambda_2, \ldots, \lambda_n]$. (Hint: Prove first that, for any $x \in X$, $\Phi(x) \geq \Phi(P_Y x)$.)

Rayleigh-Ritz-Galerkin

If Y is lss of ips X and $y = \operatorname{argmin} \Phi_\lambda(Y)$, then y represents λ on Y, i.e., $\forall\{x \in Y\}\ \lambda x = \langle x, y\rangle$. What is the relationship of y to the representer λ^{-c} (if any) of λ on all of X? We must have

$$\forall\{x \in Y\}\ \ \langle x, y\rangle = \lambda x = \langle x, \lambda^{-c}\rangle,$$

i.e., $\lambda^{-c} - y \perp Y$, while $y \in Y$. Therefore, $y = P_Y \lambda^{-c}$, i.e., *the representer of λ on Y is the best approximation from Y to the representer of λ.* This shows the **optimality** of the Ritz method: Having chosen to seek an approximation to λ^{-c} from some lss Y, minimization of Φ_λ over Y leads to the best approximation to λ^{-c} from Y.

Here is an abstract description of the **Rayleigh-Ritz method**. The problem is, given X ips, $A \in L(X)$, $g \in X$, to find $f \in X$ such that

$$Af = g.$$

This is equivalent to the **weak formulation**

$$\text{find } f \in X \text{ s.t. } \forall\{x \in X\}\ \langle x, Af\rangle = \langle x, g\rangle.$$

Now suppose that A is hermitian and positive definite, i.e.,

$$\langle , \rangle_A : X \times X \to \mathbb{F} : (x, y) \mapsto \langle x, Ay\rangle$$

is an inner product. Write

$$X_A := (X, \|\cdot\|_A)$$

with $\|\cdot\|_A := \langle\cdot,\cdot\rangle_A^{1/2}$. Then the weak formulation is equivalent to looking for a representer of the linear functional $\lambda := g^c$ wrto the ip $\langle , \rangle_A$. Is $\lambda \in X_A^*$? A sufficient condition for this would be to have the two norms $\|\cdot\|$ and $\|\cdot\|_A$ equivalent. There are two parts to this, viz. $0 < \inf_x \|x\|_A/\|x\|$ and $\sup_x \|x\|_A/\|x\| < \infty$.

One calls $\langle,\rangle_A$ **coercive** in case $0 < \inf \|x\|_A/\|x\|$. Coercivity of $\langle,\rangle_A$ implies that A is bounded below (since $\|x\|_A^2/\|x\|^2 \le \|Ax\|/\|x\|$ by CBS). Also, it implies that $g^c \in X_A^*$, since, setting

$$c := \inf \|x\|_A/\|x\| > 0,$$

we have $\|x\| \le \|x\|_A/c$, so $|g^c x| = |\langle x, g\rangle| \le \|x\|\|g\| \le (\|g\|/c)\|x\|_A$, i.e., $\|g^c\|_A \le \|g\|/c$.

Assuming $\langle,\rangle_A$ to be coercive, there exists, by (11)Riesz-Fischer, one and only one f in the *completion* of X_A for which

$$\forall\{x \in X_A = X\} \;\; \langle x, f\rangle_A = \langle x, g\rangle,$$

and then $Af = g$ *in this sense.* If $f \in X$, all is well, and we conclude that, in fact,

$$Af = g.$$

This will happen in case X is Hs and $\langle,\rangle_A$ is also **bounded**, i.e., $\sup \|x\|_A/\|x\| < \infty$, since then the two norms $\|\cdot\|$ and $\|\cdot\|_A$ are equivalent, therefore X_A is complete, hence $f \in X = \operatorname{dom} A$. Note that $\langle,\rangle_A$ is bounded in case $A \in bL(X)$ since then $\|x\|_A^2 = \langle x, Ax\rangle \le \|x\|\|Ax\| \le \|x\|^2\|A\|$. But if $\langle,\rangle_A$ fails to be bounded, then (by OMT) X_A fails to be complete, and f may not be in $X_A = \operatorname{dom} A$. Then we may not be able to apply A to f. In that case, f is called a **weak** solution of the original problem. The same language is used in the slightly more general case that A is only defined on some lss of the ips X.

Whether or not X_A is complete and/or $\operatorname{dom} A = X$, minimization of

$$\Phi : x \mapsto \langle x, Ax\rangle - 2\operatorname{Re}\langle x, g\rangle$$

over the finite-dimensional lss Y provides the unique $f_Y := P_Y f$, and $\|f_{Y_n} - f\|_A \xrightarrow[n\to\infty]{} 0$ in case we choose a sequence (Y_n) of subspaces that becomes eventually dense, i.e., so that $\liminf Y_n = X_A$. One way to insure this is to use $Y_n := \operatorname{ran}[x_1, x_2, \ldots, x_n]$, with $(x_j : j = 1, 2, \ldots)$ a complete orthonormal sequence (see below).

H.P.(16) Provide all the missing details in the following examples.

** example: Poisson **

A concrete example is provided by **Poisson's equation**: $G \subseteq \mathbb{R}^2$ and sufficiently nice (e.g., open, bounded, with piecewise smooth boundary), $g \in \mathbf{L}_2(G)$. Find f on G^- s.t.

$$-\Delta f = g, \quad f|_{\partial G} = 0. \tag{17}$$

Here, Δ is the **Laplacian**, i.e., $\Delta := D_1^2 + D_2^2$, with

$$D_i := \text{partial differentiation wrto the } i\text{th argument.}$$

Then $A := -\Delta$ is defined on the lss $\operatorname{dom} A := C_0^{(2)}(G) := \{u \in C^{(2)}(G^-) : u|_{\partial G} = 0\}$ of the ips $X = C(G^-)$ with inner product

$$\langle u, v\rangle := \int_G uv := \int_G u(t)v(t)\,\mathrm{d}t$$

and corresponding norm $\|\cdot\| := \|\cdot\|_2$. (Note that I have chosen $\mathbb{F}=\mathbb{R}$.) The first thing is to bring $\langle,\rangle_A$ into a more obviously symmetric form. (Actually, we'll be bringing it only back into the form in which it arose in the modeling of certain physical situations and from which earlier Mathematics triumphantly derived Poisson's equation.) We use **integration by parts**:

$$\int_G D_z f = \int_{\partial G} (n^c z) f,$$

where $D_z := \sum_i z(i) D_i$, n is the outward unit normal to G, and $n^c z$ is its scalar product with the vector z. When applied to

$$D_z(uv) = (D_z u)v + u(D_z v),$$

this gives $\int_G \{(D_z u)v + u(D_z v)\} = \int_{\partial G} (n^c z) uv$, hence

$$\int_G (D_i^2 u)v = \int_{\partial G} n(i)(D_i u)v - \int_G (D_i u)(D_i v),$$

therefore

$$\langle \Delta u, v\rangle = \int_{\partial G} (D_n u)v - \langle Du, Dv\rangle,$$

with the understanding that $\langle Du, Dv\rangle := \int_G (Du)^c (Dv)$ is the integral over G of the scalar product of the *gradient* Du of u with the gradient of v.

Since $x \in X$ vanishes on the boundary ∂G of G, we get

$$\langle u, v\rangle_A = \langle u, -\Delta v\rangle = \langle Du, Dv\rangle,$$

a formulation in which u and v appear in more obvious symmetry. This formulation also makes it obvious that $\langle,\rangle_A$ *is coercive*: Since $u \in X$ vanishes on ∂G, we may extend it to a continuous, piecewise $C^{(1)}$ function on all of $\mathbb{R}^2$ by setting $u := 0$ off G. Since G is bounded, we may assume that $G \subseteq [0\,.\,.\,a]^2$ for some positive a. For $t \in [0\,.\,.\,a]^2$,

$$u(t) = \int_0^{t_1} (D_1 u)(r, t_2)\,\mathrm{d}r \le \int_0^a |D_1 u|(r, t_2)\,\mathrm{d}r \le \sqrt{a}\|D_1 u(\cdot, t_2)\|_2,$$

(the last inequality by Hölder), so

$$\|u\|^2 = \int_G |u|^2 \ \le\ a \int_0^a \int_0^a \|Du_1(\cdot, t_2)\|_2^2 \,\mathrm{d}t_2\,\mathrm{d}t_1 = a^2\|D_1 u\|^2 \le a^2 \|Du\|^2.$$

We conclude that, for arbitrary $g \in \mathbf{L}_2(G)$, the Poisson problem (17) has a unique weak solution f in the *completion* $\mathbf{L}_{2,0}^{(1)}(G)$ of $X = C_0^{(2)}(G)$ wrto the **Dirichlet** norm $\|\cdot\|_A = \|D\cdot\|_2$. If g is only in $\mathbf{L}_2$, this solution cannot possibly lie in $C^{(2)}$, hence the expression $-\Delta f$ may not make pointwise sense.

Use of a completion has ensured existence of a solution of sorts. The subsequent task of ascertaining just how smooth this weak solution is may be nontrivial.

**** example: second-order elliptic ****

One treats a more general second-order elliptic differential operator

$$A := -\sum_{i,j} D_i(a_{ij}D_j)$$

by relating it to the Laplace operator. This means that one considers the ips $X = (C_0^{(2)}(G), \|D\cdot\|_2)$ with the ip now given by the bilinear form

$$\langle u, v\rangle = \int_G (Du)^c(Dv).$$

On this space, we consider the bilinear form

$$F(u,v) := \int_G \sum_{i,j} a_{ij}\ (D_iu)(D_jv) = \int_G u\,A\,v,$$

the last equality by integration by parts, hence hope to solve the equation

$$F(\cdot, ?)\ =\ \lambda := \int_G \cdot\ g.$$

This bilinear form fails to be symmetric in general. Yet the same kind of analysis can be applied, provided one makes the right assumptions. These include: the coefficients are in $C^{(1)}(G^-)$, hence uniformly bounded, and are such that A is **uniformly elliptic**, i.e.,

$$m := \inf_\xi \inf_{x\in G} \sum_{i,j} a_{ij}(x)\xi(i)\xi(j)/\|\xi\|_2^2 > 0.$$

The uniform ellipticity implies that

$$\forall\{u\in X\}\quad F(u,u) \ge m\|u\|^2, \qquad \text{hence } F \text{ is } \textit{coercive},$$

while the uniform boundedness of the coefficients implies that

$$|F(u,v)| \le \sum_{i,j} \|a_{ij}\|_\infty \int_G |Du|^c|Dv| \le M\|u\|\|v\|, \qquad \text{hence } F \text{ is } \textit{bounded.}$$

**** abstract nonsense ****

The rest of the discussion is applicable to any (skew-)bilinear functional (or form) F on some ips Y that is **bounded** (i.e., $\sup_{x,y} |F(x,y)|/(\|x\|\|y\|) < \infty$) and **coercive** (i.e., $\inf_y F(y,y)/\|y\|^2 > 0$). If Y is not complete, we first complete it and extend F to the completion Z by continuity.

H.P.(17) Prove that *the definition*

$$F(u,v) := \lim F(u_n,v_n) \text{ with } (u_n,v_n) \in Y \times Y \text{ and } (u,v) = \lim(u_n,v_n)$$

is independent of the particular sequence $((u_n,v_n))$ *used and provides an extension of* F *to a bounded, bilinear, coercive functional on* Z.

Assume that Z is complete and consider the map

$$\varphi : Z \to Z^* : y \mapsto F(\cdot, y).$$

This map is defined and bounded (by the boundedness of F), and is linear. It is also bounded below: $\|F(\cdot,z)\| = \sup_y |F(y,z)|/\|y\| \ge F(z,z)/\|z\|$, hence $\inf \|F(\cdot,z)\|/\|z\| > 0$ (by coercivity). Consequently, $\operatorname{ran}\varphi$ is closed (by H.P.(V.20)). Hence, φ is onto provided $\operatorname{ran}\varphi$ is dense, i.e. (cf. Cor.2 to (IV.33)Duality in Approximation Theory),

$$\{0\} = (\operatorname{ran}\varphi)^\perp = \{z \in Z^{**} : \forall\{\lambda \in \operatorname{ran}\varphi\}\ z\lambda = 0\}. \tag{18}$$

However, Z is Hs by assumption, hence $Z^{**} = Z$, and so $(\operatorname{ran}\varphi)^\perp = (\operatorname{ran}\varphi)_\perp = \{z \in Z : \forall\{\lambda \in \operatorname{ran}\varphi\}\ \lambda z = 0\} = \{z \in Z : \forall\{x \in Z\}\ F(z,x) = 0\} \subseteq \{z \in Z : F(z,z) = 0\} = \{0\}$, the last equality by coercivity. This shows that

$$(\operatorname{ran}\varphi)^\perp = \{0\}.$$

H.P.(18) Give an example of a proper closed lss L in the dual of a Bs X for which $L_\perp = \{0\}$, yet $L \ne X^*$. (Hint: Consider $\ker M$, with $M \in X^{**}\backslash JX$.)

This proves the

(19) Lax-Milgram lemma. *If F is a bounded, (skew-)bilinear, coercive functional on some Hs Z, then the map $Z \to Z^* \simeq Z : z \mapsto F(\cdot, z)$ is bounded and boundedly invertible. In particular, for given $\lambda \in Z^*$, the equation*

$$F(\cdot, f) = \lambda \tag{20}$$

has a unique solution $f \in Z$, and f depends continuously on λ.

**** numerics ****

The numerical solution of equation (20) proceeds in expected ways: One picks a lss Y of dimension n and determines $y \in Y$ s.t. $\forall\{z \in Y\}\ F(z,y) = \lambda z$. This means that we look for $y \in Y$ so that $F(\cdot,y) = \lambda$ on Y. Since F is coercive, so is $F|_{Y\times Y}$, hence $Y \to Y^* : y \mapsto F(\cdot,y)$ is 1-1, therefore onto (since $\dim Y^* = \dim Y = n$). Writing the solution in terms of some basis $V \in L(\mathbb{F}^n, Y)$ for Y as $y = Va$, we determine its coefficient vector a as the unique solution of the linear system

$$\sum_{j=1}^n F(w_i, v_j)\overline{a(j)} = \lambda w_i, \qquad i = 1, \ldots, n, \tag{21}$$

with $W \in L(\mathbb{F}^n, Y)$ any particular convenient basis for Y. Finally, if we choose $Y = Y_n$ with $\liminf Y_n = Z$, then the solution $y := Va$ of (21) converges to f as $n \to \infty$.

**** generalization to bilinear fl on two Hilbert spaces ****

It is instructive to compare the development so far with the more general situation that $F : Z \times Y \to \mathbb{R}$ is bilinear and bounded, with both Z and Y Hs's. Coercivity doesn't make sense anymore. Rather, we say that *the (skew-)bilinear functional F on the nls pair (Z, Y) is* **bounded below** in case it satisfies the **Babuška-Brezzi Condition**, i.e., in case

$$\inf_{y \in Y} \sup_{z \in Z} \frac{|F(z,y)|}{\|z\|\|y\|} > 0,$$

since this is equivalent to having the map

$$\varphi : Y \to Z^* : y \mapsto F(\cdot, y)$$

bounded below, hence 1-1. But there is now no reason why φ should be onto. In fact, it may now happen that $(\operatorname{ran} \varphi)_\perp$ is not trivial. One deals with this by bringing in an additional bounded bilinear fl

$$G : Z \times Z \to \mathbb{R}$$

that one assumes to be coercive on $(\operatorname{ran} \varphi)_\perp$, and can then conclude that the enlarged linear map

$$\psi : Z \times Y \to Z^* \times Y^* : (z, y) \mapsto (G(z, \cdot) + F(\cdot, y), F(z, \cdot))$$

is 1-1 and onto. If $\psi(z, y) = (\lambda, \mu)$ and G happens to be hermitian, then z minimizes $x \mapsto G(x, x) - 2 \operatorname{Re} \lambda x$ over $(\varphi^*)^{-1}(\mu)$. For this, note that the dual φ^* of φ is given by the map $Z \to Y^* : z \mapsto F(z, \cdot)$, hence having φ 1-1 implies that φ^* is onto.

**** generalization to Bs ****

A further generalization is possible, to the situation where F is a bounded bilinear functional on $Z \times Y$, with Z, Y mere Bs's. Assuming F to be bounded below, we get again that $\varphi : Y \to Z^* : y \mapsto F(\cdot, y)$ is bounded below, hence 1-1 and with $\operatorname{ran} \varphi$ closed. But having φ onto requires additional assumptions. The lack of the coercivity assumption also makes itself felt when considering discretizations. For *example*, if $Z = Y = \ell_2$ and $F : (x, y) \mapsto y{*}x$, then F is bounded below, yet, with $Y_n := \operatorname{ran}[e_1, \ldots, e_n]$ and $Z_n := \operatorname{ran}[e_1, \ldots, e_{n-1}, e_{n+1}]$, the discretization

$$\text{find } y \in Y_n \text{ s.t. } \forall \{z \in Z_n\}\ F(z, y) = \lambda z$$

is for some λ *not* solvable, no matter how large we take n. In effect, Z_n and Y_n must be paired properly.

(22) Proposition. *If $F : Z \times Y \to \mathbb{R}$ is bounded below, then $\forall\{$finite-dim.lss $Y_n \subset Y\}\ \exists\{$lss $Z_n \subset Z\}\ \forall\{\lambda \in Z^*\}$ the linear system*

$$F(\cdot, ?) = \lambda \text{ on } Z_n$$

has exactly one solution in Y_n.

Proof: $\varphi : Y \to Z^* : y \mapsto F(\cdot, y)$ is 1-1, hence, with $V =: [v_1, v_2, \ldots, v_n]$ a basis for Y, φV is 1-1, therefore (by (I.33)Corollary) can find $W =: [w_1, w_2, \ldots, w_n] \in L(\mathbb{F}^n, Z)$ so that the Gramian $(\varphi(v_j)w_i) = (F(w_i, v_j))$ is invertible, hence $Y_n \to Z_n^* : y \mapsto F(y, \cdot)|_{Z_n}$ is invertible. □

c.o.n. sets

A simple way to generate sequences (Y_n) of lss's of the ips X for which $\liminf Y_n = X$ is to choose $Y_n := \mathrm{ran}[f_1, f_2, \ldots, f_n]$, with (f_i) a complete orthonormal sequence in X.

(23) Definition. *$F \subseteq X$ is* **orthonormal** *(=:* **o.n.***) $:\iff \forall\{f, g \in F\}\ \langle f, g\rangle = \delta_{fg}$.*
An o.n. F is **complete** *if it is fundamental, i.e., if $\mathrm{ran}[F]$ is dense.*

We already proved the basic fact about o.n. sets: *If $[f_1, f_2, \ldots, f_n]$ is an o.n. basis for the lss Y of X, then $\sum_i |\langle x, f_i\rangle|^2 = \|P_Y x\|^2$*, hence, since $\|x\|^2 = \|P_Y x\|^2 + \|x - P_Y x\|^2$,

$$\sum_i |\langle x, f_i\rangle|^2 \le \|x\|^2 \text{ with equality iff } x \in Y \quad (\text{i.e., } x = P_Y x).$$

This implies

(24) Bessel's inequality. *For an arbitrary o.n. set F,*

$$\sum_{f\in F} |\langle x, f\rangle|^2 := \sup_{\#G<\infty} \sum_{g\in G} |\langle x, g\rangle|^2 \ \le\ \|x\|^2 \quad \text{with equality iff } x \in (\mathrm{ran}[F])^-.$$

Proof: We know that, for *finite* $G \subset F$, $\sum_{g\in G} |\langle x, g\rangle|^2 = \|x\|^2 - d(x, \mathrm{ran}[G])^2$. Hence $\sum_{f\in F} |\langle x, f\rangle|^2 \le \|x\|^2$, with equality if and only if $\inf_{\#G<\infty} d(x, \mathrm{ran}[G]) = 0$, i.e., $x \in (\mathrm{ran}[F])^-$. $\square$

H.P.(19) Prove that, for any x and any o.n. set F in the ips X, the sum $\sum_{f\in F} |\langle x, f\rangle|^2$ has at most countably many nonzero terms.

(25) Corollary. *The o.n. F in the ips X is complete if and only if* **Parseval's Identity**

$$\sum_{f\in F} |\langle x, f\rangle|^2 \ =\ \|x\|^2$$

holds for every $x \in X$.

H.P.(20) (a) Prove that the bivariate Haar's system is orthonormal but not complete. Here, **Haar's system** consists of the functions $\varphi_{k,n} : \mathbb{R} \to \mathbb{R} : t \mapsto 2^{k/2}\varphi(2^k t - n)$ for $k, n \in \mathbb{Z}$, with $\varphi := \chi_{(1/2..1)} - \chi_{(0..1/2)}$, and χ_I the **characteristic function of the set** I, i.e.,

$$\chi_I(t) := \begin{cases} 1, & t \in I; \\ 0, & t \notin I, \end{cases}$$

and it is complete (in $\mathbf{L_2}(\mathbb{R})$). The *bivariate* system consists of the functions $\psi_{k,n} : \mathbb{R}^2 \to \mathbb{R} : (s,t) \mapsto 2^k\varphi(2^k s - n_1)\varphi(2^k t - n_2)$, with $k \in \mathbb{Z}$, $n = (n_1, n_2) \in \mathbb{Z}^2$. (Hint: Is the function $x = \chi_{[0..1]^2}$ in the closure of $\mathrm{ran}[\psi_{k,n} : k \in \mathbb{Z}, n \in \mathbb{Z}^2]$?)
(b) Prove that $y := \chi_{[0..1]}$ *is* in the closed linear span of the *univariate* Haar's system.

The Gram-Schmidt process can be used to generate a c.o.n. sequence from a countable dense one. It's much neater, though, when such a sequence arises 'naturally', e.g., as the sequence of eigenvectors of a compact normal linear map (see Chapter IX). A prime example is the (biinfinite) sequence $\varphi_n : t \mapsto \exp(int)$, $n \in \mathbb{Z}$, for the Hs $\mathbf{L}_2[0\,.\,.2\pi]$. Another example, fit for $\mathbf{L}_2[-1\,.\,.\,1]$, is the sequence of Legendre polynomials. By completeness of such a c.o.n. sequence (φ_j), $\liminf_n \mathrm{ran}[\varphi_1, \varphi_2, \ldots, \varphi_n] = X$.