I. Preliminaries: Linear Algebra

** linear space **

Functional Analysis plays in linear spaces of functions. For all practical purposes, a **linear space** (=: **ls**) X is a (nonempty) collection of functions f, all on the same domain T, and this collection is closed under **pointwise addition** and **scalar multiplication**.

This means that, with $f, g \in X$, their sum $f+g$, i.e., the map

$$f + g : t \mapsto f(t) + g(t),$$

is also in X, as is the product αf of such f with any scalar α, i.e., the map

$$\alpha f : t \mapsto \alpha f(t).$$

For this to make sense, the maps in X must all have a common target, and it must be possible to add elements in the target and to multiply them with scalars.

The prime example is the collection

$$\mathbb{R}^T := \{f : T \to \mathbb{R}\}$$

of all real-valued functions on some set T, with addition and scalar multiplication defined pointwise, i.e., in the above way. In this case, the underlying scalar field is $\mathbb{R} :=$ the real number field. The most important special case occurs when $T = \{1, 2, \ldots, n\}$, in which case we get the n-dimensional **coordinate space**

$$\mathbb{R}^n := \mathbb{R}^{\{1,2,\ldots,n\}}$$

whose elements I will *never* write as 1-column matrices but, rather, as n-sequences $x = (x(i) : i = 1, \ldots, n) = (x(1), \ldots, x(n))$. In the discussion of inner product spaces and of eigenvalues, we will also consider complex linear spaces, i.e., linear spaces for which the scalar field is $\mathbb{C} :=$ the **complex** number field. Agreement: if nothing is said, then the scalars are real. If I don't care, I'll write $\mathbb{F}$ for the scalar field. Note that the collection

$$\mathbb{F}^0$$

of all *empty* sequences in $\mathbb{F}$ consists of exactly one element, (). Even this coordinate space is useful at times.

If X is a linear space, then the collection

$$X^T$$

of all functions f on the same domain T into X is also a linear space (under pointwise addition and scalar multiplication).

Another way to get linear spaces from linear spaces is by considering **linear subspace**s (=: **lss**'s). These are *nonempty* subsets Y that are closed under addition and scalar multiplication, i.e.,

$$Y + Y := \{y + y' : y \in Y, y' \in Y\} \subset Y, \qquad \alpha Y := \{\alpha y : y \in Y\} \subset Y.$$

E.g., for $T \subset \mathbb{R}^n$,

$$C(T) := \{f \in \mathbb{R}^T : f \text{ is continuous}\}$$

is a lss of $\mathbb{R}^T$.

H.P.(1) Verify that any sum and any intersection of lss's is again a lss.

Here, for the record, is the formal definition of a ls:

(1) Definition. *To say that X is a* **linear space** *(=:***ls***) (of* **vectors***) over the (commutative) field $\mathbb{F}$ (of* **scalars***) means that there are two maps, (i) $X \times X \to X : (x, y) \mapsto x + y$ called* **(vector) addition***; and (ii) $\mathbb{F} \times X \to X : (\alpha, x) \mapsto \alpha x =: x\alpha$ called* **scalar multiplication***, that satisfy the following rules.*

(a) X is a commutative group with respect to addition; i.e., addition
- *(a.1) is associative: $f + (g + h) = (f + g) + h$;*
- *(a.2) is commutative: $f + g = g + f$;*
- *(a.3) has neutral element: $\exists\{0\}\ \forall\{f\}\ f + 0 = f$;*
- *(a.4) has inverse: $\forall\{f\}\ \exists\{g\}\ f + g = 0$.*

(s) scalar multiplication is
- *(s.1) associative: $\alpha(\beta f) = (\alpha\beta)f$;*
- *(s.2) field addition distributive: $(\alpha + \beta)f = \alpha f + \beta f$;*
- *(s.3) vector addition distributive: $\alpha(f + g) = \alpha f + \alpha g$;*
- *(s.4) unitary: $1f = f$.*

It is standard to denote the element $g \in X$ for which $f + g = 0$ by $-f$ since such g is uniquely determined by the requirement that $f + g = 0$. I will denote the neutral element in X by the same symbol, 0, used for the zero scalar. In particular, as the sole element of $\mathbb{F}^0$, the empty sequence () is denoted by 0.

H.P.(2) Prove: *For f in the ls X, $(-1)f = -f$ and $0f = 0$. Also, $\alpha f = 0$ with $f \neq 0$ implies $\alpha = 0$.*

H.P.(3) Prove: *For any set T and any field $\mathbb{F}$, $\mathbb{F}^T$ is a ls with respect to pointwise addition and scalar multiplication.*

H.P.(4) Prove: *Any lss is a ls (with respect to the addition and scalar multiplication as restricted to the lss).*

lss's are often given as the range or the kernel of a linear map.

** linear map **

We deal extensively with **linear maps** (=: **lm**'s) (or operators, transformations, mappings, functions, etc. all much longer than 'map'), i.e., with $A : X \to U$ satisfying

$$A(f + g) = Af + Ag, \quad \text{all } f, g \in X \qquad (\textbf{additivity})$$

$$A(\alpha f) = \alpha(Af), \quad \text{all } \alpha \in \mathbb{F}, f \in X \qquad (\textbf{homogeneity})$$

where X and U are ls's. U could be X. The simplest examples are:

$$0 : X \to U : x \mapsto 0, \qquad \alpha : X \to X : x \mapsto \alpha x.$$

We denote by

$$L(X, U)$$

the collection of all lm's from the ls X to the ls U, and write

$$L(X) := L(X, X).$$

$L(X,U)$ is a ls under pointwise addition and scalar multiplication. In addition, the collection of linear maps is closed under composition: If $A \in L(X,U)$ and $C \in L(U,W)$, then

$$CA : X \to W : f \mapsto C(Af)$$

is in $L(X,W)$. Also, if $A \in L(X,U)$ is invertible (as a map), then $A^{-1} \in L(U,X)$. Composition (like all map composition) is associative and combines with addition and scalar multiplication of linear maps in the expected way. But, composition is not commutative.

H.P.(5) Prove that an additive map $A : X \to U$ is homogeneous for all *rational* scalars.

H.P.(6) Prove that *the inverse of a lm is linear*.

H.P.(7) Prove: *If X is a ls and $A : X \to U$ is a lm with respect to some addition and scalar multiplication on U, then* ran A *is a ls (even if U fails to be a ls).* (See the definition of quotient space below for an instructive example.)

** special case: column maps, especially matrices **

An important special case is $L(\mathbb{F}^n, X)$ which provides us with a first opportunity to practice a basic step in fa, namely *representation*. Here, we show that $L(\mathbb{F}^n, X)$ is nicely representable by X^n, the set of n-sequences in the ls X.

Indeed, each sequence $(v_1, \ldots, v_n) \in X^n$ gives rise to the corresponding map

$$[v_1, \ldots, v_n] : \mathbb{F}^n \to X : a \mapsto \sum_{j=1}^{n} v_j a(j)$$

which is evidently linear. As a special example, with

$$e_k := (\underbrace{0, \ldots, 0}_{k-1 \text{ zeros}}, 1, \underbrace{0, \ldots, 0}_{n-k \text{ zeros}}) \in \mathbb{F}^n$$

the kth **unit vector** in $\mathbb{F}^n$,

$$[e_1, \ldots, e_n] : \mathbb{F}^n \to \mathbb{F}^n : a \mapsto \sum_j e_j a(j) = a$$

is the identity, $1 = 1_n$, on $\mathbb{F}^n$.

If also $A \in L(X,U)$, then $A(\sum_j v_j a(j)) = \sum_j (Av_j)a(j)$, hence

$$A[v_1, \ldots, v_n] = [Av_1, \ldots, Av_n] \in L(\mathbb{F}^n, U).$$

In particular,

$$\forall \{A \in L(\mathbb{F}^n, X)\} \;\; A = A1_n = [Ae_1, \ldots, Ae_n]$$

and this shows that every $A \in L(\mathbb{F}^n, X)$ is uniquely representable as $[v_1, \ldots, v_n]$ (with $v_j = Ae_j$, all j). This sets up the invertible linear map

$$X^n \to L(\mathbb{F}^n, X) : (v_j : j = 1, \ldots, n) \mapsto [v_1, \ldots, v_n].$$

This is quite familiar for the special case that also X is a coordinate space, $X = \mathbb{F}^m$ say. In that case, each v_j is an m-vector, and one associates $[v_1, \ldots, v_n]$ with the $m \times n$ matrix whose jth column contains the entries of v_j. This sets up the invertible linear map

$$\mathbb{F}^{m\times n} = (\mathbb{F}^m)^n \to L(\mathbb{F}^n, \mathbb{F}^m) : M \to [M(:,1), \ldots, M(:,n)],$$

i.e.,

$$L(\mathbb{F}^n, \mathbb{F}^m) \simeq \mathbb{F}^{m\times n}.$$

For this reason, I will *identify* the two, i.e., refer to $M \in \mathbb{F}^{m\times n}$ as both a linear map and as the matrix representing it. If also $A \in L(\mathbb{F}^m, \mathbb{F}^r) \simeq \mathbb{F}^{r\times m}$, then, with this identification,

$$[A(:,1), \ldots, A(:,m)]M = A[M(:,1), \ldots, M(:,n)] = [AM(:,1), \ldots, AM(:,n)],$$

which is the reason why we define the matrix product AM as

$$(AM)(i,j) = \sum_k A(i,k)M(k,j), \qquad \forall\, i,j.$$

In analogy, for $v_1, \ldots, v_n$ in some arbitrary ls X, I will call the corresponding lm $[v_1, \ldots, v_n]$ a **column map**, and refer to v_j as its jth **column**. Such terminology is entirely nonstandard (but very helpful).

** lss's often come as ker or ran **

For a lm $A : X \to U$, the **kernel** or **nullspace** of A, i.e.,

$$\ker A := \{f \in X : Af = 0\},$$

is important because

$$A \textit{ is 1-1} \iff \ker A = \{0\}.$$

It is also important since a lss is usually specified as the range or the kernel of a linear map.

The definition of a lss as the *range* of a linear map is *constructive* in that it is easy to write down all of its elements (assuming that we have a description of the domain of that map). On the other hand, it may be hard to test whether a given element lies in the lss.

By contrast, the definition of a lss as the *kernel* of a linear map makes it very easy to test whether a given element lies in it, but it is *not constructive* (though perhaps more elegant): Offhand, we know no element of such a lss other than 0.

It is best to have a description as both a range and a kernel (as happens for the intermediate spaces in exact sequences).

(2) Example $\Pi_k :=$ polynomials of degree $\le k$ in one real variable. Constructive definition: With

$$()^j : \mathbb{R} \to \mathbb{R} : t \mapsto t^j, \quad \text{all } j,$$

the **monomials**, we have

$$\Pi_k := \operatorname{ran}[()^0, \ldots, ()^k] = \{\sum_0^k ()^j a(j) : a \in \mathbb{R}^{\{0,\ldots,k\}}\}.$$

The alternative definition

$$\Pi_k := \ker D^{k+1},$$

with D the linear map carrying a continuously differentiable f to its (first) derivative, is nonconstructive. (The Fundamental Theorem of Calculus verifies the equivalence of these two definitions.)

More generally,

$$\Pi_k(\mathbb{R}^d) \ := \ \mathrm{ran}[()^\alpha : |\alpha| \le k] = \{ \sum_{|\alpha|\le k} ()^\alpha \, a(\alpha) : a(\alpha) \in \mathbb{R}\}$$

denotes the space of polynomials in d arguments of total degree $\le k$, with

$$()^\alpha := \prod_{j=1}^{d} ()^{\alpha(j)} : x \mapsto x(1)^{\alpha(1)} \cdots x(d)^{\alpha(d)}, \quad \alpha \in \mathbb{Z}_+^d := \{\alpha \in \mathbb{Z}^d : \alpha(j) \ge 0, \text{all } j\}$$

and

$$|\alpha| := \alpha(1) + \cdots + \alpha(d).$$

This space can be shown to be the kernel of the lm $f \mapsto (D^\alpha f : |\alpha| = k+1)$, as a map on $C^{(k+1)}(\mathbb{R}^d)$ to $(C(\mathbb{R}^d)^{\#\{|\alpha|=k+1\}}$ say.

H.P.(8) Check the definition you learned of $C(\mathbb{R})$ to see whether it describes $C(\mathbb{R})$ as the range or as the kernel of a lm.

H.P.(9) Show that $C(\mathbb{R}) \subset \mathbb{R}^\mathbb{R}$ is the intersection of kernels of (one or more) (extended) seminorms. (A **seminorm** on a ls X is any map $\lambda : X \to \mathbb{R}+$ that is **subadditive** (i.e., $\forall\{x, y \in X\}\ \lambda(x + y) \le \lambda(x) + \lambda(y)$) and **absolutely homogeneous** (i.e., $\forall\{\alpha \in \mathbb{R}, x \in X\}\ \lambda(\alpha x) = |\alpha|\lambda(x)$) . It is **extended** if it is allowed to take the value $+\infty$, in which case $0 \cdot \infty := 0$.)

** quotient space **

We have occasion to use one other source of ls's, namely the construction of the **quotient space** X/Y of a ls X and its lss Y, and this uses the definitions

$$M \pm N := \{m \pm n : m \in M,\ n \in N\}, \qquad \alpha M := \{\alpha m : m \in M\}$$

of the sum $M + N$ and the difference $M - N$ of two subsets of a ls, respectively the scalar multiple αM of a subset with a scalar. (Note that the difference $M - N$ is not at all the same as the set-theoretic 'difference' $M\backslash N := \{m \in M : m \notin N\}$.) Since $Y + Y = Y$ and $\alpha Y = Y$ for any nonzero scalar α, the map

$$x \mapsto \langle x\rangle \ := \ x + Y$$

is linear (e.g., $\langle x\rangle + \langle y\rangle = x + Y \ + \ y + Y = (x + y) \ + \ (Y + Y) = (x + y) + Y = \langle x + y\rangle$ since $Y + Y = Y$), provided we *define* $0\langle x\rangle := \langle 0\rangle = Y$. It follows from H.P.(7) that its range, i.e., the collection

$$X/Y := \{\langle x\rangle : x \in X\}$$

of subsets of X, is a ls (with respect to the addition and scalar multiplication of such subsets of X just defined). The lm

$$\langle\rangle : X \to X/Y : x \mapsto \langle x\rangle$$

is called the **quotient map**.

(3) Factor Lemma. *If $A \in L(X,Z)$ contains the lss Y in its kernel, then A has the quotient map $\langle\rangle : X \to X/Y : x \mapsto \langle x \rangle$ as a factor, i.e., $\exists\{C \in L(X/Y, Z)\}\ A = C\langle\rangle$.*

$$\begin{array}{ccc} & X \quad \stackrel{A}{\longrightarrow} & Z \\ \langle\rangle & \downarrow \qquad \nearrow_{C} & \\ & X/Y & \end{array}$$

Proof: The definition $C : \langle x \rangle \mapsto Ax$ is unambiguous since $x' \in \langle x \rangle$ implies that $x' - x \in Y \subseteq \ker A$, hence $Ax' = Ax$. □

In particular, each $A \in L(X,Z)$ induces the corresponding lm

$$A_| : X/\ker A \to Z : \langle x \rangle \mapsto Ax,$$

its **factor map**, and this map is 1-1 and onto $\operatorname{ran} A$, and satisfies $A = A_|\langle\rangle$:

$$\begin{array}{ccc} & X \quad \stackrel{A}{\longrightarrow} & Z \\ \langle\rangle & \downarrow \qquad \nearrow_{A_|} & \\ & X/\ker A & \end{array}$$

The notation $A_|$ for the factor map is *not standard* and should not be confused with the restriction $A|_Y$ of A to some subset Y of X.

H.P.(10) Show that the collection of all straight lines in the plane parallel to a fixed straight line is a linear space under set addition and (appropriately defined) scalar multiplication.

H.P.(11) As an exercise in visualizing the sum $M + N$ of two subsets M and N of a ls, draw (a) the sum of the disc $\{x \in \mathbb{R}^2 : x(1)^2 + x(2)^2 \le 1\}$ and the interval $[0\,..\,1]x := \{\alpha x \in \mathbb{R}^2 : 0 \le \alpha \le 1\}$, with $x := (1,1)$; (b) the sum of the four 'intervals' $[0\,..\,1]x$, $x \in \{(1,0),(0,1),(1,1),(-1,1)\}$; (c) the difference $M - N$, with $M = [0\,..\,1]^2$ and $N = \{(1,1),(2,2)\}$.

** linear functionals; dual **

In computations, we cannot deal with functions directly. Rather, we deal with numerical information about them, such as value at a point, derivative at a point, coefficients in some expansion, integral over some domain, limit at a point, first zero in an interval, maximum value, etc. All of these are provided by **functionals**, i.e., by maps from the ls X into the scalars $\mathbb{F}$. Among these, we find the **linear** functionals (=: **lfl**'s) particularly useful. They form the **dual** $L(X,\mathbb{F})$ of X, denoted by a prime:

$$X' := L(X,\mathbb{F}).$$

E.g., for $X = C^{(1)}[a\,..\,b]$ (:= the collection of all functions f on $[a\,..\,b]$ whose first derivative Df is continuous), the following are linear functionals:

$$f \mapsto f(t), \quad f \mapsto (Df)(t), \quad f \mapsto \lim_{t\to a} f(t), \quad f \mapsto \int_a^b f(t)w(t)\,\mathrm{d}t,$$

while

$$f \mapsto Df$$

is linear but not a functional, and

$$f \mapsto \sup{}_t f(t), \quad f \mapsto \|f\|_\infty := \sup{}_t |f(t)|, \quad f \mapsto \min f^{-1}\{0\}$$

are functionals but not linear. In Numerical Analysis, the linear functional of evaluation at a point is so important that we give it here its own special symbol:

$$\delta_t : f \mapsto f(t).$$

H.P.(12) Let X, Y be ls's, with $X \subset Y$. What is wrong with the conclusion that 'therefore' $X' \subset Y'$?

Offhand, X' is an abstract construct. For concrete work, one usually looks for a *representation* of X' (or some of its subspaces). Here is a first example:

For $c \in \mathbb{F}^n$, the map

$$c^t : \mathbb{F}^n \to \mathbb{F} : x \mapsto c^t x := \sum_{i=1}^{n} c(i)x(i) \qquad = [c(1), \ldots, c(n)]x$$

is a lfl. The resulting map

$$\mathbb{F}^n \to (\mathbb{F}^n)' : c \mapsto c^t = [c(1), \ldots, c(n)]$$

is linear, 1-1 and onto, i.e., $(\mathbb{F}^n)' \simeq \mathbb{F}^n$. I usually *identify* $(\mathbb{F}^n)'$ *with* $\mathbb{F}^n$ in this way.

For *example*, recalling that $\mathbb{F}^n = \mathbb{F}^{\{1,\ldots,n\}}$, I write δ_i on $\mathbb{F}^n$ as ${e_i}^t$, since the linear functional $\delta_i : \mathbb{F}^n \to \mathbb{F} : x \mapsto x(i)$ is *represented by* the ith unit-vector, e_i, in the sense that $\delta_i x = x(i) = {e_i}^t x$ for all $x \in \mathbb{F}^n$.

**** bidual ****

We can always think of $f \in X$ as a linear functional on X', viz. as the linear map

$$f'' : X' \to \mathbb{F} : \lambda \mapsto \lambda f.$$

The resulting map

$$J_0 : X \to X'' : f \mapsto f''$$

from X to its **bidual**

$$X'' := (X')'$$

so defined is linear. By H.P.(19), J_0 is also 1-1.

This means that J_0 provides an **embedding**, the so called **canonical** embedding, of X into X''. It can be shown (see the discussion after (14)Corollary) that J_0 is onto if and only if $X \simeq \mathbb{F}^n$ for some n. I don't care to go into that at this point. Just keep in mind that it is always possible in a natural way to think of any element f in some ls X as the linear functional f'' on the linear functionals on X. This turns out to be, at times, a very useful point of view.

**** numerical representation; basis ****

It is usually not possible to compute directly in an arbitrary linear space X, but only in an associated coordinate space, i.e., in $\mathbb{F}^n$. The association is made by linearly mapping $\mathbb{F}^n$ to X or X to $\mathbb{F}^n$. Maps from X to $\mathbb{F}^n$ *extract* numerical information from vectors (*analysis*), while maps from $\mathbb{F}^n$ to X *construct* vectors from numerical information (*synthesis*). We consider both in turn. In this discussion, it is worthwhile to keep in mind the special situation when X itself is a coordinate space, in which case these two ways correspond to looking at a matrix in terms of its rows, respectively its columns.

Column maps: $\mathbb{F}^n$ into X

We have already discussed the invertible lm

$$X^n \to L(\mathbb{F}^n, X) : (v_j) \mapsto [v_1, v_2, \ldots, v_n]$$

which associates with each n-sequence $(v_1, \ldots, v_n)$ in X the column map

$$[v_1, v_2, \ldots, v_n] : \mathbb{F}^n \to X : a \mapsto \sum_j v_j a(j), \tag{4}$$

and so identifies each $V \in L(\mathbb{F}^n, X)$ as the column map $[Ve_1, \ldots, Ve_n]$.

I'll denote the number of columns of such a column map V by $\#V$. Thus,

$$\#[v_1, v_2, \ldots, v_n] = n.$$

Here, n can be any nonnegative integer, *including* 0. Specifically, there is exactly one linear map from $\mathbb{F}^0$ to X, namely the map that carries the sole element of $\mathbb{F}^0$ to $0 \in X$. This map is 1-1. For obvious reasons, I denote it by

$$[\,].$$

The following notations will be convenient for work with column maps. Let V, W be column maps *with the same target*. Then, $v \in V$ means that v is a column of V, while $V \subset M$ means that all the columns of V lie in the subset M of its target. Further, $[V, W]$ denotes the column map obtained by first using the columns of V and then the columns of W, i.e., a lm from $\mathbb{F}^{\#V+\#W}$, with $[V, w]$ the special case in which we append to V just one column, w. Note that

$$V \subset \operatorname{ran} W \quad \Longrightarrow \quad \operatorname{ran} V \subset \operatorname{ran} W. \tag{5}$$

H.P.(13) Verify that for any $V \in L(\mathbb{F}^n, X)$ and any $A \in L(X, Y)$, AV is the column map $[Av_1, Av_2, \ldots, Av_n]$.

Standard terms concerning the n-sequence $(v_j : j = 1, \ldots, n)$ correspond to rather more enlightening terms concerning the corresponding map $V = [v_1, v_2, \ldots, v_n]$: It is customary to call the elements of the range

$$\operatorname{ran} V = \{\sum_{j=1}^{n} v_j a(j) : a \in \mathbb{F}^n\}$$

of V the **linear combinations** of $(v_j : j = 1, \ldots, n)$. Note that, in this sum, I have written the scalar $a(j)$ to the *right* of the corresponding vector v_j, in order to stress the fact that formation of such a linear combination amounts to the evaluation of the linear map $[v_1, v_2, \ldots, v_n]$ at the point a. Further, it is customary to call $\operatorname{ran} V$ the **span of** $(v_j : j = 1, \ldots, n)$, and to call $(v_j : j = 1, \ldots, n)$

(i) **spanning** (for X) in case V is onto,

(ii) **linearly independent** in case V is 1-1,

(iii) a **basis** (for X) in case V is 1-1 and onto, i.e., invertible.

Since the reason for considering a sequence $(v_j : j = 1, \ldots, n)$ in X in the first place is usually one's interest in the corresponding column map $V = [v_1, v_2, \ldots, v_n]$, I will usually abandon the sequence terms 'span of', 'spanning', 'linearly independent', 'basis' for the corresponding map terms 'range', 'onto', '1-1', 'invertible'. But, I will call any invertible column map V to the linear space X a **basis for** X even though, conventionally speaking, it is the sequence of columns of V that forms the basis, rather than V itself. (In such conventional terms, an invertible column map V would be called the **basis map** for the basis formed by the columns of V.)

An invertible column map (into X) is ideal for our purposes since it associates X in a linear and 1-1 manner with a coordinate space. For an invertible $V \in L(\mathbb{F}^n, X)$, we call $V^{-1}g$ the **coordinates** of $g \in X$ wrto the basis V.

The task of solving the linear system $V? = g$ is precisely the task of expressing g as a linear combination of the columns of V, in particular the task of determining the coordinates of g with respect to $(v_j : j = 1, \ldots, n)$ in case V is invertible.

At times (e.g., when dealing with bases for a space of *multi*variate functions), it is not at all convenient or natural to *order* its elements. In such a circumstance, we associate, more generally, a finite subset M with the linear map

(6) $$[M] : \mathbb{F}^M \to X : a \mapsto \sum_{v \in M} v a(v).$$

In a way, we let the elements of the set M, i.e., the columns of the map $[M]$, index themselves. With this, $\operatorname{ran}[M] = \{\sum_{v\in M} va(v) : a \in \mathbb{F}^M\}$ is the **linear hull** of the subset M of X. If M is not finite, we would have to replace $\mathbb{F}^M$ by

$$\mathbb{F}_0^M := \{a \in \mathbb{F}^M : \#\operatorname{supp} a < \infty\}$$

since we cannot form *infinite* linear combinations without some additional structure.

**** use of a basis ****

If $[M]$ is a basis for the ls X, i.e., $[M] : \mathbb{F}_0^M \to X : a \mapsto \sum_{m\in M} m\, a(m)$ is invertible, then, for any ls U and any $A \in L(X, U)$, we have $A = A[M][M]^{-1} = [A(M)][M]^{-1}$. Conversely, for any $f \in U^M$, $[f(M)][M]^{-1}$ is a linear map from X to U, and this map depends linearly on f. This shows that the map

(7) $$U^M \to L(X, U) : f \mapsto [f(M)][M]^{-1}$$

is linear and invertible, hence provides a convenient representation for $L(X, U)$, – except that we can readily find a basis for a ls only when that space is *finitely generated*, i.e., when it is the range of some column map with *finitely* many columns.

**** construction of a basis; dimension ****

(8) Lemma. *If $V \in L(\mathbb{F}^n, X)$ is 1-1 and $x \in X$, then $[V, x]$ is 1-1 iff $x \notin \operatorname{ran} V$.*

Proof: If $x \in \operatorname{ran} V$, then $x = Va$ for some a, hence $[V, x](a, -1) = 0$, i.e., $[V, x]$ is not 1-1. Conversely, if $x \notin \operatorname{ran} V$ and $[V, x](a, b) = Va + xb = 0$, then necessarily $b = 0$ (since otherwise $x = V(-ab^{-1}) \in \operatorname{ran} V$, a contradiction), therefore already $Va = 0$, hence also $a = 0$ (since V is 1-1 by assumption). □

In particular, a 1-1 $V \in L(\mathbb{F}^n, X)$ is onto, i.e., a basis for X, if and only if it is **maximally 1-1**, i.e., for any $x \in X$, $[V, x]$ fails to be 1-1.

(9) Corollary. *If $V \in L(\mathbb{F}^n, X)$ is 1-1, and $W \in L(\mathbb{F}^m, X)$ is onto, then there exists $U \subset W$ so that $[V, U]$ is a basis (for X).*

Proof: Subject the given V and W to the following

(10) Algorithm. U `<-` $[\,]$`;` `for` $w \in W$`:` `if` $w \notin \operatorname{ran}[V, U]$`, then` U `<-` $[U, w]$`;`

At every step of this algorithm, the column map $[V, U]$ is 1-1, by (8)Lemma. Further, for the final U, all the columns of W are contained in $\operatorname{ran}[V, U]$, therefore $X = \operatorname{ran} W \subseteq \operatorname{ran}[V, U] \subseteq X$, i.e., $[V, U]$ is also onto. □

For the choice $V = [\,]$, (9)Corollary implies

(11) Corollary. *Any column map can be thinned to a basis for its range.*

In particular, any **finitely generated** ls, i.e., the range of any $[v_1, \dots, v_n]$, has a basis. Further, again by (9)Corollary, any 1-1 column map into a finitely generated space can be extended to a basis. However, we are still missing one very important fact, namely that any two (finite) bases for the same ls have the same cardinality. This follows from the following.

(12) Lemma. *If $V \in L(\mathbb{F}^n, X)$ is 1-1 and $W \in L(\mathbb{F}^m, X)$ is onto, then $n \le m$.*

Proof: Since W is onto, we can find, for each column v_j of V, some m-vector c_j so that $v_j = W c_j$. This shows that $V = WC$, with $C := [c_1, \dots, c_n] \in \mathbb{F}^{m \times n}$. If now $n > m$, then C would not be 1-1 (since any *homogeneous* linear system with more unknowns than equations always has nontrivial solutions), hence V would not be 1-1, contrary to assumption. □

H.P.(14) Give as elementary and as short a proof as you can of the basic linear algebra fact used above, that *a homogeneous linear system with more unknowns than equations always has nontrivial solutions.*

We conclude that two (finite) bases for X have the same number of columns. This number is called the **dimension** of X, and written $\dim X$. E.g., $\dim \mathbb{F}^n = n$ (since $\mathbb{F}^n \to \mathbb{F}^n : a \mapsto a$ is trivially invertible).

The **codimension** of a lss Y of X is the smallest possible dimension of a lss Z for which $X = Y + Z$. Any such smallest lss Z is a(n **algebraic) complement** of Y (in X), and necessarily also satisfies $Y \cap Z = \{0\}$, a fact denoted by

$$X = Y \dot{+} Z,$$

and this is called a **direct sum decomposition** (of X).

In these terms, (11)Corollary implies that, *for any column map V*, $\dim \operatorname{ran} V \le \#V$. For *example*, this says that $\dim \Pi_k \le k + 1$ since $\Pi_k = \operatorname{ran}[()^0, \dots, ()^k]$.

(13) Corollary. *Let X be a lss of the finite-dimensional ls Y. Then $\dim X \le \dim Y$, with equality iff $X = Y$.*

H.P.(15) Prove (13)Corollary. (The only subtle point is to show that X *has* a basis.)

(14) Corollary. *If X is a ls of dimension n, and $A \in L(X, U)$ is $\genfrac{}{}{0pt}{}{\text{onto}}{\text{1-1}}$, then $n \genfrac{}{}{0pt}{}{\ge}{\le} \dim U$, with equality in either if and only if A is both 1-1 and onto.*

H.P.(16) The map $V = [()^0, \ldots, ()^k/k!] : \mathbb{R}^{k+1} \to C(\mathbb{R})$ has Π_k as its range, by definition of Π_k. Prove that V is, in fact, 1-1. (Hint: Make up some lm $\Lambda^t : C^{(k)}(\mathbb{R}) \to \mathbb{R}^{k+1}$ for which $\Lambda^t V$ is invertible.)

As we already pointed out, the notion of (algebraic) basis extends to spaces that are not finite-dimensional. It can be shown, using Hausdorff's Maximality Theorem, that any subset M of a ls X, for which $[M]$ is 1-1, can be extended to a subset H for which $[H] : \mathbb{F}_0^H \to X : a \mapsto \sum_{h \in H} h a(h)$ is invertible. Such H (or, perhaps, the linear map $[H]$) is called a **Hamel basis** for X. Even for concrete infinite-dimensional ls's, such a basis is usually not constructible, hence is only of theoretical interest. However (see (7)), such a basis provides a ready description of $L(X, U)$ for any U and, in particular, for $U = \mathbb{F}$, where it provides the invertible linear map $\mathbb{F}^H \to X' : f \mapsto f^t$, with

$$f^t : X \to \mathbb{F} : x \mapsto \sum_{h \in H} f(h)([H]^{-1}x)(h).$$

Note that $f^t x$ is just the scalar product of f with the coordinates $[H]^{-1}x$ of x.

If now X is finite-dimensional, then $\#H < \infty$, hence $\mathbb{F}^H = \mathbb{F}_0^H$. In particular, then $\dim X' = \dim X$. But if X is not finite-dimensional, then $\dim X'$ is much larger than $\dim X$. Further, $[{e_h}^t : h \in H]$ is 1-1, hence can be extended to a basis $[H']$ for X'. With this, $X' \simeq \mathbb{F}_0^{H'}$ and $X'' \simeq \mathbb{F}^{H'}$ and, in this last correspondence, the elements of $J_0(X)$ correspond to functions on H' whose restriction to the $({e_h}^t : h \in H)$ part of H' has finite support (and determines them on all of H'). In particular, $J_0(X) \ne X''$.

H.P.(17) Use the preceding discussion to prove: *For any lss Z of any ls X and any $x \in X \backslash Z$, there exists $\lambda \in X'$ with $\lambda x = 1$ and $\lambda(Z) = \{0\}$.*

H.P.(18) Use the preceding discussion to prove: *For any lss Z of any ls X, the restriction map $X' \to Z' : \lambda \mapsto \lambda\big|_Z$ is onto, i.e., any $\mu \in Z'$ can be extended to a lfl on all of X.*

H.P.(19) Use H.P.(17) to show that $J_0 : X \to X'' : x \mapsto (J_0 x : X' \to \mathbb{F} : \lambda \mapsto \lambda x)$ is 1-1.

** basic wisdom **

All the wisdom of elementary linear algebra has been distilled into one formula:

(15) Dimension Formula. *For $A \in L(X, U)$,* $\dim \ker A + \dim \operatorname{ran} A = \dim \operatorname{dom} A.$

Proof: If $\dim \ker A \not< \infty$, then also $\dim X \not< \infty$ and there is nothing to prove. So, assume that $\dim \ker A < \infty$. Let Y be any finite-dimensional lss of X containing $\ker A$. By (9)Corollary, there is a basis $V = [W, R]$ for Y with W a basis for $\ker A$. Hence $[AW, AR]$ is onto $A(Y)$. Since $AW = 0$, it follows that AR is also onto $A(Y)$. But AR is also 1-1, since $ARa = 0$ implies that $Ra \in \ker A = \operatorname{ran} W$, i.e., $Ra = Wb$ for some b, $V(-b, a) = -Wb + Ra = 0$, therefore $(-b, a) = 0$, and so $a = 0$. Consequently, $\dim A(Y) = \#AR = \#R = \#V - \#W = \dim Y - \dim \ker A$.

If $\dim X < \infty$, then the choice $Y = X$ finishes the proof. In the contrary case, we can find subspaces Y containing $\ker A$ of as large a dimension as we wish, hence conclude that $\dim \operatorname{ran} A \not< \infty$, thus verifying the formula for this case, too. □

H.P.(20) What is the dimension of $\{f \in \Pi_2(\mathbb{R}^2) : f\big|_T = 0\}$ in case $T \subset \mathbb{R}^2$ consists of four collinear points?

As an *example*, (15) supplies the statement that, for a lss L of X', $\dim \bigcap_{\lambda \in L} \ker \lambda = \dim X - \dim L$, using in (15) the lm $A : x \mapsto (\lambda_i x : i = 1, \dots, n)$ for some basis $[\lambda_1, \dots, \lambda_n]$ of L.

The Dimension Formula also supplies the

(16) Fredholm Alternative. *For* $\dim X = \dim U < \infty$*:* *A is 1-1* $\iff$ *A is onto.*

I.e., such A is invertible iff A is either 1-1 or onto. Further, the formula shows that, for $\dim X < \dim U$, $A \in L(X, U)$ cannot be onto, while, for $\dim X > \dim U$, A cannot be 1-1.

Row maps: X into $\mathbb{F}^m$

Each linear map

$$A : X \to \mathbb{F}^m \tag{17}$$

is characterized by the m-sequence

$$\lambda_i := \delta_i A, \qquad i = 1, \dots, m, \tag{17'}$$

of lfl's on X in the sense that, given any sequence $(\lambda_i : i = 1, \dots, m)$ of lfl's on X, there is exactly one lm (17) for which (17′) holds, i.e., for which

$$\forall \{g \in X\} \qquad Ag = (\lambda_i g : i = 1, \dots, m).$$

This makes it convenient to use for it the notation

$$A = [\lambda_1, \dots, \lambda_m]^t = \Lambda^t$$

to signify this correspondence, and refer to λ_i as "the ith **row** of Λ^t" since that is exactly what λ_i is when X is itself a coordinate space and, correspondingly, Λ^t is a matrix. We call any such linear map Λ^t to some coordinate space a **row map** (or, **data map**). The resulting map

$$(X')^m \to L(X, \mathbb{F}^m) : (\lambda_i) \mapsto [\lambda_1, \lambda_2, \dots, \lambda_m]^t$$

is invertible and linear, hence $(X')^m \simeq L(X, \mathbb{F}^m)$. Note that also $(X')^m \simeq L(\mathbb{F}^m, X')$ via $(\lambda_i) \mapsto \Lambda := [\lambda_1, \dots, \lambda_m]$ and, correspondingly,

$$\forall \{c \in \mathbb{F}^m\} \;\; c^t \Lambda^t = \sum_i c(i) \lambda_i = \Lambda(c).$$

A *standard example* is provided by

$$\Lambda^t : C^{(m-1)}(\mathbb{R}) \to \mathbb{R}^m : f \mapsto (D^{i-1} f(0) : i = 1, \dots, m).$$

H.P.(21) Use the fact that $\Lambda^t V = 1$ for this Λ^t, with $V = [()^0, \dots, ()^{m-1}/(m-1)!]$, to prove the linear independence of the columns of V. Can you also deduce the linear independence of the rows of Λ^t?

The interplay between column maps and row maps

For any $V \in L(\mathbb{F}^n, X)$ and $\Lambda \in L(\mathbb{F}^m, X')$, the composition $\Lambda^t V$ is always defined. This linear map, carrying $\mathbb{F}^n$ to $\mathbb{F}^m$, is (therefore) an $m\times n$ matrix, also called the **Gramian (matrix)** of the sequence $(\lambda_i : i = 1,\ldots,m)$ in X' and the sequence $(v_j : j = 1,\ldots,n)$ in X, and is often written more explicitly

$$\Lambda^t V = \big(\lambda_i v_j\big) = (\lambda_i v_j : i = 1,\ldots,m; j = 1,\ldots,n).$$

The sequence $(\lambda_i : i = 1,\ldots,m)$ is said to be **dual** to $(v_j : j = 1,\ldots,n)$ in case their Gramian is the identity, i.e., $\Lambda^t V = 1$. The word "dual" is pleasantly short. Another common way of describing the situation that is more symmetric is to say that $(\lambda_i : i = 1,\ldots,m)$ and $(v_j : j = 1,\ldots,n)$ are **bi-orthonormal** in that case. The condition is often written out with the aid of the **Kronecker delta**:

$$\lambda_i v_j = \delta_{ij} \quad := \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

The terminology is, once again, taken from the model situation $X = \mathbb{F}^m$.

**** the inverse of a basis ****

If $V = [v_1, v_2, \ldots, v_n] \in L(\mathbb{F}^n, X)$ is invertible, then $V^{-1} \in L(X, \mathbb{F}^n)$. Its rows are the **coordinate fl**'s for the basis $(v_j : j = 1,\ldots,n)$, i.e., $\mu_i v_j = \delta_{ij}$ for $[\mu_1,\ldots,\mu_n]^t := V^{-1}$.

Let, more generally, $V = [v_1, v_2, \ldots, v_n] \in L(\mathbb{F}^n, X)$ be 1-1, hence a basis for the lss $F := \operatorname{ran} V$ of X. How does one find the coordinates of a given $f \in F$ wrto the basis V?

Offhand, we solve the linear system $V? = f$; its unique solution, c, is the coordinate vector for f. But that is not the same thing as having a concrete formula for the n-vector c in terms of f.

Of course, we can always write

$$c = V^{-1} f. \tag{18}$$

If $F = X = \mathbb{F}^n$, then V^{-1} is a matrix; in this case, (18) is an explicit formula. However, if F is only a linear subspace of some $\mathbb{F}^m$ or worse, then (18) is merely a formal expression.

Here is a recipe (in fact the *only* one available) for getting an explicit formula. It does require you to know *some* linear map Λ^t that carries F *onto* $\mathbb{F}^n$. However, the recipe works with any such map.

Take any $\Lambda \in L(\mathbb{F}^n, X')$ with $\Lambda^t(F) = \mathbb{F}^n$. Then $\Lambda^t|_F$ maps the n-dimensional ls F onto the n-dimensional ls $\mathbb{F}^n$, hence is invertible (by (14)Corollary). Therefore, the Gramian $\Lambda^t V$ must be invertible. Consequently, with $f \in F$,

$$Vc = f \iff \Lambda^t|_F V c = \Lambda^t|_F f \iff \Lambda^t V c = \Lambda^t f \iff c = (\Lambda^t V)^{-1} \Lambda^t f,$$

and the last statement is the promised formula. In effect, we have used Λ^t to convert the abstract equation $V? = f$ into the numerical equation $(\Lambda^t V)? = \Lambda^t f$.

(19) Proposition. *If V is a basis for the n-dimensional lss F of the ls X, and $\Lambda^t \in L(X, \mathbb{F}^n)$ carries F onto $\mathbb{F}^n$, then*

$$V^{-1} = (\Lambda^t V)^{-1} \Lambda^t|_F.$$

In practice, one may not know *a priori* that Λ^t maps F onto $\mathbb{F}^n$. In that case, one simply computes $\Lambda^t V$. Since V is a basis for F, $\Lambda^t(F) = \mathbb{F}^n$ iff the Gramian $\Lambda^t V$ is invertible. For *example*, with $V := [()^0, \ldots, ()^k]$ taken as a basis for $\Pi_k \subset C^{(k)}(\mathbb{R})$, choose $\Lambda = [\delta_0 D^i : i = 0, \ldots, k]$. Then $\Lambda^t V = \operatorname{diag}\lceil i! : i = 0, \ldots, k\rfloor$, hence is invertible, therefore $f = \sum_{i=0}^k ()^i (i!)^{-1} D^i f(0)$ for all $f \in \Pi_k$.

** **linear projectors** **

It follows that

$$P := V(\Lambda^t V)^{-1} \Lambda^t \tag{20}$$

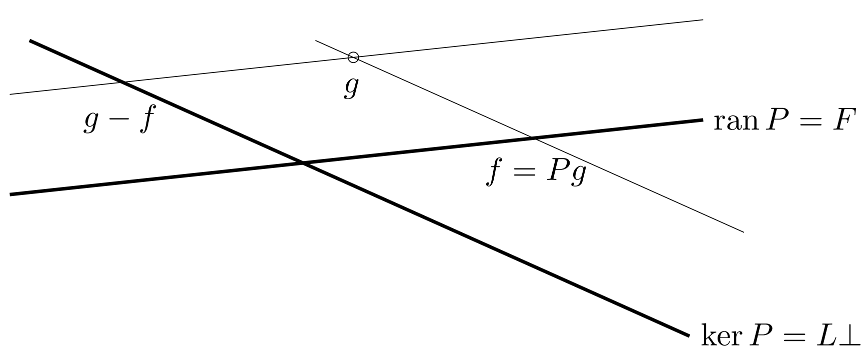
is the identity on its range, F, since $\operatorname{ran} P \subseteq \operatorname{ran} V = F \subseteq \{x \in X : Px = x\}$, while $\{x \in X : Px = x\} \subseteq \operatorname{ran} P$ is immediate, and so

$$\operatorname{ran} P = \{x \in X : Px = x\} = \operatorname{ran} V.$$

Therefore, $PP = P$, i.e., P is **idempotent** or, a linear **projector**. In particular, for any $x \in X$, $P(Px) = Px$, hence $\ker P \cap \operatorname{ran} P = \{0\}$, while $x = Px + (1-P)x$ with $(1-P)x \in \ker P$ since $P(1-P) = P - P = 0$. In other words (see (22) Figure below), $Px + (1-P)x$ is the *unique* way of writing $x \in X$ as the sum of an element from $\operatorname{ran} P$ and an element from $\ker P$. In short, we have the *direct sum decomposition* of X,

$$X = \operatorname{ran} P \dot{+} \ker P = \operatorname{ran} P \dot{+} \operatorname{ran}(1-P). \tag{21}$$

The second equality uses the fact that $\ker P = \operatorname{ran}(1-P)$ since, as already observed, $\operatorname{ran}(1-P) \subseteq \ker P$, while $(1-P) = 1$ on $\ker P$, hence also $\ker P \subseteq \operatorname{ran}(1-P)$. Note that also $1-P$ is a lprojector.



(22) Figure. Interpolation and linear projector

Further note that, directly from (20), $PV = V$ and $\Lambda^t P = \Lambda^t$. The latter implies that $\ker P \subset \ker \Lambda^t$, while (20) implies $\ker P \supset \ker \Lambda^t$. Therefore $\ker P = \ker \Lambda^t$. With (21), this implies that, *for any $g \in X$, Pg is the unique element $f \in \operatorname{ran} P$ that* **interpolates** *g in the sense that $\Lambda^t f = \Lambda^t g$.*

This (as yet nonstandard) language derives from the standard example of polynomial interpolation (see (46)), in which $X = C[a\,..\,b]$, $V = [()^{j-1} : j = 1, \ldots, n]$, and $\Lambda = [\delta_{t_i} : i = 1, \ldots, n]$ for some n-set $\{t_1, \ldots, t_n\}$ in the interval $[a\,..\,b]$, hence $Pg = V(\Lambda^t V)^{-1}\Lambda^t g$ is the unique polynomial of degree $< n$ that matches or interpolates g at the points $t_1, \ldots, t_n$.

H.P.(22) Prove: *if* $\ker \Lambda^t \cap F = \{0\}$ *for some* $\Lambda \in L(\mathbb{F}^n, X')$ *and some n-dim. lss F of X, then* $P := (\Lambda^t\big|_F)^{-1}\Lambda^t$ *is a well-defined lm on X, with* $\operatorname{ran} P = F$, $\ker P = \ker \Lambda^t$, *and* $P^2 = P$.

** factorization and rank **

In order to compute with $A \in L(X, U)$, we have to factor it through a coordinate space, i.e., we have to write it as $A = V\Lambda^t$ for some column map V. We call $\#V$ the **order** of this factorization.

H.P.(23) Use the familiar map D of differentiation, say as an element of $L(\Pi_k)$, to illustrate the point just made that one must factor a linear map through coordinate space in order to be able to compute with it.

The smaller the order of the factorization $A = V\Lambda^t$, the cheaper the calculation of Ag via $V\Lambda^t g = \sum_j v_j\, \lambda_j g$. But there is a limit to how small we can make the order. The smallest possible order is called the **rank** of A and the corresponding factorization is called **minimal**. In these terms, each linear map of the form $[v]\lambda$ with $v \in Y\backslash 0$ and $\lambda \in X'\backslash 0$ is a rank-one map, from X to Y, and the rank of $A \in L(X, Y)$ is the smallest number of terms in any sum of rank-one maps that equals A.

H.P.(24) Prove that the rank of the lm $X \to U : x \mapsto 0$ is zero.

(23) Proposition. *$A = V\Lambda^t$ is minimal if and only if V is a basis for* $\operatorname{ran} A$. *In particular,*

$$\operatorname{rank} A = \dim \operatorname{ran} A.$$

Proof: For any factorization $A = V\Lambda^t$, $\operatorname{ran} A \subseteq \operatorname{ran} V$, hence

$$\dim \operatorname{ran} A \le \dim \operatorname{ran} V \le \#V,$$

with equality in the first '$\le$', by (13)Corollary, iff $\operatorname{ran} A = \operatorname{ran} V$, and in the second '$\le$' iff V is 1-1. Thus, $\dim \operatorname{ran} A \le \#V$, with equality iff V is a basis for $\operatorname{ran} A$. □

It follows that any A with $\dim \operatorname{ran} A < \infty$ *has* a (minimal) factorization since, for any basis V for $\operatorname{ran} A$, $A = V(V^{-1}A)$.

The dual of a linear map

Each linear map $A \in L(X,U)$ induces a map, called its **dual** and denoted by A', by the prescription
$$A' : U' \to X' : \lambda \mapsto \lambda A.$$

H.P.(25) Let $A \in L(X,U)$. Verify: (i) $A' \in L(U',X')$; (ii) if $C \in L(Y,X)$, then $C'A'$ is defined and equals $(AC)'$ (what if you only know that $C \in L(Y,W)$ for some $W \subseteq X$?); (iii) if A is a matrix, i.e., both X and U are coordinate spaces, then A' is the **transpose** of A.

The dual of a linear map is of interest because of the connections that exist between range and kernel of a lm and those of its dual. For example, for any $A \in L(X,U)$, $\lambda A = 0$ iff $\forall\{x \in X\}\ (\lambda A)x = 0$ iff $\forall\{x \in X\}\ \lambda(Ax) = 0$ iff $\lambda(\operatorname{ran} A) = \{0\}$. In particular, if A is onto, then A' is necessarily 1-1. More subtle conclusions of this kind (e.g., (27), (29), (32) below) make use of the following notion of orthogonality between elements of a ls and elements of its dual.

**** orthogonality ****

Because of the special case $(\mathbb{R}^n)' \simeq \mathbb{R}^n$, we say that $\lambda \in X'$ and $f \in X$ are **orthogonal** to one another in case $\lambda f = 0$, and write this
$$\lambda \perp f.$$

More generally, for $L \subseteq X'$, we denote by
$$L\perp := \{f \in X : \forall \lambda \in L \;\; \lambda \perp f\} = \bigcap_{\lambda \in L} \ker \lambda =: \ker L$$
the **kernel** of L, and, for $F \subseteq X$, by
$$\perp F := \{\lambda \in X' : \forall f \in F \;\; \lambda \perp f\} \;=\; \{\lambda \in X' : \ker \lambda \supseteq F\}$$
the **annihilator** of F. For example (as just observed), for $A \in L(X,U)$,

(24) $$\ker A' = \{\lambda \in U' : \lambda A = 0\} = \perp \operatorname{ran} A.$$

The complementary assertion: $\operatorname{ran} A' = \perp \ker A$ requires, in general, Hausdorff's Maximality Theorem; see the proof of (29)Proposition below.

The notations $L_\perp$ and L_0 for $L\perp$ are quite common, as are the notations $F^\perp$ and F^0 for $\perp F$. The notation used here reflects the fact that, in applying $\lambda \in X'$ to $x \in X$, we write λx, i.e., write the lfl to the left of the element it is being applied to. In particular, $\perp N$ is the set of lfl's (on whatever ls N is a subset of) that vanish on N, while $N\perp$ is the set of elements (of whatever ls the elements of N are defined on) on which all the elements of N vanish.

H.P.(26) Verify that $L\perp$ and $\perp F$ are lss's (but see H.P.(31)).

While it is obvious that
$$\perp X = \{0\},$$

the assertion

$$X'\perp = \{0\} \tag{25}$$

is obvious only when X is a function space, $X \subseteq \mathbb{F}^T$ say, for then

$$X' \perp f \quad \Longrightarrow \quad \forall\{t \in T\} \;\; f(t) = \delta_t f = 0.$$

For more abstract spaces, an application of Hausdorff's Maximality Theorem is usually needed to verify that X' is rich enough to distinguish between elements of X, i.e., for (25) to hold. Here is the basic claim.

(26) Proposition. *For any $F \subset X$,*

$$(\perp F)\perp \supseteq F,$$

with equality iff F is a lss.

Proof: The containment is immediate as is the claim that equality implies that F is a lss. For the converse, assume that F is a lss and let $x \in X \backslash F$. Then, by H.P.(17) (which uses Hamel bases), there exists $\lambda \in \perp F$ with $\lambda x = 1$, hence $x \notin (\perp F)\perp$. □

This, together with (24), implies that

(27) $\forall\{A \in L(X,U)\}$ $\ker A'\perp = \operatorname{ran} A$; in particular: *$A'$ is 1-1* $\iff$ *A is onto.*

H.P.(27) Prove (25) and (27).

** the duals of row maps and column maps **

For $\Lambda \in L(\mathbb{F}^m, X')$, and with the identification $\mathbb{F}^m \simeq (\mathbb{F}^m)'$ via $a \mapsto a^t$, we have

$$(\Lambda^t)' : (\mathbb{F}^m)' \simeq \mathbb{F}^m \to X' : c \mapsto c^t\Lambda^t : x \mapsto \sum_i c(i)\lambda_i x = (\Lambda c)x,$$

hence $(\Lambda^t)' = \Lambda$. Also, $\Lambda' \in L(X'', \mathbb{F}^m)$, and $\Lambda' J_0 = \Lambda^t$. In particular, we can think of Λ^t as the restriction of Λ' to $J_0(X) \simeq X$. In this sense, $\Lambda' = \Lambda^t$ if and only if $\dim X < \infty$.

Further, with the same identification $\mathbb{F}^m \simeq (\mathbb{F}^m)'$, we get for $V = [v_1, v_2, \ldots, v_n] \in L(\mathbb{F}^n, X)$ that $V' : \lambda \mapsto \lambda V = [\lambda v_1, \lambda v_2, \ldots, \lambda v_n] = (\lambda v_i : 1, \ldots, n)^t$, hence we conclude that the 'rows' of V' are the lfl's v_j'', i.e., the v_j as they act on X' via $v_j'' = J_0 v_j : \lambda \mapsto \lambda v_j$, i.e.,

$$V' = [J_0 v_1, \ldots, J_0 v_n]^t, \qquad \text{therefore } \; V'' = [v_1'', v_2'', \ldots, v_n''] = J_0 V. \tag{28}$$

** use of minimal factorization **

Since $(\Lambda^t)' = \Lambda$, we observe that $A = V\Lambda^t$ iff $A' = \Lambda V'$, hence conclude that, in case $X = \operatorname{dom} A$ is finite-dimensional, hence $X'' \simeq X$, the factorization $A = V\Lambda^t$ is minimal iff $A' = \Lambda V'$ is minimal. This implies that

$$\dim \operatorname{ran} A = \operatorname{rank} A = \operatorname{rank} A' = \dim \operatorname{ran} A',$$

and, with (23)Proposition, that $A = V\Lambda^t$ is minimal iff Λ is a basis for $\operatorname{ran} A'$.

For example, (20) provides a minimal factorization for P since V is 1-1; hence $\Lambda = (\Lambda^t)'$ is necessarily a basis for $\operatorname{ran} P'$.

H.P.(28) Let $A \in L(X, U)$, and $\dim \operatorname{ran} A = n$. Prove that $A = V\Lambda^t$ for some $V \in L(\mathbb{F}^n, U)$ and some $\Lambda \in L(\mathbb{F}^n, X')$, both 1-1.

(29) Proposition. *For any* $A \in L(X, U)$, $\operatorname{ran} A' = \perp \ker A$.

Proof: For any $\lambda \in U'$, $\lambda A \perp \ker A$, hence we only need to prove that

$$\operatorname{ran} A' \supseteq \perp \ker A. \tag{30}$$

Further, if X is finite-dimensional, A has a minimal factorization, $A = V\Lambda^t$ say. But then, also $A = \Lambda V'$ is a minimal factorization and, by (23)Proposition, V is 1-1, hence $Ax = 0$ iff $\Lambda^t x = 0$, i.e., $\ker A = \ker \Lambda^t$, while Λ is a basis for $\operatorname{ran} A'$, and (31)Lemma below finishes the proof, since it shows that $\mu \perp \ker \Lambda^t$ implies that $\mu \in \operatorname{ran} \Lambda = \operatorname{ran} A'$.

In the general case, if $\mu \perp \ker A$, then $\ker(\mu) \supset \ker A$, hence, by (3) Factor Lemma and the discussion following it, the map $(A_|)^{-1}A : X \to X/\ker A$ is a factor of μ, i.e., $\mu = \nu A_|^{-1} A$, with $\nu (A_|)^{-1} \in (\operatorname{ran} A)'$, while, by H.P.(18), ν can be extended to a lfl λ on all of U. □

Here is a special case of (30) of independent interest which can be proved without use of Hamel bases.

(31) Lemma. *For* $\mu, \lambda_1, \ldots, \lambda_m \in X'$: $\mu \in \operatorname{ran}[\lambda_1, \lambda_2, \ldots, \lambda_m] \iff \ker \mu \supseteq \bigcap_{i=1}^m \ker \lambda_i$.

H.P.(29) Give an example to show that having *finitely* many λ_i's here is essential. (Hint: Try $X = C[0..1]$.)

Proof: With $\Lambda := [\lambda_1, \lambda_2, \ldots, \lambda_m]$, we have $\ker \Lambda^t = \cap_i \ker \lambda_i$, hence we want to show that $\ker \mu \supset \ker \Lambda^t$ implies that $\mu \in \operatorname{ran} \Lambda$. We make that job only harder if we omit some λ_j. So, after omitting any λ_j for which $\ker \lambda_j \supseteq \bigcap_{i \neq j} \ker \lambda_i$, we may assume, without loss of the condition $\ker \mu \supset \cap_i \ker \lambda_i$, that, for all j, $\ker \lambda_j \not\supseteq \bigcap_{i \neq j} \ker \lambda_i$, i.e., for all j, $\big(\cap_{i \neq j} \ker \lambda_i\big) \backslash \ker \lambda_j \neq \{\}$. This means that, for each j, there exists $v_j \in X$ so that $\lambda_i v_j = 0$ for all $i \neq j$, while $\lambda_j v_j \neq 0$, hence, after dividing v_j by $\lambda_j v_j$, all j, we have $\Lambda^t V = 1$, hence $P := V\Lambda^t$ is a lprojector. Since $\operatorname{ran}(1 - P) = \ker P = \ker \Lambda^t \subseteq \ker \mu$, we have $\mu(1 - P) = 0$, hence $\mu = \mu P$, and so $\mu = \mu V \Lambda^t \in \operatorname{ran} \Lambda$. □

(32) Proposition. *For any* $A \in L(X, U)$, *A is* $\begin{smallmatrix}\text{1-1}\\ \text{onto}\end{smallmatrix}$ *iff A' is* $\begin{smallmatrix}\text{onto}\\ \text{1-1}\end{smallmatrix}$. *In particular, A is invertible iff A' is invertible.*

Proof: By (29), A' is onto iff $\perp \ker A = X'$, i.e., by H.P.(17), iff $\ker A = \{0\}$. The other equivalence was already observed in (27). □

H.P.(30) Prove: $A = V\Lambda^t$ *is minimal iff* V *is 1-1 and* Λ^t *is onto.*

H.P.(31) Prove that, for a finite-dimensional lss L of X', $\perp(L\perp) = L$. Show, by an example, that $\dim L < \infty$ is needed here (Hint: H.P.(17) or H.P.(29)).

** tests for linear independence **

(33) Corollary. *For* $\Lambda \in L(\mathbb{F}^m, X')$, *$\Lambda$ is 1-1 iff* $\exists \{V \in L(\mathbb{F}^m, X)\}$ $\Lambda^t V$ *is invertible.*

Proof: By (32)Proposition, $\Lambda = (\Lambda^t)'$ is 1-1 iff Λ^t is onto, while Λ^t, being linear, is onto iff it has a linear right inverse. □

H.P.(32) Prove that, for any sequence $t_0 < \cdots < t_k$, the sequence $\delta_{t_0}, \ldots, \delta_{t_k}$ is linearly independent over (i.e., as functions on) Π_k. (Hint: Try $v_j := \prod_{i<j}(\cdot - t_i)$ or $\ell_j := \prod_{i\neq j}(\cdot - t_i)$.) Conclude that $\dim \Pi_k = k+1$.

H.P.(33) Let $\Lambda^t : \Pi_2(\mathbb{R}^2) \to \mathbb{R}^4 : p \mapsto (p(a), p(b), p(c), p(d))$. Prove that Λ is 1-1 (i.e., the four point evaluations are linearly independent on $\Pi_2(\mathbb{R}^2)$) iff the four points $a, b, c, d \in \mathbb{R}^2$ do not all lie on the same straight line. (Hints: Consider products $pq : x \mapsto p(x)q(x)$, with both p and q a linear polynomial that vanishes on two of the four points. Also, if the four points all lie on some straight line, ℓ say, then $\dim \Lambda^t(\Pi_2(\mathbb{R}^2)) \le \dim \Pi_2(\mathbb{R}^2)\big|_\ell$.)

(34) Corollary. *$V \in L(\mathbb{F}^n, X)$ is 1-1 iff $\exists\{\Lambda \in L(\mathbb{F}^n, X')\}$ $\Lambda^t V$ is invertible.*

Proof: By (32)Proposition, V is 1-1 iff V' is onto, and this is equivalent to having some $\Lambda \in L(\mathbb{F}^n, X')$ for which $V'\Lambda$ is invertible, i.e., by (32)Proposition, for which $\Lambda^t V = (V'\Lambda)'$ is invertible.

H.P.(34) Let $t_1 < t_2 < \cdots$. Prove that the sequence $v_j := \prod_{i=j}^{j+k-1}(\cdot - t_i)$, $j = 1, \ldots, k+1$, is linearly independent. (Hint: The most versatile matrix class in which invertibility is easily checked is the class of triangular matrices.)

H.P.(35) Let $A \in L(X, Y)$ be such that both $k := \dim\ker A$ and $r := \dim\ker A'$ are finite, and let V be a basis for $\ker A$ and M a basis for $\ker A'$.

(a) Prove that there is a column map Λ (into X') dual to V and a column map W (into Y) dual to M.

(b) Show that $\operatorname{ran} A \cap \operatorname{ran} W = \{0\}$ and that $\ker A \cap \ker \Lambda^t = \{0\}$, and that $\operatorname{ran} A = \ker \mathrm{M}^t$.

(c) Prove that the lm

$$\hat{A} := \begin{bmatrix} A & W \\ \Lambda^t & 0 \end{bmatrix} : X \times \mathbb{F}^r \to Y \times \mathbb{F}^k : (x, \alpha) \mapsto (Ax + W\alpha, \Lambda^t x)$$

is 1-1 and onto.

Application: approximate evaluation of linear functionals; interpolation

Since functional information is what we usually have about a function, Numerical Analysis is much concerned with the following

(35) Problem. *Given $\Lambda^t g$ for some $\Lambda \in L(\mathbb{F}^m, X')$, what can be said about μg for $\mu \in X'$?*

(36) Example. If we know $g(a), Dg(a), \ldots, D^k g(a)$, then we have learned to think that we have a good idea of what $g(t)$ is for t near a from the **truncated Taylor series**:

$$\mu g := g(t) \sim g(a) + Dg(a)(t-a) + \cdots + D^k g(a)(t-a)^k/k! =: \sum c(i)\lambda_i g = c^t \Lambda^t g.$$

(37) Example. If we know $g(a), g(a+h), g(a+2h), \ldots, g(b)$ with $h := (b-a)/N$, then we have learned to think that we get some idea about $\int_a^b g(t)\,\mathrm{d}t$ from the **composite trapezoidal rule**:

$$\mu g := \int_a^b g(t)\,\mathrm{d}t \sim \big(g(a)/2 + g(a+h) + g(a+2h) + \cdots + g(b)/2\big)h =: \sum c(i)\lambda_i g = c^t \Lambda^t g.$$

**** inadequacy of rules ****

In fact, in both examples, the approximation may be way off since it incorporates only a *finite* amount of *linear* information about g. To make this precise, look at it abstractly. All we know about g is the vector $\Lambda^t g$, i.e., that $g \in (\Lambda^t)^{-1}\{\Lambda^t g\} = g + \ker \Lambda^t$, hence

$$\mu g \ \in \ \mu g + \mu(\ker \Lambda^t).$$

There are just two cases.

If $\mu \in \operatorname{ran} \Lambda$, then $\mu = \Lambda(c) = c^t \Lambda^t$ for some c and, for that c, $c^t \Lambda^t g$ provides the exact value for μg. Correspondingly, by (31)Lemma, $\mu(\ker \Lambda^t) = \{0\}$ in this case.

In the contrary case,

$$\mu \ \notin \ \operatorname{ran} \Lambda, \tag{38}$$

and, by (31)Lemma, then $\mu x \neq 0$ for some $x \in \ker \Lambda^t$, hence now $\mu(\ker \Lambda^t) = \mathbb{F}$ and $\Lambda^t g$ tells us *nothing* about μg.

In both examples, (38) holds. By (31)Lemma, this is demonstrated by making up a function g for which $\Lambda^t g = 0$, while $\mu g \neq 0$.

For (36)Example, take $g := (\cdot - a)^{k+1}$.

For (37)Example, take $g := (\cdot - a)^2 (\cdot - a - h)^2 (\cdot - a - 2h)^2 \cdots (\cdot - b)^2$.

**** rule construction ****

In both examples, the approximation $\Lambda(c) = c^t \Lambda^t$ to μ is customarily derived using *interpolation*, an idea that goes back to Newton (at least): The **rule for** μ is determined as the particular element

$$\lambda := c^t \Lambda^t = \Lambda(c)$$

of $\operatorname{ran} \Lambda$ that matches μ at certain elements $v_1, \ldots, v_n$ of X. Since both μ and Λ^t are *linear* maps, this means that $\lambda V = \mu V$ for $V := [v_1, \ldots, v_n]$, or

$$c^t \Lambda^t V = \mu V. \tag{39}$$

The rule in (36)Example can be obtained by using for V a basis for Π_k.

H.P.(36) Give a choice of V that results in the rule in (37)Example (and prove that it works).

Assume that $\Lambda^t V$ is invertible. Then (39) has exactly one solution,

$$[c(1), \ldots, c(n)] = c^t = \mu V (\Lambda^t V)^{-1},$$

hence the resulting rule is

$$\lambda = c^t \Lambda^t = \mu V (\Lambda^t V)^{-1} \Lambda^t. \tag{40}$$

What if the Gramian $\Lambda^t V$ is not invertible? This can have many causes. E.g., if Λ is not 1-1, then Λ^t will fail to be onto, hence $\Lambda^t V$ cannot be onto. But this is an avoidable failure. After all, we are not interested in the coefficient vector c. Rather, we are interested in finding some $\lambda \in \operatorname{ran} \Lambda$ that agrees with μ on the v_j's. Also, having $\lambda V = \mu V$ is equivalent to having

$$\lambda v = \mu v \quad \text{for all } v \in \operatorname{ran} V.$$

Thus the rule construction task does not depend on the Gramian $\Lambda^t V$, but only on the lss's

$$L := \operatorname{ran} \Lambda \qquad \text{and} \quad F := \operatorname{ran} V,$$

and reads in such terms as follows:

(41) Rule Construction Problem (L,F)**.** *Given the lss's* $L \subseteq X'$ *and* $F \subseteq X$*, determine, for given* $\mu \in X'$*, a* $\lambda \in L$ *so that*

(42) $$\lambda - \mu \ \perp\ F.$$

Call this problem **correct** if it has exactly one solution λ for every $\mu \in X'$.

(43) Lemma. *Let* V *be any basis for the lss* F *of* X*, and let* Λ *be any basis for the lss* L *of* X'*.*

Then, the RCP(L,F) is correct iff the Gramian $\Lambda^t V$ *is invertible.*

Proof: Since V and Λ are bases, the RCP(L,F) is correct iff the linear system

(44) $$?^t \Lambda^t V = \mu V$$

has exactly one solution for every $\mu \in X'$.

Since V is 1-1, the map $\mu \mapsto \mu V$ is onto (by (32)Proposition). Hence, the RCP(L,F) has a solution for every μ iff the equation $?^t \Lambda^t V = z^t$ has a solution for every $z \in \mathbb{F}^n$, i.e., iff $c \mapsto c^t \Lambda^t V$ is onto $\mathbb{F}^{1\times n}$.

Since Λ is 1-1, we have $c^t \Lambda^t = 0 \iff c = 0$. Hence, the RCP$(L,F)$ has at most one solution iff (44) has at most one solution, i.e., iff $c \mapsto c^t \Lambda^t V$ is 1-1.

Thus, the RCP(L,F) is correct iff $\Lambda^t V$ is invertible. □

H.P.(37) Let $V = [()^0, ()^1, ()^2]$, $c := (a+b)/2$. For each of the following choices of Λ, determine whether or not the RCP(ran Λ,ran V) is correct: (a) $\Lambda = [\delta_a, \delta_b, \delta_c]$; (b) $\Lambda = [\delta_a, \delta_b, (\delta_a + \delta_b)/2]$; (c) $\Lambda = [\delta_a, \delta_b, ((\delta_a + \delta_b)/2)D]$.

**** interpolation ****

The solution (44) comes to us in the striking form

$$\lambda = \mu P,$$

with $P := V(\Lambda^t V)^{-1}\Lambda^t$ the linear projector (20) we encountered while providing a formula for the inverse of a basis (see (19)). We noted there that Pg solves the

(45) Linear Interpolation Problem (F,L)**.** *Determine, for given* $g \in X$*, an* $f \in F$ *that* **agrees with** g **on** L *in the sense that*

$$L \ \perp\ g - f.$$

We call the LIP(F,L) **correct** if it has exactly one solution for every $g \in X$. This will happen exactly when its **dual** problem, the RCP(L,F), is correct. In these terms, a rule provides an approximation to μg by applying μ to the interpolant Pg for g.

Remark. A linear projector is customarily characterized by its range and its kernel, because of the direct sum decomposition $X = \operatorname{ran} P \dot{+} \ker P$ mentioned earlier. I (as a numerical analyst) prefer to characterize such a linear projector by its range and the range of its dual, $\operatorname{ran} P' = \perp \ker P = L = \operatorname{ran}\Lambda$, since $\operatorname{ran} P'$ consists of all the $\lambda \in X'$ for which $\lambda = \lambda P$, i.e., on which g and Pg agree for every $g \in X$. For that reason, I will refer to $\operatorname{ran} P'$ as the set of **interpolation functionals** for P, while $\operatorname{ran} P$ is its set of (possible) interpolants.

H.P.(38) Prove that, with $V \in L(\mathbb{F}^n, X)$ and $\Lambda \in L(\mathbb{F}^n, X')$, *the lm* $V(\Lambda^t V)^{-1}\Lambda^t$ *(if defined) only depends on* $\operatorname{ran} V$ *and* $\operatorname{ran}\Lambda$.

H.P.(39) Let μ be a lfl on functions on some domain in $\mathbb{R}^d$. One says that the rule $\lambda = \sum_{t\in T} w(t)\delta_t$ is of **degree** k, or, has **precision** k if $\lambda = \mu$ on $\Pi_k(\mathbb{R}^d)$. Such a rule of degree k is called **interpolatory** if it is the only rule of degree k for μ based on the point set T. Example: $d = 1$, $\mu = \int_{-1}^{1} \cdot$, $\lambda = 2\delta_0$ the Midpoint rule, hence $k = 1$.

Show that the adjective 'interpolatory' is appropriate by proving that an interpolatory rule of degree k for μ is necessarily of the form μP for some linear projector with interpolation functionals $\{\delta_t : t \in T\}$, i.e., with $\operatorname{ran} P' = \operatorname{ran}[\delta_t : t \in T]$. (Hint: (32)Proposition and (33)Corollary.)

(46) Example. The *standard example* is **polynomial interpolation**: $X = C[a\,..\,b]$, $v_j = ()^{j-1}, \lambda_i = \delta_{t_i}$ (with $t_i \neq t_j$ for $i \neq j$), $i, j = 1, \ldots, n$. This LIP is correct since, e.g., with $\hat{v}_j := \prod_{i\neq j}(\cdot - t_i)$, $j = 1, \ldots, n$, the column map $\hat{V}$ maps into $\Pi_{<n} = \operatorname{ran} V = F$ and the Gramian $\Lambda^t\hat{V}$ is invertible (by inspection, since it is diagonal with nonzero diagonal elements), hence $\hat{V}$ must be 1-1, and, since $\#\hat{V} = \#V$ and $\operatorname{ran}\hat{V} \subseteq \operatorname{ran} V$, it follows that $\operatorname{ran}\hat{V} = \operatorname{ran} V$. It follows that the Gramian $\Lambda^t V$ is also invertible; it is called the **Vandermonde** matrix.

Polynomial interpolation is the workhorse of numerical approximation. All the standard rules for numerical integration and differentiation use it.

H.P.(40) Prove: *For every lss* F *of* $\mathbb{R}^T$ *of dimension* n *there exists* $(t_i)_1^n$ *in* T *so that the RCP(*$\operatorname{ran}[\delta_{t_1}, \ldots, \delta_{t_n}], F$*) is correct, i.e., so that the LIP(*$F, \operatorname{ran}[\delta_{t_1}, \ldots, \delta_{t_n}]$*) is correct.* (Hint: You might prove first that, if P is the lprojector corresponding to a correct LIP($F, \operatorname{ran}[\delta_{t_1}, \ldots, \delta_{t_n}]$), then $g \notin F$ implies that $g - Pg \neq 0$. This is useful in an inductive argument. I don't know how to do this homework *without* induction.)

H.P.(41) Prove: *if* $T \subset \mathbb{R}^2$ *lies on no conic section, then some subset* U *of* T *with* $\#U = 6$ *lies on no conic section. (Here,* **conic section** *is, by definition, the zeroset of any polynomial of exact degree 2.)*

** numerics **

Whether constructing rules or interpolants, we have to evaluate an expression (either μP or Pg) that involves the inverse of the Gramian $\Lambda^t V$. This is invariably done by **factoring** the Gramian,

$$\Lambda^t V \;=\; AC$$

say, with A and C square matrices, hence (why?) invertible, and usually more easily invertible than $\Lambda^t V$, e.g., triangular. Then

$$\hat{\Lambda}^t\hat{V} := (A^{-1}\Lambda^t)(VC^{-1}) \;=\; A^{-1}ACC^{-1} = 1,$$

with $\hat{\Lambda} = (A^{-1}\Lambda^t)' = \Lambda(A^{-1})'$, $\hat{V} = VC^{-1}$ again bases for L, F respectively. They are special, though, in that they are dual to each other. Therefore, P takes the simple form

$$P = \hat{V}\hat{\Lambda}^t = \sum_j [\hat{v}_j]\hat{\lambda}_j.$$

For the standard example (46), the linear projector of polynomial interpolation has several standard representations $P = \sum_j [\hat{v}_j]\hat{\lambda}_j$ corresponding to the different ways the

Gramian is factored as $\Lambda^t V = AC$. For example, since $Pg = \sum_j \ell_j g(t_j)$ (the **Lagrange form**), with

$$\ell_j := \prod_{i\neq j} \frac{\cdot - t_i}{t_j - t_i}$$

the jth **Lagrange polynomial**, therefore necessarily $[\ell_1, \dots, \ell_n] = V(\Lambda^t V)^{-1}$, i.e., the inverse of the Vandermonde matrix provides the power form for the Lagrange polynomials, corresponding to the choice $A = 1$, hence $C = \Lambda^t V$. If we choose, instead, AC to be the LU factorization of $\Lambda^t V$ into a lower triangular A and a *unit* upper triangular C (as would be obtained when applying Gauss elimination to the interpolation equations $\Lambda^t V? = (g(t_j))$), the resulting form $P = \sum_j [\hat v_j]\hat\lambda_j$ is the **Newton form**: Now

$$\hat v_j := \prod_{i<j} (\cdot - t_i)$$

(since C is unit upper triangular, hence so is C^{-1}, hence $\hat v_j$, as the jth column of VC^{-1}, has the leading term t^{j-1}, while $\Lambda^t \hat V = \Lambda^t V C^{-1} = A$ is lower triangular, hence $\hat v_j$ must vanish at $t_1, \dots, t_j$), and

$$\hat\lambda_j g =: \delta_{t_1,\dots,t_j} g \qquad (47)$$

is, by definition, the **divided difference of** g **at** $t_1, \dots, t_j$.

H.P.(42) Verify that the notation $\delta_{t_1,\dots,t_i}$ reflects the situation accurately, i.e., that the above $\hat\lambda_i g$ depends only on g and $t_1, \dots, t_i$. Also verify that $\hat\lambda_i$ vanishes on Π_{i-2} and that $\hat\lambda_i ()^{i-1} = 1$ (as would be expected of the divided difference at $t_1, \dots, t_i$).

H.P.(43) Derive from the above definition of $\delta_{t_1,\dots,t_k}$ the standard recurrence for the divided difference (which accounts for its name).

(48) Example. An equally important *example* is **least-squares approximation** in which Pg is chosen from F so that the error $g - Pg$ be perpendicular to F. E.g., if $X = C[a\,..\,b]$, this means that

$$\forall\{f \in F\} \qquad \int_a^b f(t)\big(g - Pg\big)(t)\,\mathrm{d}t \;=\; 0.$$

In other words, the collection L of interpolation functionals consists of all lfl's of the form

$$\int f\cdot \;: X \to \mathbb{R} : g \mapsto \int f(t)g(t)\,\mathrm{d}t$$

for some $f \in F$.

If V is a basis for F, then $\Lambda = [\dots, \int v_j \;\cdot\;, \dots]$ is a basis for L, and the Gramian $\Lambda^t V = \big(\int v_i v_j\big)$ is the coefficient matrix for the **normal equations**

$$\sum_j \int v_i v_j a(j) \;=\; \int v_i g, \qquad i = 1, \dots, n.$$

H.P.(44) Prove that the LIP(F, L) with $F \subseteq C[a\,..\,b]$, $\dim F < \infty$, and $L := \{\int f\cdot : f \in F\}$ is correct. (Hint: With V a basis for F, prove that the Gramian $\Lambda^t V = (\int v_i v_j)$ is 1-1 by considering $c^t \Lambda^t V c$.)

The application of Gauss elimination to this system is, in effect, **Gram-Schmidt-orthogonalization**: On factoring $\Lambda^t V$ as AC, but this time with $C = A'$ (which is possible since $\Lambda^t V$ is symmetric and positive definite), we obtain a new basis $\hat{V}$ for F and a new basis $\hat{\Lambda}$ for L, and these are dual to each other. Since also $\hat{\lambda}_i = \int \hat{v}_i\cdot$ (since $C = A'$), it follows that $\int \hat{v}_i \hat{v}_j = \delta_{ij}$, i.e., $(\hat{v}_j)$ is an **orthonormal** sequence.

