A formula for the coordinate map

Let $V \in L(\mathbb{F}^n, X)$ be a basis for the vector space X. How do we find the coordinates

$$a = V^{-1}x \tag{5.3}$$

for given $x \in X$?

Offhand, we solve the (linear) equation $V? = x$ for a. Since V is a basis, we know that this equation has exactly one solution. But that is not the same thing as having a concrete formula for a in terms of x.

If $X = \mathbb{F}^n$, then V^{-1} is a matrix; in this case, (5.3) is an explicit formula. However, even if $X \subset \mathbb{F}^n$ but $X \neq \mathbb{F}^n$, then (5.3) is merely a formal expression.

(5.4) Example: If V is a basis for some linear subspace X of $\mathbb{F}^n$, then we can obtain a formula for V^{-1} via elimination as follows.

Act as if V were invertible, i.e., apply elimination to $[V, \mathrm{id}_n]$. Let $r := \#V$. Since V is 1-1, the first r columns in $[V, \mathrm{id}_n]$ are bound, hence we are able to produce, via elimination, an equivalent matrix R for which $R(\mathbf{q}, 1{:}r) = \mathrm{id}_r$, for some r-sequence $\mathbf{q}$. Since we obtain R from $[V, \mathrm{id}_n]$ by (invertible) row operations, we know that $R = M[V, \mathrm{id}_n] = [MV, M]$ for some invertible matrix M. In particular,

$$\mathrm{id}_r = R(\mathbf{q}, 1{:}r) = (MV)(\mathbf{q}, :) = M(\mathbf{q}, :)V,$$

showing $M(\mathbf{q}, :) = R(\mathbf{q}, r + (1{:}n))$ to be a left inverse for V, hence equal to V^{-1} when restricted to $\operatorname{ran} V$.

Suppose, in particular, that we carry elimination all the way through, to obtain $R = \mathrm{rref}([V, \mathrm{id}_n])$. Then, $\mathbf{q} = 1{:}r$ and, with $r + \mathtt{b}$ and $r + \mathtt{f}$ the bound and free columns of $[V, \mathrm{id}_n]$ other than the columns of V, we necessarily have $M(\mathbf{q}, \mathtt{b}) = 0$, hence, for this choice of M, we get

$$V^{-1}x = M(\mathbf{q}, :)x = M(\mathbf{q}, \mathtt{f})x(\mathtt{f}), \qquad x \in X := \operatorname{ran} V.$$

In effect, we have replaced here the equation $V? = x$ by the *equivalent* equation

$$V(\mathtt{f}, :)? = x(\mathtt{f})$$

whose coefficient matrix is invertible. (In particular, $\#\mathtt{f} = \#V$; see H.P.(5.3).) □

5.2 For each of the following bases V of the linear subspace $\operatorname{ran} V$ of $\mathbb{F}^n$, give a matrix U for which Ux gives the coordinates of $x \in \operatorname{ran} V$ with respect to the basis V. How would you check your answer?

(a) $V = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$; (b) $V = [e_2, e_1, e_3] \in \mathbb{R}^{3\times 3}$; (c) $V = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 0 & 6 \end{bmatrix}$; (d) $V = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ -1 & 1 \\ 2 & -2 \end{bmatrix}$.

5.3 Prove the claim at the end of (5.4)Example.

This reduction, of the abstract linear equation $V? = x$ to a uniquely solvable square linear system, also works in the general setting.

To obtain a concrete expression, we **discretize** the abstract equation $V? = x$ by considering instead the *numerical* equation

$$\Lambda^{\mathrm{t}} V? = \Lambda^{\mathrm{t}} x$$

for some suitable data map $\Lambda^{\mathrm{t}} \in L(Y, \mathbb{F}^n)$ defined on some vector space $Y \supset X$. Here, 'suitable' means that the *matrix* $\Lambda^{\mathrm{t}} V$ is invertible, for then the unique solution of this equation must be the sought-for coordinate vector for $x \in X$ with respect to the basis V, i.e.,

$$a = V^{-1}x = (\Lambda^{\mathrm{t}} V)^{-1} \Lambda^{\mathrm{t}} x.$$

In (5.4)Example, we simply chose the linear map $y \mapsto y(\mathtt{f})$ as our Λ^{t}, i.e., $\Lambda^{\mathrm{t}} = \mathrm{id}_n(\mathtt{f}, :) = [e_j : j \in \mathtt{f}]^{\mathrm{t}}$, with $\mathtt{f}$ chosen, in effect, to ensure that $\Lambda^{\mathrm{t}} V = V(\mathtt{f}, :)$ is invertible. We indeed obtained there V^{-1} as

$$x \mapsto U(:, \mathtt{f})x(\mathtt{f}) = V(\mathtt{f}, :)^{-1} x(\mathtt{f}) = (\Lambda^{\mathrm{t}} V)^{-1} \Lambda^{\mathrm{t}} x.$$

How would one find a 'suitable' data map in general? That depends on the particular circumstances. For example, if $V \in L(\mathbb{F}^n, Y)$ and $\Lambda^t \in L(Y, \mathbb{F}^n)$, and we somehow know that Λ^t maps $X := \operatorname{ran} V = V(\mathbb{F}^n)$ *onto* $\mathbb{F}^n$, then we know that $\Lambda^t V$ maps $\mathbb{F}^n$ onto $\mathbb{F}^n$, hence, being a square matrix, $\Lambda^t V$ must be invertible. Conversely, if $\Lambda^t V$ is invertible, then V must be 1-1, hence provides a basis for its range, and Λ^t must map $\operatorname{ran} V$ onto $\mathbb{F}^n$.

(5.5) Proposition: If the linear map $V : \mathbb{F}^n \to X \subset Y$ is onto, and $\Lambda^t \in L(Y, \mathbb{F}^n)$ is such that their **Gramian matrix**, $\Lambda^t V$, is invertible, then V is a basis for X, and its inverse is

$$V^{-1} : X \to \mathbb{F}^n : x \mapsto (\Lambda^t V)^{-1} \Lambda^t x.$$

Change of basis

To be sure, under the assumptions of (5.5)Proposition, we also know that Λ^t maps X onto $\mathbb{F}^n$, hence, since both X and $\mathbb{F}^n$ are of the same finite dimension, the restriction of Λ^t to X must be invertible as a linear map to $\mathbb{F}^n$. Consequently, there must be an invertible $W \in L(\mathbb{F}^n, X)$, i.e., a basis W for X, with $\Lambda^t W = \mathrm{id}_n$.

Hence, the right side in our numerical equation $\Lambda^t V? = \Lambda^t x$ is the coordinate vector for $x \in X$ with respect to this basis W for X. In other words, our great scheme for computing the coordinates of $x \in X$ with respect to the basis V for X requires us to know the coordinates of x with respect to some basis for X. In other words, the entire calculation is just a *change of basis*, with $\Lambda^t V = W^{-1} V$ the socalled **transition matrix** that carries the V-coordinates of x to the W-coordinates of x.

However, this in no way diminishes its importance. For, we really have no choice in this matter. We cannot compute without having numbers to start with. Also, we often have ready access to the coordinate vector $\Lambda^t x$ without having in hand the corresponding basis W. At the same time, we may much prefer to know the coordinates of x with respect to a different basis.

For example, we know from (3.21)Proposition that, with $(\tau_1, \ldots, \tau_k)$ any sequence of pairwise distinct real numbers, the linear map $\Lambda^t : p \mapsto (p(\tau_1), \ldots, p(\tau_k))$ is 1-1 on the k-dimensional space $\Pi_{<k}$, hence provides the coordinates of $p \in \Pi_{<k}$ with respect to a certain basis W of $\Pi_{<k}$, namely the socalled **Lagrange basis** whose columns can be verified to be the so-called **Lagrange fundamental polynomials**

$$\ell_j : t \mapsto \prod_{h \neq j} \frac{t - \tau_h}{\tau_j - \tau_h}, \qquad j = 1{:}k. \tag{5.6}$$

However, you can imagine that it is a challenge to differentiate or integrate a polynomial written in terms of this basis. Much better for that to have the coordinates of the polynomial with respect to the power basis $V = [()^0, \ldots, ()^{k-1}]$.

5.4 What are the coordinates of $p \in \Pi_k$ with respect to the Lagrange basis for $\Pi_{<k}$ for the points $\tau_1, \ldots, \tau_k$?

5.5 Find the value at 0 of the quadratic polynomial p, for which $p(-1) = p(1) = 3$ and $Dp(1) = 6$.

5.6 Find a formula for $p(0)$ in terms of $p(-1)$, $p(1)$ and $Dp(1)$, assuming that p is a quadratic polynomial.

5.7 Find the coordinates for the polynomial $q(t) = 3 - 4t + 2t^2$ with respect to the basis $W := [()^0, ()^0 + ()^1, ()^0 + ()^1 + ()^2]$ of the space of quadratic polynomials. (Hint: you are given the coordinates for q wrto $V := [()^0, ()^1, ()^2] = W(W^{-1}V)$ and can easily determine $W^{-1}V)^{-1} = V^{-1}W$.)

5.8 Under the assumptions of (5.5)Proposition, the restriction of the data map Λ^t to that subspace X must be invertible. Prove that its inverse is $W := V(\Lambda^t V)^{-1}$.

5.9 Find the coordinates for the polynomial $q(t) = 3 - 4t + 2t^2$ with respect to the basis $W := [()^0, ()^0 + ()^1, ()^0 + ()^1 + ()^2]$ of the space of quadratic polynomials. (Hint: you are given the coordinates for q wrto $V := [()^0, ()^1, ()^2] = W(W^{-1}V)$ and can easily determine $W^{-1}V)^{-1} = V^{-1}W$.)

5.10 Let $v_1, \ldots, v_n$ be a sequence of $(n-1)$-times continuously differentiable functions, all defined on the interval $[a\,..\,b]$. For $x \in [a\,..\,b]$, the matrix

$$W(v_1, \ldots, v_n; x) := (D^{i-1} v_j(x) : i, j = 1{:}n)$$

is called the **Wronski matrix at** x for the sequence $(v_j : j = 1{:}n)$.

Prove that $V := [v_1, \ldots, v_n]$ is 1-1 in case, for some $x \in [a\,..\,b]$, $W(v_1, \ldots, v_n; x)$ is invertible. (Hint: Consider the Gram matrix $\Lambda^{\rm t} V$ with $\Lambda^{\rm t} f := (f(x), f'(x), \ldots, D^{n-1} f(x))$.)

Interpolation and linear projectors

As (3.20)Example already intimates, our formula in (5.5) for the inverse of a basis $V \in L(\mathbb{F}^n, X)$ can be much more than that. It is useful for *interpolation* in the following way. Assuming that $\Lambda^{\rm t} V$ is invertible, it follows that, for any $y \in Y$, $x = V(\Lambda^{\rm t} V)^{-1} \Lambda^{\rm t} y$ is the unique element in X that **agrees with** y **at** $\Lambda^{\rm t}$ in the sense that

$$\Lambda^{\rm t} x = \Lambda^{\rm t} y.$$

To recall the specifics of (3.20)Example, if $X = \Pi_{<k}$ and $\Lambda^{\rm t} : f \mapsto (f(\tau_i) : i = 1{:}k)$, with $\tau_1 < \cdots < \tau_k$, then, by (3.21)Proposition, for arbitrary $f : \mathbb{R} \to \mathbb{R}$, there is exactly one polynomial p of degree $< k$ for which $p(\tau_i) = f(\tau_i)$, $i = 1{:}k$.

One can readily imagine other examples.

Example: In **Hermite interpolation**, one specifies not only values but also derivatives. For example, in **two-point** Hermite interpolation from $\Pi_{<k}$, one picks two points, $t \neq u$, and two nonnegative integers r and s with $r + 1 \; + s + 1 = k$, and defines

$$\Lambda^{\rm t} : f \mapsto (f(t), Df(t), \ldots, D^r f(t), f(u), Df(u), \ldots, D^s f(u)).$$

Now the requirement that $\Lambda^{\rm t} p = \Lambda^{\rm t} f$ amounts to looking for $p \in \Pi_{<k}$ that agrees with f in the sense that p and f have the same derivative values of order $0, 1, \ldots, r$ at t and the same derivative values of order $0, 1, \ldots, s$ at u. □

Example: Recall from Calculus the bivariate **Taylor series**

$$f(s,t) = f(0) + D_s f(0)\, s + D_t f(0)\, t + \bigl({D_s}^2 f(0) s^2 + D_s D_t f(0) st + D_t D_s f(0) ts + {D_t}^2 f(0) t^2\bigr)/2 + h.o.t.$$

In particular, for any smooth function f, the quadratic polynomial

$$p : (s,t) \mapsto f(0) + D_s f(0)\, s + D_t f(0)\, t + \bigl({D_s}^2 f(0) s^2 + 2 D_s D_t f(0) st + {D_t}^2 f(0) t^2\bigr)/2$$

is the unique quadratic polynomial that matches the information about f given by the data map

$$\Lambda^{\rm t} : f \mapsto (f(0), D_s f(0), D_t f(0), D_s^2 f(0), D_s D_t f(0), D_t^2 f(0)).$$

□

Example: When dealing with **Fourier series**, one uses the data map

$$\Lambda^{\rm t} : f \mapsto (\int_0^{2\pi} f(t) \operatorname{cis}(jt)\, {\rm d}t : j = 0{:}N),$$

with cis standing for ‘sine and cosine’. One looks for a **trigonometric polynomial**

$$p = [\operatorname{cis}(j\cdot) : j = 0{:}N] a$$

that satisfies $\Lambda^{\rm t} p = \Lambda^{\rm t} f$, and finds it in the **truncated Fourier series** for f. □

Directly from (5.5)Proposition, we obtain (under the assumptions of that proposition) the following pretty formula

$$x = Py := V(\Lambda^{t}V)^{-1}\Lambda^{t}y \tag{5.7}$$

for the interpolant $x \in X$ to given $y \in Y$ with respect to the data map Λ^{t}. The linear map $P := V(\Lambda^{t}V)^{-1}\Lambda^{t}$ so defined on Y is very special:

(5.8) Proposition: Let the linear map $V : \mathbb{F}^n \to Y$ be onto $X \subset Y$, and let $\Lambda^{t} \in L(Y, \mathbb{F}^n)$ be such that their Gramian matrix, $\Lambda^{t}V$, is invertible. Then $P := V(\Lambda^{t}V)^{-1}\Lambda^{t}$ is a linear map on Y with the following properties:

(i) P is the identity on $X = \operatorname{ran} V$.

(ii) $\operatorname{ran} P = \operatorname{ran} V = X$.

(iii) P is a **projector** or **idempotent**, i.e., $PP = P$, hence $P(\operatorname{id} - P) = 0$.

(iv) $\operatorname{null} P = \operatorname{null} \Lambda^{t} = \operatorname{ran}(\operatorname{id} - P)$.

(v) Y is the direct sum of $\operatorname{ran} P$ and $\operatorname{null} P$, i.e., $Y = \operatorname{ran} P \dotplus \operatorname{null} P$.

Proof: (i) $PV = V(\Lambda^{t}V)^{-1}\Lambda^{t}\, V = V \operatorname{id} = V$, hence $P(Va) = Va$ for all $a \in \mathbb{F}^n$.

(ii) Since $P = VA$ for some A, we have that $\operatorname{ran} P \subset \operatorname{ran} V$, while $PV = V$ implies that $\operatorname{ran} P \supset \operatorname{ran} V$.

(iii) By (i) and (ii), P is the identity on its range, hence, in particular, $PP = P$, or, equivalently, $P(\operatorname{id} - P) = 0$.

(iv) The fact that $P = A\Lambda^{t}$ for some A implies that $\operatorname{null} P \supset \operatorname{null} \Lambda^{t}$, while also

$$\Lambda^{t}P = \Lambda^{t}\, V(\Lambda^{t}V)^{-1}\Lambda^{t} = \operatorname{id}_n \Lambda^{t} = \Lambda^{t},$$

hence also $\operatorname{null} P \subset \operatorname{null} \Lambda$. As to $\operatorname{null} P = \operatorname{ran}(\operatorname{id} - P)$, note that $x \in \operatorname{null} P$ implies that $x = x - Px = (\operatorname{id} - P)x \in \operatorname{ran}(\operatorname{id} - P)$, while, conversely, $\operatorname{null} P \supset \operatorname{ran}(\operatorname{id} - P)$ since, by (iii), $P(\operatorname{id} - P) = 0$.

(v) For any $y \in Y$, $y = Py + (\operatorname{id} - P)y \in \operatorname{ran} P + \operatorname{null} P$, by (iv), hence $Y = \operatorname{ran} P + \operatorname{null} P$. If also $y = x + z$ for some $x \in \operatorname{ran} P$ and some $z \in \operatorname{null} P$, then, by (i) and (iv), $Py = P(x + z) = Px + Pz = x$, therefore also $z = y - x = y - Py = (\operatorname{id} - P)y$, showing such a decomposition to be unique. □

5.11 Consider the linear map Q given on $X = \{f : \mathbb{R} \to \mathbb{R}\}$ by $Qf(t) = (f(t) + f(-t))/2$. Prove that Q is a linear projector. Also, give a succinct description of its range and its nullspace. (Hint: consider the map $F : X \to X$ defined by $(Ff)(t) = -f(t)$)

(5.9) Example: We specialize the general situation of (5.8)Proposition to the case $V : \mathbb{R}^1 \to X \subset \mathbb{R}^2$, so we can draw a figure like (5.10)Figure.

Take $Y = \mathbb{R}^2$, and let $v \in \mathbb{R}^2 \neq 0$, hence $X := \operatorname{ran} V$ with $V := [v]$ is 1-dimensional. The general linear map $\Lambda^{t} : \mathbb{R}^2 \to \mathbb{R}^1$ is of the form $[w]^{t}$ for some $w \in \mathbb{R}^2$, and the requirement that $\Lambda^{t}V$ be invertible reduces to the requirement that $[w]^{t}[v] = w^{t}v \neq 0$.

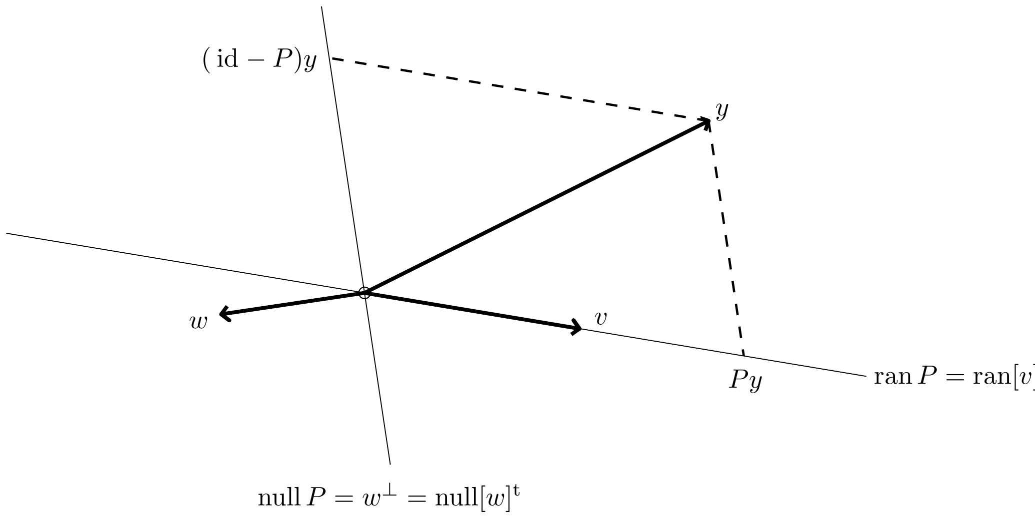
With $V = [v]$ and $\Lambda^{t} = [w]^{t}$ so chosen, the linear projector P becomes

$$P := \frac{vw^{t}}{w^{t}v} : y \mapsto v\frac{w^{t}y}{w^{t}v}.$$

We readily verify directly that

$$PP = \frac{vw^{t}}{w^{t}v}\frac{vw^{t}}{w^{t}v} = \frac{v\, w^{t}v\, w^{t}}{(w^{t}v)\,(w^{t}v)} = \frac{vw^{t}}{w^{t}v} = P,$$

i.e., that P is a linear projector. Its range equals ran$[v]$, i.e., the straight line through the origin in the direction of v. Its nullspace equals null$[w]^t$ and this is necessarily also 1-dimensional, by (4.15)Dimension Formula, hence is the straight line through the origin perpendicular to w. The two lines have only the origin in common since $y \in \operatorname{ran} P \cap \operatorname{null} P$ implies that $y = v\alpha$ for some scalar α, therefore $0 = w^t y = w^t v\alpha$ and this implies that $\alpha = 0$ since $w^t v \neq 0$ by assumption.



(5.10) Figure. The direct sum decomposition provided by a certain linear projector. Compare this to (4.28)Figure.

We can locate the two summands in the split

$$y = Py + (\operatorname{id} - P)y$$

graphically (see (5.10)Figure): To find Py, draw the line through y parallel to null P; its unique intersection with ran P = ran$[v]$ is Py. The process of locating $(\operatorname{id} - P)y$ is the same, with the roles of ran P and null P reversed: Now draw the line through y parallel to ran P; its unique intersection with null P is the element $(\operatorname{id} - P)y$.

This shows graphically that, for each y, Py is the unique element of ran P for which $w^t Py = w^t y$, i.e., the unique point in the intersection of ran P and $y + \operatorname{null}[w]^t$. □

It is useful to note that, for any linear projector P, also $(\operatorname{id} - P)$ is a linear projector (since $(\operatorname{id} - P)(\operatorname{id} - P) = \operatorname{id} - P - P + PP = \operatorname{id} - P$), and that any direct sum decomposition $Y = X \dot{+} Z$ of a finite-dimensional Y necessarily has $X = \operatorname{ran} P$ and $Z = \operatorname{null} P$ for some linear projector P. The following is a more general such claim, of use later.

(5.11) Proposition: Let $X_1, \ldots, X_r$ be linear subspaces of the finite-dimensional vector space Y. Then the following are equivalent.

(i) Y is the direct sum of the X_j, i.e., $Y = X_1 \dot{+} \cdots \dot{+} X_r$.

(ii) There exist $P_j \in L(Y)$ with $\operatorname{ran} P_j = X_j$ so that

$$\mathrm{id}_Y = P_1 + \cdots + P_r \tag{5.12}$$

and

$$P_j P_k = \begin{cases} P_j = P_k & \text{if } j = k; \\ 0 & \text{otherwise.} \end{cases} \tag{5.13}$$

In particular, each P_j is a linear projector.

Also, the conditions in (ii) uniquely determine the P_j.

Proof: Let V_j be a basis for X_j, all j. By (4.26)Proposition, (i) is equivalent to having $V := [V_1, \ldots, V_r]$ be a basis for Y.

'(i) $\Longrightarrow$ (ii)': By assumption, V is a basis for Y. Let $V^{-1} =: \Lambda^{\mathrm{t}} =: [\Lambda_1, \ldots, \Lambda_r]^{\mathrm{t}}$ be its inverse, grouped correspondingly. Then

$$\mathrm{id}_{\dim Y} = \Lambda^{\mathrm{t}} V = [\Lambda_1, \ldots, \Lambda_r]^{\mathrm{t}} [V_1, \ldots, V_r] = ({\Lambda_i}^{\mathrm{t}} V_j : i, j = 1{:}r),$$

i.e.,

$${\Lambda_i}^{\mathrm{t}} V_j = \begin{cases} \mathrm{id} & \text{if } i = j; \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the linear maps

$$P_j := V_j {\Lambda_j}^{\mathrm{t}}, \quad j = 1{:}r,$$

satisfy (5.13), and $\operatorname{ran} P_j = X_j$, for all j. But also

$$\mathrm{id}_Y = V \Lambda^{\mathrm{t}} = [V_1, \ldots, V_r][\Lambda_1, \ldots, \Lambda_r]^{\mathrm{t}} = \sum_j V_j {\Lambda_j}^{\mathrm{t}},$$

showing (5.12).

'(ii) $\Longrightarrow$ (i)': By assumption, $\operatorname{ran} P_j = \operatorname{ran} V_j$, all j. Therefore, by assumption (5.13),

$$P_j V_i = \begin{cases} V_j & \text{if } j = i; \\ 0 & \text{otherwise.} \end{cases} \tag{5.14}$$

Therefore, $0 = Va = \sum_i V_i a_i$ implies, for any particular j, that $0 = P_j 0 = P_j V a = \sum_i P_j V_i a_i = P_j V_j a_j = V_j a_j$, hence $a_j = 0$ (since V_j is 1-1). It follows that V is 1-1. On the other hand, the assumption (5.12) implies that V is onto. Hence, V is a basis for Y.

Finally, to prove the uniqueness of the P_j satisfying (ii), notice that (5.14) pins down P_j on all the columns of V. Since (ii) implies that V is a basis for Y, this therefore determines P_j uniquely (by (4.2)Proposition). □

Returning to the issue of interpolation, this gives the following

(5.15)Corollary: If $V \in L(\mathbb{F}^n, Y)$ is 1-1, and $\Lambda^{\mathrm{t}} \in L(Y, \mathbb{F}^n)$ is such that $\operatorname{ran} V \cap \operatorname{null} \Lambda^{\mathrm{t}} = \{0\}$, then $P := V(\Lambda^{\mathrm{t}} V)^{-1}\Lambda^{\mathrm{t}}$ is well-defined; it is the *unique* linear projector P with

(5.16) $$\operatorname{ran} P = \operatorname{ran} V, \qquad \operatorname{null} P = \operatorname{null} \Lambda^{\mathrm{t}}.$$

In particular, then Λ^{t} is onto, and

(5.17) $$Y = \operatorname{ran} V \dotplus \operatorname{null} \Lambda^{\mathrm{t}}.$$

For an arbitrary abstract vector space, it may be very hard to come up with suitable concrete data maps. For that reason, we now consider a particular kind of vector space for which it is very easy to provide suitable data maps, namely the inner product spaces.

6. Inner product spaces

Definition and examples

Inner product spaces are vector spaces with an additional operation, the *inner product*. Here is the definition.

(6.1) Definition: An **inner product space** is a vector space Y (over the field $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) and an **inner product**, meaning a map

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

that is

(a) **positive definite**, i.e., $\|x\|^2 := \langle x, x\rangle \ge 0$, with equality iff $x = 0$;

(b) **linear in its first argument**, i.e., $\langle \cdot, y\rangle \in L(Y, \mathbb{F})$;

(c) **hermitian**, or **skew-symmetric**, i.e., $\langle y, x\rangle = \overline{\langle x, y\rangle}$.

You already know an inner product space, namely n-dimensional Euclidean space, i.e., the space of n-vectors with the inner product

$$\langle x, y\rangle := \overline{y}^{\rm t} x = \sum_j x_j \overline{y_j} =: y^{\rm c} x,$$

though you may know it under the name **scalar product** or **dot product**. In particular, (b) and (c) are obvious in this case. As to (a), observe that, for any complex number $z = u + {\rm i}v$,

$$\overline{z}z = (u - {\rm i}v)(u + {\rm i}v) = u^2 + v^2 = |z|^2 \ge 0,$$

with equality if and only if $u = 0 = v$, i.e., $z = 0$. Hence, for any $x \in \mathbb{F}^n$,

$$\langle x, x\rangle = \overline{x}^{\rm t} x = |x_1|^2 + \cdots + |x_n|^2 \ge 0,$$

with equality iff all the x_j are zero, i.e., $x = 0$.

Of course, if the scalar field is $\mathbb{R}$, we can forget about taking complex conjugates since then $\overline{x} = x$. But if $\mathbb{F} = \mathbb{C}$, then it is essential that we define $\langle x, y\rangle$ as $y^{\rm c} x = \overline{y}^{\rm t} x$ rather than as $y^{\rm t} x$ since we would not get positive definiteness otherwise. Indeed, if z is a complex number, then there is no reason to think that z^2 is nonnegative, and the following calculation

$$(1, {\rm i})^{\rm t}(1, {\rm i}) = 1^2 + ({\rm i})^2 = 1 - 1 = 0$$

shows that, for a complex x, $x^{\rm t} x$ can be zero without x being zero.

So, why not simply stick with $\mathbb{F} = \mathbb{R}$? Work on eigenvalues requires consideration of *complex* scalars (since it relies on zeros of polynomials, and a polynomial may have complex zeros even if all its coefficients are real). For this reason, we have taken the trouble all along to take into account the possibility that $\mathbb{F}$ might be $\mathbb{C}$. It is a minor nuisance at this point, but will save time later.

Another example of an inner product space of great practical interest is the space $Y = \overset{\circ}{C}$ of all continuous 2π-periodic functions, with the inner product

$$\langle f, g\rangle := \int_0^{2\pi} f(t)\overline{g(t)}\,{\rm d}t.$$

