Of course, we can also think of the space $C[a\,..\,b]$ as an inner product space, with respect to the inner product

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

Often, it is even useful to consider on $C[a\,..\,b]$ the more general inner product

$$\langle f, g\rangle := \int_a^b f(t)\overline{g(t)}w(t)\,\mathrm{d}t$$

with w some positive function on $[a\,..\,b]$, and there are analogous inner product spaces consisting of functions of several variables.

In order to stress the fact that a general inner product space Y behaves just like $\mathbb{F}^n$ with the standard inner product, I will use the notation

$$y^{\mathrm{c}} : Y \to \mathbb{F} : x \mapsto \langle x, y\rangle, \qquad \forall y \in Y,$$

for the linear functional provided, according to (6.1)(b), by the inner product, hence will feel free to write $y^{\mathrm{c}}x$ rather than $\langle x, y\rangle$ for the inner product of x with y. Correspondingly, you can read the rest of this chapter as if we were just talking about the familiar space of n-vectors with the dot product, yet be certain that, when the time comes, you will have in hand very useful facts about an arbitrary inner product space, for example the space $\overset{\circ}{C}$.

The conjugate transpose

Here is the promised ready supply of data maps available in an inner product space.

Any column map $W = [w_1, \ldots, w_n] \in L(\mathbb{F}^n, Y)$ into an inner product space Y provides the corresponding data map

$$W^{\mathrm{c}} : Y \mapsto \mathbb{F}^n : x \mapsto ({w_j}^{\mathrm{c}}x : j = 1{:}n),$$

called its **conjugate transpose** or **Hermitian**.

The terminology comes from the special case $Y = \mathbb{F}^m$. In that case, $W \in \mathbb{F}^{m\times n}$, and then W^{c} is, indeed, just the conjugate transpose of the matrix W since then $w_j = W(:, j)$, hence

$${w_j}^{\mathrm{c}}x = W(:, j)^{\mathrm{c}}x = \sum_k \overline{W(k, j)}x_k = \sum_k (W^{\mathrm{c}})(j, k)x_k = (W^{\mathrm{c}}x)_j.$$

Further, if $W \in L(\mathbb{F}^n, Y)$ and $A \in \mathbb{F}^{n\times m}$, then, with $WA = [u_j := \sum_k w_k A(k, j) : j = 1{:}n]$, one verifies that

$$((WA)^{\mathrm{c}}x)_j = {u_j}^{\mathrm{c}}x = \sum_k \overline{A(k, j)}{w_k}^{\mathrm{c}}x = \sum_k A^{\mathrm{c}}(j, k){w_k}^{\mathrm{c}}x = (A^{\mathrm{c}}(W^{\mathrm{c}}x))_j.$$

This proves

(6.2): If $W \in L(\mathbb{F}^n, Y)$ and $A \in \mathbb{F}^{n\times m}$, then $WA \in L(\mathbb{F}^m, Y)$ and $(WA)^{\mathrm{c}} = A^{\mathrm{c}}W^{\mathrm{c}}$.

This observation shows that *the above definition of the conjugate transpose of a column map is a special case of the abstract definition of the conjugate transpose of $A \in L(X, Y)$ as the unique map $A^{\mathrm{c}} : Y \to X$ (necessarily linear) for which*

(6.3) $$\langle x, A^{\mathrm{c}}y\rangle = \langle Ax, y\rangle, \qquad \forall (x, y) \in X \times Y.$$

Indeed, if also $\langle x, z\rangle = \langle Ax, y\rangle$ for all $x \in X$, then $\langle x, z - A^c y\rangle = 0$ for all $x \in X$, including $x = z - A^c y$, hence, by the definiteness of the inner product, $z - A^c y = 0$, showing that $A^c y$ is uniquely determined by (6.3). With that, it is easy to see that A^c is a linear map, and that the conjugate transpose of an n-column map into Y is, indeed, the conjugate transpose in the sense of (6.3) (with $X = \mathbb{F}^n$), and that

$$(BA)^c = A^c B^c \tag{6.4}$$

in case BA makes sense, hence, in particular,

$$A^{-c} := (A^{-1})^c = (A^c)^{-1}. \tag{6.5}$$

The only fly in the ointment is the fact that, for some $A \in L(X, Y)$, there may not be any map $A^c : Y \to X$ satisfying (6.3) unless X is 'complete', a condition that is beyond the scope of these notes. However, if both X and Y are finite-dimensional inner-product spaces, then, with V and W bases for X and Y, respectively, we can write any $A \in L(X, Y)$ as $A = W\widehat{A}V^{-1}$ (using the *matrix* $\widehat{A} := W^{-1}AV$), hence, with (6.4), have available the formula

$$A^c = (W\widehat{A}V^{-1})^c = V^{-c}\widehat{A}^c W^c$$

for the conjugate transpose of A, – another nice illustration of the power of the basis concept.

With that, we are ready for the essential fact about the conjugate transpose needed now.

(6.6) Lemma: If the range of the 1-1 column map V is contained in the range of some column map W, then $W^c V$ is 1-1, i.e., W^c is 1-1 on $\operatorname{ran} V$.

Proof: Assume that $W^c V a = 0$ and let $b := Va$. Then $b \in \operatorname{ran} V \subset \operatorname{ran} W$, hence we must have $b = Wc$ for some vector c. Therefore, using (6.2),

$$0 = c^c 0 = c^c\, W^c V a = (Wc)^c V a = b^c b.$$

By the definiteness of the inner product, this implies that $b = 0$, i.e., $Va = 0$, therefore that $a = 0$, since V is assumed to be 1-1. □

By taking now, in particular, $W = V$, it follows that, for any basis V of the linear subspace X of the inner product space Y, the linear map $(V^c V)^{-1} V^c$ is well-defined, hence provides a formula for V^{-1}.

In `MATLAB`, the conjugate transpose of a matrix `A` is obtained as `A'`, hence the corresponding formula is `inv(V'*V)*V'`. It is, in effect, used there to carry out the operation

```
V\
```

for a matrix `V` that is merely 1-1.

□

Orthogonal projectors and least-squares approximation

We conclude that, with V a basis for the linear subspace X of the inner product space Y, the linear projector

$$P_V := V(V^{\rm c}V)^{-1}V^{\rm c}$$

is well-defined. Moreover, by (5.8), $\operatorname{null} P_V = \operatorname{null} V^{\rm c} = \{y \in Y : V^{\rm c}y = 0\}$. Since $x \in \operatorname{ran} P_V = \operatorname{ran} V$ is necessarily of the form $x = Va$, it follows that, for any $x \in \operatorname{ran} P_V$ and any $y \in \operatorname{null} P_V$,

$$x^{\rm c}y = (Va)^{\rm c}y = a^{\rm c}(V^{\rm c}y) = 0.$$

In other words, $\operatorname{ran} P_V$ and $\operatorname{null} P_V = \operatorname{ran}(\operatorname{id} - P_V)$ are perpendicular or orthogonal to each other, in the sense of the following definition.

Definition: We say that the elements u, v of the inner product space Y are **orthogonal** or **perpendicular** to each other, and write this

$$u \perp v,$$

in case $\langle u, v\rangle = 0$.

More generally, for any $F, G \subset Y$, we write $F \perp G$ to mean that, $\forall f \in F, g \in G,\ \ f \perp g$.

Note that $u \perp v$ iff $v \perp u$ since $\langle v, u\rangle = \overline{\langle u, v\rangle}$.

Because of the orthogonality

$$\operatorname{null} P_V = \operatorname{ran}(\operatorname{id} - P_V)\ \perp\ \operatorname{ran} P_V$$

just proved, P_V is called the **orthogonal** projector onto $\operatorname{ran} V$. Correspondingly, we write

(6.7) $$Y = \operatorname{ran} P_V \oplus \operatorname{null} P_V$$

to stress the fact that, in this case, the summands in this direct sum are orthogonal to each other. This orthogonality, as we show in a moment, has the wonderful consequence that, for any $y \in Y$, $P_V y$ is the unique element of $\operatorname{ran} P_V = \operatorname{ran} V$ that is closest to y in the sense of the **(Euclidean) norm**

(6.8) $$\|\cdot\| : Y \to \mathbb{R} : y \mapsto \sqrt{y^{\rm c}y}.$$

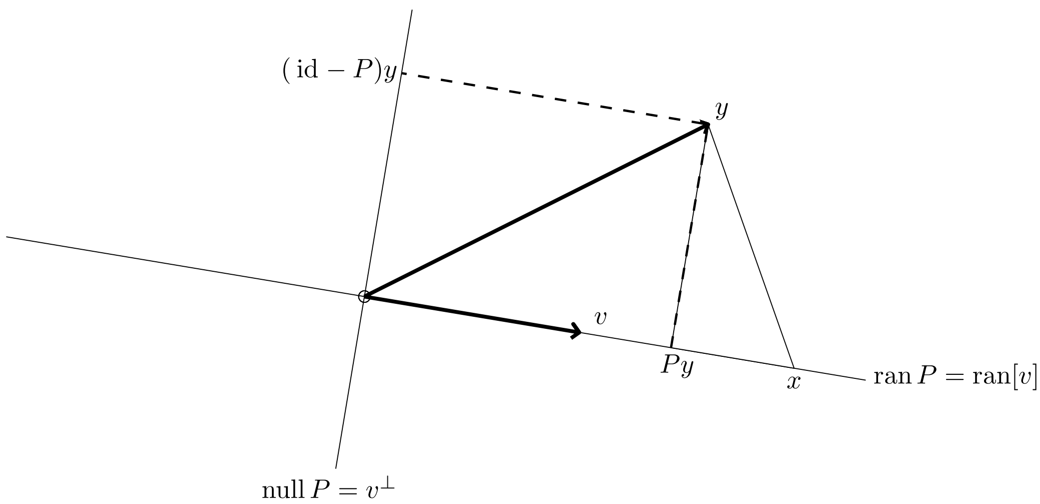
Thus, whether or not $y \in Y$ is in $\operatorname{ran} V$, the coordinate vector $a = (V^{\rm c}V)^{-1}V^{\rm c}y$ supplied by our formula gives the coordinates of the point in $\operatorname{ran} V$ closest to y. If $y \in \operatorname{ran} V$, then this is, of course, y itself.

(6.9) Example: We continue with (5.9)Example. In that example, the choice $\Lambda^{\rm t} = V^{\rm c}$ amounts to choosing $w = v$. Now P becomes $P = vv^{\rm c}/v^{\rm c}v$, and, correspondingly,

$$Py = v\frac{v^{\rm c}y}{v^{\rm c}v},$$

which we recognize as the standard formula for the orthogonal projection of the vector y onto the line spanned by the vector v.

Correspondingly, (5.10)Figure changes to the following.



(6.10) Figure. If $y - Py$ is perpendicular to $\operatorname{ran} P$, then Py is the closest point to y from $\operatorname{ran} P$ since then, for any $x \in \operatorname{ran} P$, $\|y-x\|^2 = \|y-Py\|^2 + \|x-Py\|^2$.

□

The *proof* that, for any $y \in Y$, $P_V y$ is the unique element of $\operatorname{ran} V$ closest to y in the sense of the norm (6.8) is based on nothing more than the following little calculation.

$$\|u+v\|^2 = (u+v)^{\mathrm{c}}(u+v) = \|u\|^2 + v^{\mathrm{c}}u + u^{\mathrm{c}}v + \|v\|^2.$$

Since $v^{\mathrm{c}}u = \overline{u^{\mathrm{c}}v}$, this proves

(6.11) Pythagoras: $u \perp v \implies \|u+v\|^2 = \|u\|^2 + \|v\|^2.$

Since, for any $x \in X$, $y-x = (y-P_V y) + (P_V y - x)$, while $(y-P_V y) \in \operatorname{null} P_V \perp \operatorname{ran} P_V = X \ni (P_V y - x)$ we conclude that

(6.12) $$\|y-x\|^2 = \|y-P_V y\|^2 + \|P_V y - x\|^2.$$

Here, the first term on the right is *independent of* x. This shows that $\|y-x\|$ is uniquely minimized over $x \in X$ by the choice $x = P_V y$, as we claimed.

Here is the formal statement.

(6.13) Theorem: For any basis V for the linear subspace X of the inner product space Y, the linear map

$$P_V = V(V^{\mathrm{c}}V)^{-1}V^{\mathrm{c}}$$

is the orthogonal projector onto X in the sense that, for all $y \in Y$, $P_V y \in X$ and $y - P_V y \perp X$. Therefore, Y is the **orthogonal direct sum**

$$Y = \operatorname{ran} V \oplus \operatorname{null} V^{\mathrm{c}} = \operatorname{ran} P_V \oplus \operatorname{null} P_V = X \oplus \operatorname{ran}(\operatorname{id} - P_V),$$

and

$$\forall\{x \in X,\ y \in Y\} \quad \|y-x\| \ge \|y - P_V y\|,$$

with equality if and only if $x = P_V y$.

Note that $P_V y = Va$, with the coefficient vector a the unique solution to the linear equation

$$V^c V a = V^c y.$$

This equation is also referred to as the **normal equation** since it requires that $V^c(y - Va) = 0$, i.e., that the residual, $y - Va$, be perpendicular or *normal* to every column of V, hence to all of $\operatorname{ran} V$ (see (6.10)Figure). In effect, given that the equation $V? = y$ doesn't have a solution for arbitrary $y \in Y$, our particular $Va = P_V y$ gives us the closest thing to a solution.

In particular, if $y \in Y = \mathbb{R}^n$ and $V \in \mathbb{R}^{n \times r}$ is 1-1, then $P_V y$ minimizes $\|y - Va\|$ over all $a \in \mathbb{R}^r$. For that reason, the coefficient vector $a := V^{-1} P_V y$ is called the **least-squares solution** to the (usually inconsistent or overdetermined) linear system $V? = y$.

A practically very important special case of this occurs when $t := (t_1, \ldots, t_n)$ is a sequence of pairwise distinct points in some set T, with

$$Q_t : f \mapsto (f(t_j) : j = 1{:}n)$$

the corresponding data map of evaluation at the t_j, and F is a finite-dimensional vector space of real functions on that set T. Then, with V a basis for F and y an arbitrary real n-vector, we could look for solutions to the linear system

$$Q_t V? = y.$$

If $Q_t V a = y$, then $f := Va$ is an element in F that interpolates the data values y_j, i.e., that satisfies

$$f(t_j) = y_j, \quad j = 1{:}n.$$

But, unless

$$V_t := Q_t V$$

is invertible, we will not be able to find solutions for every $y \in \mathbb{R}^n$.

The next best thing is to assume that $V_t = Q_t V$ is 1-1, hence a basis for

$$F_t := Q_t F.$$

In that case, we now know that $a = P_{V_t} y$ minimizes $\|y - V_t a\|$, hence is the least-squares solution to the equation $V_t? = y$. Correspondingly, the element $f := V P_{V_t} y$ of F is called a discrete least-squares approximation from F to the data $((t_j, y_j) : j = 1{:}n)$.

In `MATLAB`, the vector $P_V y$ is computed as `V*(V\y)`, in line with the fact mentioned earlier that the action of the matrix $(V^c V)^{-1} V^c$ is provided by the operator `V\`, i.e., (up to roundoff and for any vector `y`) the three vectors

```
a1 = V\y, a2 = inv(V'*V)*V'*y, a3 = (V'*V)\(V'*y)
```

are all the same. However, the first way is preferable since it avoids actually forming the matrix `V'*V` (or its inverse) and, therefore, is less prone to roundoff effects.

□

Incidentally, by choosing $x = 0$ in (6.12), – legitimate since $\operatorname{ran} V$ is a linear subspace, – we find the following very useful fact.

(6.14) Proposition: For any 1-1 column map V into Y and any $y \in Y$,

$$\|y\| \ge \|P_V y\|,$$

with equality if and only if $y = P_V y$, i.e., if and only if $y \in \operatorname{ran} V$.

This says that P_V *strictly reduces norms*, except for those elements that it doesn't change at all.

In particular, this says that, for an arbitrary nonzero $x \in Y$, the orthogonal projection $P_{[x]}y = x(x^c y)/\|x\|^2$ onto ran$[x]$ of an arbitrary $y \in Y$ has norm smaller than $\|y\|$ unless $y = P_{[x]}y$. In other words, $\|y\| \geq \|x(x^c y)/\|x\|^2\| = |x^c y|/\|x\|$, hence

$$|\langle y, x\rangle| = |x^c y| \leq \|y\|\|x\|, \tag{6.15}$$

with equality if and only if $[y, x]$ is not 1-1. This is the famous **Cauchy(-Bunyakovski-Schwarz) Inequality** (or **CBS Inequality**), and it holds trivially (as an equality) in case $x = 0$. Be sure to remember not only the *in*equality, but also exactly when it is an *equal*ity.

In the derivation of (6.15), we used the fact that the norm is **absolutely homogeneous**, meaning that

$$\|\alpha y\| = |\alpha|\|y\|, \quad \forall(\alpha, y) \in \mathbb{F} \times Y.$$

This makes it possible to **normalize** any nonzero $y \in Y$ by dividing it by its norm and, in this way, obtain the vector $y/\|y\|$ that points in the same direction as y but is of norm 1.

6.1 Construct the orthogonal projection of the vector $(1, 1, 1)$ onto the line $L = \mathrm{ran}[1; -1; 1]$.

6.2 Construct the orthogonal projection of the vector $x := (1, 1, 1)$ onto the straight line $y + \mathrm{ran}[v]$, with $y = (2, 0, 1)$ and $v = (1, -1, 1)$. (Hint: you want to minimize $\|x - (y + \alpha v)\|$ over all $\alpha \in \mathbb{R}$.)

6.3 Compute the distance between the two straight lines $y + \mathrm{ran}[v]$ and $z + \mathrm{ran}[w]$, with $y = (2, 0, 1)$, $v = (1, 1, 1)$, $z = (-1, 1, -1)$ and $w = (0, 1, 1)$. (Hint: you want to minimize $\|y + \alpha v - (z + \beta w)\|$ over α, β.)

6.4 With $v_1 = (1, 2, 2)$, $v_2 = (-2, 2, -1)$, (a) construct the matrix that provides the orthogonal projection onto the subspace $\mathrm{ran}[v_1, v_2]$ of $\mathbb{R}^3$; (b) compute the orthogonal projection of the vector $y = (1, 1, 1)$ onto $\mathrm{ran}[v_1, v_2]$.

6.5 Taking for granted that the space $Y := C[-1\,..\,1]$ of real-valued continuous functions on the interval $[-1\,..\,1]$ is an inner product space with respect to the inner product

$$\langle f, g\rangle := \int_{-1}^{1} f(t)g(t)\,\mathrm{d}t,$$

do the following: (a) Construct (a formula for) the orthogonal projector onto $X := \Pi_1$, using the power basis, $V = [()^0, ()^1]$ for X. (b) Use your formula to compute the orthogonal projection of $()^2$ onto Π_1.

6.6 (a) Prove: *If* $\mathbb{F} = \mathbb{R}$*, then* $u \perp v$ *if and only if* $\|u + v\|^2 = \|u\|^2 + \|v\|^2$. (b) What goes wrong with your argument when $\mathbb{F} = \mathbb{C}$?

6.7 Compute the discrete least squares approximation by straight lines (i.e., from Π_1) to the data (j, j^2), $j = 1{:}10$ using (a) the basis $[()^0, ()^1]$; (b) the basis $[()^0, ()^1 - 5.5()^0]$. (c) Why might one prefer (b) to (a)?

6.8 For each of the following maps $f : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$, determine whether or not it is an inner product.

(a) $\mathbb{F} = \mathbb{R}$, $n = 3$, and $f(x, y) = x_1y_1 + x_3y_3$; (b) $\mathbb{F} = \mathbb{R}$, $n = 3$, and $f(x, y) = x_1y_1 - x_2y_2 + x_3y_3$; (c) $\mathbb{F} = \mathbb{R}$, $n = 2$, and $f(x, y) = x_1^2 + y_1^2 + x_2y_2$; (d) $\mathbb{F} = \mathbb{C}$, $n = 3$, and $f(x, y) = x_1y_1 + x_2y_2 + x_3y_3$; (e) $\mathbb{F} = \mathbb{R}$, $n = 3$, and $f(x, y) = x_1y_2 + x_2y_3 + x_3y_1$;

6.9 Prove that, for any invertible $A \in \mathbb{F}^{n\times n}$, $\langle\cdot,\cdot\rangle : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F} : (x, y) \mapsto (Ay)^c Ax = y^c(A^cA)x$ is an inner product on $\mathbb{F}^n$.

$\operatorname{ran} A$ and $\operatorname{null} A^{c}$ form an orthogonal direct sum for $\operatorname{tar} A$

The two basic linear subspaces associated with $A \in L(X,Y)$ are its range, $\operatorname{ran} A$, and its kernel or nullspace, $\operatorname{null} A$. However, when X and Y are inner product spaces, it is also very useful to consider the range of A and the nullspace of the (conjugate) transpose A^{c} of A together. For, then, by the definiteness of the inner product, $A^{c}y = 0$ iff $\langle x, A^{c}y\rangle = 0$ for all $x \in X$, while, by (6.3), $\langle x, A^{c}y\rangle = \langle Ax, y\rangle$, hence

$$\operatorname{null} A^{c} = \{y \in Y : y \perp \operatorname{ran} A\}.$$

With the standard notation

$$M^{\perp} := \{y \in Y : y \perp M\}$$

for the **orthogonal complement** of the subset M of Y, we get the following.

(6.16) Proposition: For any $A \in L(X,Y)$, $(\operatorname{ran} A)^{\perp} = \operatorname{null} A^{c}$.

(6.17) Corollary: For any $A \in L(X,Y)$, Y is the *orthogonal* direct sum $Y = \operatorname{ran} A \oplus \operatorname{null} A^{c}$. Hence

$$\dim \operatorname{tar} A = \dim \operatorname{ran} A + \dim \operatorname{null} A^{c}.$$

Proof: Let V be any basis for $\operatorname{ran} A$. By (6.13)Theorem,

$$Y = \operatorname{ran} V \oplus \operatorname{null} V^{c},$$

while, by choice of V, $\operatorname{ran} V = \operatorname{ran} A$, and so, by (6.16), $\operatorname{null} V^{c} = (\operatorname{ran} V)^{\perp} = (\operatorname{ran} A)^{\perp} = \operatorname{null} A^{c}$. □

In particular, A is onto if and only if A^{c} is 1-1. Further, since $(A^{c})^{c} = A$, we also have the following complementary statement.

(6.18) Corollary: For any $A \in L(X,Y)$, X is the *orthogonal* direct sum $X = \operatorname{ran} A^{c} \oplus \operatorname{null} A$. Hence,

$$\dim \operatorname{dom} A = \dim \operatorname{ran} A^{c} + \dim \operatorname{null} A.$$

In particular, A^{c} is onto if and only if A is 1-1. Also, on comparing (6.18) with the Dimension Formula, we see that $\dim \operatorname{ran} A = \dim \operatorname{ran} A^{c}$.

The fact (see (6.17)Corollary) that $\operatorname{tar} A = \operatorname{ran} A \oplus \operatorname{null} A^{c}$ is often used as a characterization of the elements $y \in \operatorname{tar} A$ for which the equation $A? = y$ has a solution. For, it says that $y \in \operatorname{ran} A$ if and only if $y \perp \operatorname{null} A^{c}$. Of course, since $\operatorname{null} A^{c}$ consists exactly of those vectors that are orthogonal to all the columns of A, this is just a special case of the fact that the orthogonal complement of the orthogonal complement of a linear subspace is that linear subspace itself.

Orthonormal column maps

The formula

$$P_V = V(V^c V)^{-1} V^c$$

for the orthogonal projector onto the range of the 1-1 column map V becomes particularly simple in case

$$V^c V = \text{id}; \tag{6.19}$$

it then reduces to

$$P_V = V V^c.$$

We call V **orthonormal** (or, **o.n.**, for short) in this case since, written out entry by entry, (6.19) reads

$$\langle v_j, v_k \rangle = \left\{ \begin{matrix} 1 & \text{if } j = k; \\ 0 & \text{otherwise,} \end{matrix} \right\} =: \delta_{jk}.$$

In other words, each column of V is *normalized*, meaning that it has norm 1, and different columns are orthogonal to each other. Such bases are special in that they provide their own inverse, i.e.,

$$x = V(V^c x), \quad \forall x \in \operatorname{ran} V.$$

We now show that every finite-dimensional linear subspace of an inner-product space Y has o.n. bases.

(6.20) Proposition: For every 1-1 $V \in L(\mathbb{F}^n, Y)$, there exists an o.n. $Q \in L(\mathbb{F}^n, Y)$ so that, for all j, $\operatorname{ran}[q_1, q_2, \ldots, q_j] = \operatorname{ran}[v_1, v_2, \ldots, v_j]$, hence $V = QR$ with R (invertible and) upper triangular, a **QR factorization** for V.

Proof: For $j = 1{:}n$, define $u_j := v_j - P_{V_{<j}} v_j$, with $V_{<j} := V_{j-1} := [v_1, \ldots, v_{j-1}]$. By (6.13)Theorem, $u_j \perp \operatorname{ran} V_{<j}$, all j, hence $u_j \perp u_k$ for $j \neq k$. Also, each u_j is nonzero (since $u_j = V_j(a, 1)$ for some $a \in \mathbb{F}^{j-1}$, and V_j is 1-1), hence $q_j := u_j/\|u_j\|$ is well-defined and, still, $q_j \perp q_k$ for $j \neq k$.

It follows that $Q := [q_1, \ldots, q_n]$ is o.n., hence, in particular, 1-1. Finally, since $q_j = u_j/\|u_j\| \in \operatorname{ran} V_j$, it follows that, for each j, the 1-1 map $[q_j, \ldots, q_j]$ has its range in the j-dimensional space $\operatorname{ran} V_j$, hence must be a basis for it. □

Since $Q_{<j} = [q_1, \ldots, q_{j-1}]$ is an o.n. basis for $\operatorname{ran} V_{<j}$, it is of help in constructing q_j since it gives

$$u_j = v_j - P_{V_{<j}} v_j, \quad \text{with} \quad P_{V_{<j}} v_j = P_{Q_{<j}} v_j = \sum_{k<j} q_k \langle v_j, q_k \rangle = \sum_{k<j} u_k \frac{\langle v_j, u_k \rangle}{\langle u_k, u_k \rangle}. \tag{6.21}$$

For this reason, it is customary to construct the u_j or the q_j's one by one, from the first to the last, using (6.21). This process is called **Gram-Schmidt orthogonalization**.

Since any 1-1 column map into a finite-dimensional vector space can be extended to a basis for that vector space, we have also proved the following.

(6.22) Corollary: Every o.n. column map Q into a finite-dimensional inner product space can be extended to an o.n. basis for that space.

An o.n. column map Q has many special properties, all of which derive from the defining property, $Q^c Q = \text{id}$, by the observation that therefore, for any $a, b \in \mathbb{F}^n$,

$$\langle Qa, Qb \rangle = \langle Q^c Q a, b \rangle = \langle a, b \rangle.$$

For example, any o.n. $Q \in L(\mathbb{F}^n, X)$ is an **isometry** in the sense that

$$\forall a \in \mathbb{F}^n \quad \|Qa\| = \|a\|. \tag{6.23}$$

More than that, any o.n. $Q \in L(\mathbb{F}^n, X)$ is **inner-product preserving** in the sense that $\langle Qa, Qb\rangle = \langle a, b\rangle$ for any $a, b \in \mathbb{F}^n$. This is also called **angle-preserving** since a standard definition of the **angle** φ between two real nonzero n-vectors x and y is the following implicit one:

$$\cos(\varphi) := \frac{\langle x, y\rangle}{\|x\|\|y\|}.$$

Note that this definition makes good sense since, by (6.15)Cauchy's Inequality, the righthand side lies in the interval $[-1\,..\,1]$.

Given any 1-1 matrix `V`, the `MATLAB` command `[q,r] = qr(V,0)` provides an o.n. basis, `q`, for ran `V`, along with the upper triangular matrix `r` for which `q*r` equals `V`. The (simpler) statement `[Q,R]=qr(V)` provides a **unitary**, i.e., a *square* o.n., matrix `Q` and an upper triangular matrix `R` so that `Q*R` equals `V`. If `V` is itself square, then `q` equals `Q`. In the contrary case, `Q` equals `[q,U]` for some o.n. basis `U` of the orthogonal complement of ran`V`. Finally, the simplest possible statement, `p = qr(V)`, gives the most complicated result, namely a matrix `p` of the same size as `V` that contains `r` in its upper triangular part and complete information about the various Householder matrices used in its strictly lower triangular part.

While, for each $j = 1{:}\#$`V`, ran`V`$(:, [1{:}j])$ = ran`Q`$(:, [1{:}j])$, the construction of `q` or `Q` does not involve the Gram-Schmidt algorithm, as that algorithm is not reliable numerically when applied to an arbitrary 1-1 matrix `V`. Rather, the matrix `V` is factored column by column with the aid of certain elementary matrices, the so-called **Householder reflections** $\mathrm{id} - 2ww^{\mathrm{c}}/w^{\mathrm{c}}w$.

□

It is customary to call a *real* unitary matrix **orthogonal**. However, the columns of such an orthogonal matrix are not just orthogonal to each other, they are also normalized. Thus it would be better to call such a matrix 'orthonormal', freeing the term 'orthogonal matrix' to denote one whose columns are merely orthogonal to each other. But such naming conventions are hard to change. I will simply not use the term 'orthogonal matrix', but use 'real unitary matrix' instead.

6.10 Prove that $V = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 0 & 1 & 2 \end{bmatrix}$ is a basis for $\mathbb{R}^3$ and compute the coordinates of $x := (1, 1, 1)$ with respect to V.

6.11 Verify that $V = \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ is an orthogonal basis for its range, and extend it to an orthogonal basis for $\mathbb{R}^4$.

6.12 (a) Use the calculations in H.P. 6.7 to construct an orthogonal basis for Π_2 from the power basis $V = [()^0, ()^1, ()^2]$ with respect to the (discrete) inner product in 6.7.

(b) Use (a) to compute the discrete least-squares approximation from Π_2 to the data (j, j^3), $j = 1{:}10$.

6.13 Taking for granted that the space $C[-1\,..\,1]$ of real-valued continuous functions on the interval $[-1\,..\,1]$ is an inner product space with respect to the inner product

$$\langle f, g\rangle := \int_{-1}^{1} f(t)g(t)\,\mathrm{d}t,$$

consider the orthogonal projector onto Π_1.

(a) Verify that $[()^0, ()^1]$ is an orthogonal basis for Π_1. (b) Use (a) to construct the orthogonal projection of $()^2$ onto Π_1. (c) Use (a) and (b) to construct an o.n. basis for Π_2.

6.14 What is the angle between $(1, 2, 2)$ and $(3, -1, -2)$?

6.15 Consider the **Vandermonde** matrix

$$A := [\delta_{z_0}, \ldots, \delta_{z_k}]^{\mathrm{c}}[()^0, \ldots, ()^k] = (z_i^j : i, j = 0{:}k)$$

for some sequence $z_0, \ldots, z_k$ of complex numbers.

Prove that A is a scalar multiple of a unitary matrix if and only if, after some reordering and for some real α,

$$\{z_0, \ldots, z_k\} = \{\exp(2\pi\mathrm{i}(\alpha + i/(k+1))) : i = 0{:}k\}.$$

The inner product space $\mathbb{F}^{m\times n}$ and the trace of a matrix

At the outset of these notes, we introduced the space $\mathbb{F}^{m\times n}$ as a special case of the space $\mathbb{F}^T$ of all scalar-valued functions on some set T, namely with

$$T = \underline{m} \times \underline{n}.$$

This set being finite, there is a natural inner product on $\mathbb{F}^{m\times n}$, namely

$$\langle A, B\rangle := \sum_{i,j} \overline{B(i,j)} A(i,j).$$

This inner product can also be written in the form

$$\langle A, B\rangle = \sum_{i,j} B^{\mathrm{c}}(j,i)A(i,j) = \sum_j (B^{\mathrm{c}}A)(j,j) = \mathrm{trace}(B^{\mathrm{c}}A).$$

Here, the **trace** of a square matrix C is, by definition, the sum of its diagonal entries,

$$\mathrm{trace}\, C := \sum_j C(j,j).$$

The norm in this inner product space is called the **Frobenius norm**,

$$\|A\|_F := \sqrt{\mathrm{trace}\, A^{\mathrm{c}}A} = \sum_{i,j} |A(i,j)|^2. \tag{6.24}$$

The Frobenius norm is **compatible** with the Euclidean norm $\|\ \|$ on $\mathbb{F}^n$ and $\mathbb{F}^m$ in the sense that

$$\|Ax\| \le \|A\|_F \|x\|, \quad x \in \mathbb{F}^n. \tag{6.25}$$

Not surprisingly, the map $\mathbb{F}^{m\times n} \to \mathbb{F}^{n\times m} : A \mapsto A^{\mathrm{t}}$ is unitary, i.e., inner-product preserving:

$$\langle A^{\mathrm{t}}, B^{\mathrm{t}}\rangle = \sum_{i,j} \overline{B^{\mathrm{t}}(i,j)} A^{\mathrm{t}}(i,j) = \sum_{i,j} \overline{B(j,i)} A(j,i) = \langle A, B\rangle. \tag{6.26}$$

In particular,

$$\mathrm{trace}(B^{\mathrm{c}}A) = \mathrm{trace}(AB^{\mathrm{c}}), \quad A, B \in \mathbb{F}^{m\times n}. \tag{6.27}$$

6.16 T/F

(a) $(x,y) \mapsto y^{\mathrm{c}} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x$ is an inner product on $\mathbb{R}^2$.

(b) $\|x+y\|^2 \le \|x\|^2 + \|y\|^2$;