6C *Orthogonal Complements and Minimization Problems*

Orthogonal Complements

6.46 definition: *orthogonal complement,* $U^{\perp}$

If U is a subset of V, then the *orthogonal complement* of U, denoted by $U^{\perp}$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^{\perp} = \{v \in V : \langle u, v \rangle = 0 \text{ for every } u \in U\}.$$

The orthogonal complement $U^{\perp}$ depends on V as well as on U. However, the inner product space V should always be clear from the context and thus it can be omitted from the notation.

6.47 example: *orthogonal complements*

- If $V = \mathbf{R}^3$ and U is the subset of V consisting of the single point $(2, 3, 5)$, then $U^{\perp}$ is the plane $\{(x, y, z) \in \mathbf{R}^3 : 2x + 3y + 5z = 0\}$.
- If $V = \mathbf{R}^3$ and U is the plane $\{(x, y, z) \in \mathbf{R}^3 : 2x + 3y + 5z = 0\}$, then $U^{\perp}$ is the line $\{(2t, 3t, 5t) : t \in \mathbf{R}\}$.
- More generally, if U is a plane in $\mathbf{R}^3$ containing the origin, then $U^{\perp}$ is the line containing the origin that is perpendicular to U.
- If U is a line in $\mathbf{R}^3$ containing the origin, then $U^{\perp}$ is the plane containing the origin that is perpendicular to U.
- If $V = \mathbf{F}^5$ and $U = \{(a, b, 0, 0, 0) \in \mathbf{F}^5 : a, b \in \mathbf{F}\}$, then
$$U^{\perp} = \{(0, 0, x, y, z) \in \mathbf{F}^5 : x, y, z \in \mathbf{F}\}.$$
- If $e_1, \dots, e_m, f_1, \dots, f_n$ is an orthonormal basis of V, then
$$\big(\operatorname{span}(e_1, \dots, e_m)\big)^{\perp} = \operatorname{span}(f_1, \dots, f_n).$$

We begin with some straightforward consequences of the definition.

6.48 *properties of orthogonal complement*

(a) If U is a subset of V, then $U^{\perp}$ is a subspace of V.

(b) $\{0\}^{\perp} = V$.

(c) $V^{\perp} = \{0\}$.

(d) If U is a subset of V, then $U \cap U^{\perp} \subseteq \{0\}$.

(e) If G and H are subsets of V and $G \subseteq H$, then $H^{\perp} \subseteq G^{\perp}$.

Proof

(a) Suppose U is a subset of V. Then $\langle u, 0 \rangle = 0$ for every $u \in U$; thus $0 \in U^\perp$.

Suppose $v, w \in U^\perp$. If $u \in U$, then

$$\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle = 0 + 0 = 0.$$

Thus $v + w \in U^\perp$, which shows that $U^\perp$ is closed under addition.

Similarly, suppose $\lambda \in \mathbf{F}$ and $v \in U^\perp$. If $u \in U$, then

$$\langle u, \lambda v \rangle = \overline{\lambda} \langle u, v \rangle = \overline{\lambda} \cdot 0 = 0.$$

Thus $\lambda v \in U^\perp$, which shows that $U^\perp$ is closed under scalar multiplication. Thus $U^\perp$ is a subspace of V.

(b) Suppose that $v \in V$. Then $\langle 0, v \rangle = 0$, which implies that $v \in \{0\}^\perp$. Thus $\{0\}^\perp = V$.

(c) Suppose that $v \in V^\perp$. Then $\langle v, v \rangle = 0$, which implies that $v = 0$. Thus $V^\perp = \{0\}$.

(d) Suppose U is a subset of V and $u \in U \cap U^\perp$. Then $\langle u, u \rangle = 0$, which implies that $u = 0$. Thus $U \cap U^\perp \subseteq \{0\}$.

(e) Suppose G and H are subsets of V and $G \subseteq H$. Suppose $v \in H^\perp$. Then $\langle u, v \rangle = 0$ for every $u \in H$, which implies that $\langle u, v \rangle = 0$ for every $u \in G$. Hence $v \in G^\perp$. Thus $H^\perp \subseteq G^\perp$. ■

Recall that if U and W are subspaces of V, then V is the direct sum of U and W (written $V = U \oplus W$) if each element of V can be written in exactly one way as a vector in U plus a vector in W (see 1.41). Furthermore, this happens if and only if $V = U + W$ and $U \cap W = \{0\}$ (see 1.46).

The next result shows that every finite-dimensional subspace of V leads to a natural direct sum decomposition of V. See Exercise 16 for an example showing that the result below can fail without the hypothesis that the subspace U is finite-dimensional.

6.49 *direct sum of a subspace and its orthogonal complement*

Suppose U is a finite-dimensional subspace of V. Then

$$V = U \oplus U^\perp.$$

Proof First we will show that

$$V = U + U^\perp.$$

To do this, suppose that $v \in V$. Let $e_1, \dots, e_m$ be an orthonormal basis of U. We want to write v as the sum of a vector in U and a vector orthogonal to U.

We have

6.50
$$v = \underbrace{\langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m}_{u} + \underbrace{v - \langle v, e_1\rangle e_1 - \cdots - \langle v, e_m\rangle e_m}_{w}.$$

Let u and w be defined as in the equation above (as was done in the proof of 6.26). Because each $e_k \in U$, we see that $u \in U$. Because $e_1, \ldots, e_m$ is an orthonormal list, for each $k = 1, \ldots, m$ we have

$$\begin{aligned}\langle w, e_k\rangle &= \langle v, e_k\rangle - \langle v, e_k\rangle \\ &= 0.\end{aligned}$$

Thus w is orthogonal to every vector in $\operatorname{span}(e_1, \ldots, e_m)$, which shows that $w \in U^{\perp}$. Hence we have written $v = u + w$, where $u \in U$ and $w \in U^{\perp}$, completing the proof that $V = U + U^{\perp}$.

From 6.48(d), we know that $U \cap U^{\perp} = \{0\}$. Now equation $V = U + U^{\perp}$ implies that $V = U \oplus U^{\perp}$ (see 1.46). ■

Now we can see how to compute $\dim U^{\perp}$ from $\dim U$.

6.51 *dimension of orthogonal complement*

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U^{\perp} = \dim V - \dim U.$$

Proof The formula for $\dim U^{\perp}$ follows immediately from 6.49 and 3.94. ■

The next result is an important consequence of 6.49.

6.52 *orthogonal complement of the orthogonal complement*

Suppose U is a finite-dimensional subspace of V. Then

$$U = (U^{\perp})^{\perp}.$$

Proof First we will show that

6.53
$$U \subseteq (U^{\perp})^{\perp}.$$

To do this, suppose $u \in U$. Then $\langle u, w\rangle = 0$ for every $w \in U^{\perp}$ (by the definition of $U^{\perp}$). Because u is orthogonal to every vector in $U^{\perp}$, we have $u \in (U^{\perp})^{\perp}$, completing the proof of 6.53.

To prove the inclusion in the other direction, suppose $v \in (U^{\perp})^{\perp}$. By 6.49, we can write $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. We have $v - u = w \in U^{\perp}$. Because $v \in (U^{\perp})^{\perp}$ and $u \in (U^{\perp})^{\perp}$ (from 6.53), we have $v - u \in (U^{\perp})^{\perp}$. Thus $v - u \in U^{\perp} \cap (U^{\perp})^{\perp}$, which implies that $v - u = 0$ [by 6.48(d)], which implies that $v = u$, which implies that $v \in U$. Thus $(U^{\perp})^{\perp} \subseteq U$, which along with 6.53 completes the proof. ■

Exercise 16(a) shows that the result below is not true without the hypothesis that U is finite-dimensional.

Suppose U is a subspace of V and we want to show that $U = V$. In some situations, the easiest way to do this is to show that the only vector orthogonal to U is 0, and then use the result below. For example, the result below is useful for Exercise 4.

6.54 $U^\perp = \{0\} \iff U = V$ *(for U a finite-dimensional subspace of V)*

Suppose U is a finite-dimensional subspace of V. Then

$$U^\perp = \{0\} \iff U = V.$$

Proof First suppose $U^\perp = \{0\}$. Then by 6.52, $U = \left(U^\perp\right)^\perp = \{0\}^\perp = V$, as desired.

Conversely, if $U = V$, then $U^\perp = V^\perp = \{0\}$ by 6.48(c). ■

We now define an operator P_U for each finite-dimensional subspace U of V.

6.55 definition: *orthogonal projection, P_U*

Suppose U is a finite-dimensional subspace of V. The *orthogonal projection* of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as follows: For each $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then let $P_U v = u$.

The direct sum decomposition $V = U \oplus U^\perp$ given by 6.49 shows that each $v \in V$ can be uniquely written in the form $v = u + w$ with $u \in U$ and $w \in U^\perp$. Thus $P_U v$ is well defined. See the figure that accompanies the proof of 6.61 for the picture describing $P_U v$ that you should keep in mind.

6.56 example: *orthogonal projection onto one-dimensional subspace*

Suppose $u \in V$ with $u \neq 0$ and U is the one-dimensional subspace of V defined by $U = \operatorname{span}(u)$.

If $v \in V$, then

$$v = \frac{\langle v, u\rangle}{\|u\|^2}u + \left(v - \frac{\langle v, u\rangle}{\|u\|^2}u\right),$$

where the first term on the right is in $\operatorname{span}(u)$ (and thus is in U) and the second term on the right is orthogonal to u (and thus is in $U^\perp$). Thus $P_U v$ equals the first term on the right. In other words, we have the formula

$$P_U v = \frac{\langle v, u\rangle}{\|u\|^2}u$$

for every $v \in V$.

The formula above becomes $P_U u = u$ if $v = u$ and becomes $P_U v = 0$ if $v \in \{u\}^\perp$. These equations are special cases of (b) and (c) in the next result.

6.57 ***properties of orthogonal projection*** P_U

Suppose U is a finite-dimensional subspace of V. Then

(a) $P_U \in \mathcal{L}(V)$;

(b) $P_U u = u$ for every $u \in U$;

(c) $P_U w = 0$ for every $w \in U^\perp$;

(d) $\operatorname{range} P_U = U$;

(e) $\operatorname{null} P_U = U^\perp$;

(f) $v - P_U v \in U^\perp$ for every $v \in V$;

(g) $P_U^2 = P_U$;

(h) $\|P_U v\| \le \|v\|$ for every $v \in V$;

(i) if $e_1, \dots, e_m$ is an orthonormal basis of U and $v \in V$, then

$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m.$$

Proof

(a) To show that P_U is a linear map on V, suppose $v_1, v_2 \in V$. Write

$$v_1 = u_1 + w_1 \quad \text{and} \quad v_2 = u_2 + w_2$$

with $u_1, u_2 \in U$ and $w_1, w_2 \in U^\perp$. Thus $P_U v_1 = u_1$ and $P_U v_2 = u_2$. Now

$$v_1 + v_2 = (u_1 + u_2) + (w_1 + w_2),$$

where $u_1 + u_2 \in U$ and $w_1 + w_2 \in U^\perp$. Thus

$$P_U(v_1 + v_2) = u_1 + u_2 = P_U v_1 + P_U v_2.$$

Similarly, suppose $\lambda \in \mathbf{F}$ and $v \in V$. Write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then $\lambda v = \lambda u + \lambda w$ with $\lambda u \in U$ and $\lambda w \in U^\perp$. Thus $P_U(\lambda v) = \lambda u = \lambda P_U v$.

Hence P_U is a linear map from V to V.

(b) Suppose $u \in U$. We can write $u = u + 0$, where $u \in U$ and $0 \in U^\perp$. Thus $P_U u = u$.

(c) Suppose $w \in U^\perp$. We can write $w = 0 + w$, where $0 \in U$ and $w \in U^\perp$. Thus $P_U w = 0$.

(d) The definition of P_U implies that $\operatorname{range} P_U \subseteq U$. Furthermore, (b) implies that $U \subseteq \operatorname{range} P_U$. Thus $\operatorname{range} P_U = U$.

(e) The inclusion $U^\perp \subseteq \operatorname{null} P_U$ follows from (c). To prove the inclusion in the other direction, note that if $v \in \operatorname{null} P_U$ then the decomposition given by 6.49 must be $v = 0 + v$, where $0 \in U$ and $v \in U^\perp$. Thus $\operatorname{null} P_U \subseteq U^\perp$.

(f) If $v \in V$ and $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$v - P_U v = v - u = w \in U^\perp.$$

(g) If $v \in V$ and $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$(P_U{}^2)v = P_U(P_U v) = P_U u = u = P_U v.$$

(h) If $v \in V$ and $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$\|P_U v\|^2 = \|u\|^2 \leq \|u\|^2 + \|w\|^2 = \|v\|^2,$$

where the last equality comes from the Pythagorean theorem.

(i) The formula for $P_U v$ follows from equation 6.50 in the proof of 6.49. ■

In the previous section we proved the Riesz representation theorem (6.42), whose key part states that every linear functional on a finite-dimensional inner product space is given by taking the inner product with some fixed vector. Seeing a different proof often provides new insight. Thus we now give a new proof of the key part of the Riesz representation theorem using orthogonal complements instead of orthonormal bases as in our previous proof.

The restatement below of the Riesz representation theorem provides an identification of V with V'. We will prove only the "onto" part of the result below because the more routine "one-to-one" part of the result can be proved as in 6.42.

Intuition behind this new proof: If $\varphi \in V'$, $v \in V$, and $\varphi(u) = \langle u, v \rangle$ for all $u \in V$, then $v \in (\operatorname{null} \varphi)^\perp$. However, $(\operatorname{null} \varphi)^\perp$ is a one-dimensional subspace of V (except for the trivial case in which $\varphi = 0$), as follows from 6.51 and 3.21. Thus we can obtain v by choosing any nonzero element of $(\operatorname{null} \varphi)^\perp$ and then multiplying by an appropriate scalar, as is done in the proof below.

6.58 *Riesz representation theorem, revisited*

Suppose V is finite-dimensional. For each $v \in V$, define $\varphi_v \in V'$ by

$$\varphi_v(u) = \langle u, v \rangle$$

for each $u \in V$. Then $v \mapsto \varphi_v$ is a one-to-one function from V onto V'.

Proof To show that $v \mapsto \varphi_v$ is surjective, suppose $\varphi \in V'$. If $\varphi = 0$, then $\varphi = \varphi_0$. Thus assume $\varphi \neq 0$. Hence $\operatorname{null} \varphi \neq V$, which implies that $(\operatorname{null} \varphi)^\perp \neq \{0\}$ (by 6.49 with $U = \operatorname{null} \varphi$).

Caution: *The function $v \mapsto \varphi_v$ is a linear mapping from V to V' if $\mathbf{F} = \mathbf{R}$. However, this function is not linear if $\mathbf{F} = \mathbf{C}$ because $\varphi_{\lambda v} = \bar{\lambda}\varphi_v$ if $\lambda \in \mathbf{C}$.*

Let $w \in (\operatorname{null} \varphi)^\perp$ be such that $w \neq 0$. Let

6.59 $$v = \frac{\overline{\varphi(w)}}{\|w\|^2} w.$$

Then $v \in (\operatorname{null} \varphi)^\perp$. Also, $v \neq 0$ (because $w \notin \operatorname{null} \varphi$).

Taking the norm of both sides of 6.59 gives

6.60
$$\|v\| = \frac{|\varphi(w)|}{\|w\|}.$$

Applying φ to both sides of 6.59 and then using 6.60, we have

$$\varphi(v) = \frac{|\varphi(w)|^2}{\|w\|^2} = \|v\|^2.$$

Now suppose $u \in V$. Using the equation above, we have

$$u = \left(u - \frac{\varphi(u)}{\varphi(v)}v\right) + \frac{\varphi(u)}{\|v\|^2}v.$$

The first term in parentheses above is in null φ and hence is orthogonal to v. Thus taking the inner product of both sides of the equation above with v shows that

$$\langle u, v\rangle = \frac{\varphi(u)}{\|v\|^2}\langle v, v\rangle = \varphi(u).$$

Thus $\varphi = \varphi_v$, showing that $v \mapsto \varphi_v$ is surjective, as desired. ■

See Exercise 13 for yet another proof of the Riesz representation theorem.

Minimization Problems

The following problem often arises: Given a subspace U of V and a point $v \in V$, find a point $u \in U$ such that $\|v - u\|$ is as small as possible. The next result shows that $u = P_U v$ is the unique solution of this minimization problem.

The remarkable simplicity of the solution to this minimization problem has led to many important applications of inner product spaces outside of pure mathematics.

6.61 *minimizing distance to a subspace*

Suppose U is a finite-dimensional subspace of V, $v \in V$, and $u \in U$. Then

$$\|v - P_U v\| \le \|v - u\|.$$

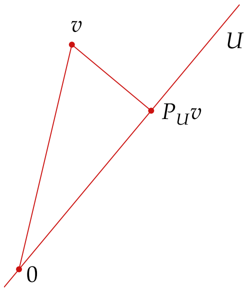
Furthermore, the inequality above is an equality if and only if $u = P_U v$.

Proof We have

6.62
$$\begin{aligned}\|v - P_U v\|^2 &\le \|v - P_U v\|^2 + \|P_U v - u\|^2\\ &= \big\|(v - P_U v) + (P_U v - u)\big\|^2\\ &= \|v - u\|^2,\end{aligned}$$

where the first line above holds because $0 \leq \|P_U v - u\|^2$, the second line above comes from the Pythagorean theorem [which applies because $v - P_U v \in U^\perp$ by 6.57(f), and $P_U v - u \in U$], and the third line above holds by simple algebra. Taking square roots gives the desired inequality.

The inequality proved above is an equality if and only if 6.62 is an equality, which happens if and only if $\|P_U v - u\| = 0$, which happens if and only if $u = P_U v$. ■



$P_U v$ is the closest point in U to v.

The last result is often combined with the formula 6.57(i) to compute explicit solutions to minimization problems, as in the following example.

6.63 example: *using linear algebra to approximate the sine function*

Suppose we want to find a polynomial u with real coefficients and of degree at most 5 that approximates the sine function as well as possible on the interval $[-\pi, \pi]$, in the sense that

$$\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \, dx$$

is as small as possible.

Let $C[-\pi, \pi]$ denote the real inner product space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

6.64
$$\langle f, g \rangle = \int_{-\pi}^{\pi} fg.$$

Let $v \in C[-\pi, \pi]$ be the function defined by $v(x) = \sin x$. Let U denote the subspace of $C[-\pi, \pi]$ consisting of the polynomials with real coefficients and of degree at most 5. Our problem can now be reformulated as follows:

Find $u \in U$ such that $\|v - u\|$ is as small as possible.

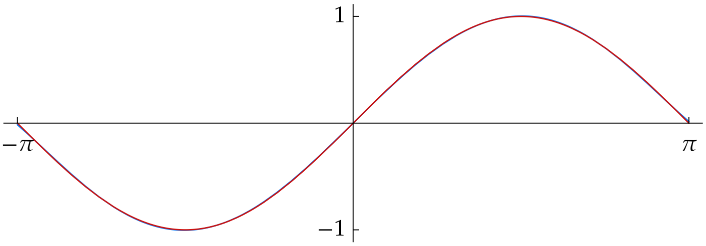
A computer that can integrate is useful here.

To compute the solution to our approximation problem, first apply the Gram–Schmidt procedure (using the inner product given by 6.64) to the basis $1, x, x^2, x^3, x^4, x^5$ of U, producing an orthonormal basis $e_1, e_2, e_3, e_4, e_5, e_6$ of U. Then, again using the inner product given by 6.64, compute $P_U v$ using 6.57(i) (with $m = 6$). Doing this computation shows that $P_U v$ is the function u defined by

6.65
$$u(x) = 0.987862x - 0.155271x^3 + 0.00564312x^5,$$

where the π's that appear in the exact answer have been replaced with a good decimal approximation. By 6.61, the polynomial u above is the best approximation to the sine function on $[-\pi, \pi]$ using polynomials of degree at most 5 (here "best approximation" means in the sense of minimizing $\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \, dx$).

To see how good this approximation is, the next figure shows the graphs of both the sine function and our approximation u given by 6.65 over the interval $[-\pi, \pi]$.



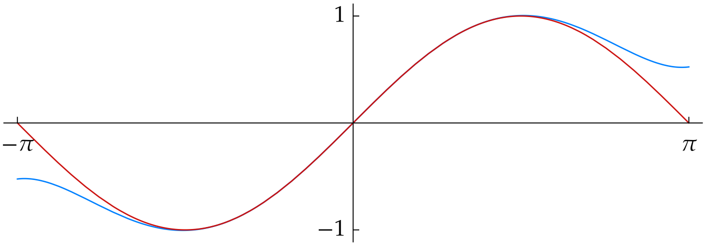
Graphs on $[-\pi, \pi]$ of the sine function (red) and its best fifth degree polynomial approximation u (blue) from 6.65.

Our approximation 6.65 is so accurate that the two graphs are almost identical—our eyes may see only one graph! Here the red graph is placed almost exactly over the blue graph. If you are viewing this on an electronic device, enlarge the picture above by 400% near π or $-\pi$ to see a small gap between the two graphs.

Another well-known approximation to the sine function by a polynomial of degree 5 is given by the Taylor polynomial p defined by

6.66
$$p(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!}.$$

To see how good this approximation is, the next picture shows the graphs of both the sine function and the Taylor polynomial p over the interval $[-\pi, \pi]$.



Graphs on $[-\pi, \pi]$ of the sine function (red) and the Taylor polynomial (blue) from 6.66.

The Taylor polynomial of degree 5 is an excellent approximation to $\sin x$ for x near 0. But the picture above shows that for $|x| > 2$, the Taylor polynomial is not so accurate, especially compared to 6.65. For example, taking $x = 3$, our approximation 6.65 estimates $\sin 3$ with an error of approximately 0.001, but the Taylor polynomial 6.66 estimates $\sin 3$ with an error of approximately 0.4. Thus at $x = 3$, the error in the Taylor polynomial is hundreds of times larger than the error given by 6.65. Linear algebra has helped us discover an approximation to the sine function that improves upon what we learned in calculus!

Pseudoinverse

Suppose $T \in \mathcal{L}(V, W)$ and $w \in W$. Consider the problem of finding $v \in V$ such that

$$Tv = w.$$

For example, if $V = \mathbf{F}^n$ and $W = \mathbf{F}^m$, then the equation above could represent a system of m linear equations in n unknowns $v_1, \dots, v_n$, where $v = (v_1, \dots, v_n)$.

If T is invertible, then the unique solution to the equation above is $v = T^{-1}w$. However, if T is not invertible, then for some $w \in W$ there may not exist any solutions of the equation above, and for some $w \in W$ there may exist infinitely many solutions of the equation above.

If T is not invertible, then we can still try to do as well as possible with the equation above. For example, if the equation above has no solutions, then instead of solving the equation $Tv - w = 0$, we can try to find $v \in V$ such that $\|Tv - w\|$ is as small as possible. As another example, if the equation above has infinitely many solutions $v \in V$, then among all those solutions we can try to find one such that $\|v\|$ is as small as possible.

The pseudoinverse will provide the tool to solve the equation above as well as possible, even when T is not invertible. We need the next result to define the pseudoinverse.

In the next two proofs, we will use without further comment the result that if V is finite-dimensional and $T \in \mathcal{L}(V, W)$, then null T, $(\operatorname{null} T)^\perp$, and range T are all finite-dimensional.

6.67 *restriction of a linear map to obtain a one-to-one and onto map*

Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then $T|_{(\operatorname{null} T)^\perp}$ is an injective map of $(\operatorname{null} T)^\perp$ onto range T.

Proof Suppose that $v \in (\operatorname{null} T)^\perp$ and $T|_{(\operatorname{null} T)^\perp} v = 0$. Hence $Tv = 0$ and thus $v \in (\operatorname{null} T) \cap (\operatorname{null} T)^\perp$, which implies that $v = 0$ [by 6.48(d)]. Hence $\operatorname{null} T|_{(\operatorname{null} T)^\perp} = \{0\}$, which implies that $T|_{(\operatorname{null} T)^\perp}$ is injective, as desired.

Clearly range $T|_{(\operatorname{null} T)^\perp} \subseteq$ range T. To prove the inclusion in the other direction, suppose $w \in$ range T. Hence there exists $v \in V$ such that $w = Tv$. There exist $u \in \operatorname{null} T$ and $x \in (\operatorname{null} T)^\perp$ such that $v = u + x$ (by 6.49). Now

$$T|_{(\operatorname{null} T)^\perp} x = Tx = Tv - Tu = w - 0 = w,$$

which shows that $w \in \operatorname{range} T|_{(\operatorname{null} T)^\perp}$. Hence range $T \subseteq \operatorname{range} T|_{(\operatorname{null} T)^\perp}$, completing the proof that $\operatorname{range} T|_{(\operatorname{null} T)^\perp} = \operatorname{range} T$. ■

To produce the pseudoinverse notation $T^\dagger$ in TeX, type `T^\dagger`.

Now we can define the pseudoinverse $T^\dagger$ (pronounced "T dagger") of a linear map T. In the next definition (and from now on), think of $T|_{(\operatorname{null} T)^\perp}$ as an invertible linear map from $(\operatorname{null} T)^\perp$ onto range T, as is justified by the result above.

6.68 definition: *pseudoinverse*, $T^\dagger$

Suppose that V is finite-dimensional and $T \in \mathcal{L}(V, W)$. The *pseudoinverse* $T^\dagger \in \mathcal{L}(W, V)$ of T is the linear map from W to V defined by

$$T^\dagger w = (T|_{(\operatorname{null} T)^\perp})^{-1} P_{\operatorname{range} T}\, w$$

for each $w \in W$.

Recall that $P_{\operatorname{range} T}\, w = 0$ if $w \in (\operatorname{range} T)^\perp$ and $P_{\operatorname{range} T}\, w = w$ if $w \in \operatorname{range} T$. Thus if $w \in (\operatorname{range} T)^\perp$, then $T^\dagger w = 0$, and if $w \in \operatorname{range} T$, then $T^\dagger w$ is the unique element of $(\operatorname{null} T)^\perp$ such that $T(T^\dagger w) = w$.

The pseudoinverse behaves much like an inverse, as we will see.

6.69 *algebraic properties of the pseudoinverse*

Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$.

(a) If T is invertible, then $T^\dagger = T^{-1}$.

(b) $TT^\dagger = P_{\operatorname{range} T} =$ the orthogonal projection of W onto $\operatorname{range} T$.

(c) $T^\dagger T = P_{(\operatorname{null} T)^\perp} =$ the orthogonal projection of V onto $(\operatorname{null} T)^\perp$.

Proof

(a) Suppose T is invertible. Then $(\operatorname{null} T)^\perp = V$ and $\operatorname{range} T = W$. Thus $T|_{(\operatorname{null} T)^\perp} = T$ and $P_{\operatorname{range} T}$ is the identity operator on W. Hence $T^\dagger = T^{-1}$.

(b) Suppose $w \in \operatorname{range} T$. Thus

$$TT^\dagger w = T(T|_{(\operatorname{null} T)^\perp})^{-1} w = w = P_{\operatorname{range} T}\, w.$$

If $w \in (\operatorname{range} T)^\perp$, then $T^\dagger w = 0$. Hence $TT^\dagger w = 0 = P_{\operatorname{range} T}\, w$. Thus $TT^\dagger$ and $P_{\operatorname{range} T}$ agree on $\operatorname{range} T$ and on $(\operatorname{range} T)^\perp$. Hence these two linear maps are equal (by 6.49).

(c) Suppose $v \in (\operatorname{null} T)^\perp$. Because $Tv \in \operatorname{range} T$, the definition of $T^\dagger$ shows that

$$T^\dagger(Tv) = (T|_{(\operatorname{null} T)^\perp})^{-1}(Tv) = v = P_{(\operatorname{null} T)^\perp} v.$$

If $v \in \operatorname{null} T$, then $T^\dagger T v = 0 = P_{(\operatorname{null} T)^\perp} v$. Thus $T^\dagger T$ and $P_{(\operatorname{null} T)^\perp}$ agree on $(\operatorname{null} T)^\perp$ and on $\operatorname{null} T$. Hence these two linear maps are equal (by 6.49). ∎

The pseudoinverse is also called the Moore–Penrose inverse.

Suppose that $T \in \mathcal{L}(V, W)$. If T is surjective, then $TT^\dagger$ is the identity operator on W, as follows from (b) in the result above. If T is injective, then $T^\dagger T$ is the identity operator on V, as follows from (c) in the result above. For additional algebraic properties of the pseudoinverse, see Exercises 19–23.

For $T \in \mathcal{L}(V, W)$ and $w \in W$, we now return to the problem of finding $v \in V$ that solves the equation

$$Tv = w.$$

As we noted earlier, if T is invertible, then $v = T^{-1}w$ is the unique solution, but if T is not invertible, then T^{-1} is not defined. However, the pseudoinverse $T^{\dagger}$ is defined. Taking $v = T^{\dagger}w$ makes Tv as close to w as possible, as shown by (a) of the next result. Thus the pseudoinverse provides what is called a *best fit* to the equation above.

Among all vectors $v \in V$ that make Tv as close as possible to w, the vector $T^{\dagger}w$ has the smallest norm, as shown by combining (b) in the next result with the condition for equality in (a).

6.70 *pseudoinverse provides best approximate solution or best solution*

Suppose V is finite-dimensional, $T \in \mathcal{L}(V, W)$, and $w \in W$.

(a) If $v \in V$, then

$$\|T(T^{\dagger}w) - w\| \leq \|Tv - w\|,$$

with equality if and only if $v \in T^{\dagger}w + \operatorname{null} T$.

(b) If $v \in T^{\dagger}w + \operatorname{null} T$, then

$$\|T^{\dagger}w\| \leq \|v\|,$$

with equality if and only if $v = T^{\dagger}w$.

Proof

(a) Suppose $v \in V$. Then

$$Tv - w = (Tv - TT^{\dagger}w) + (TT^{\dagger}w - w).$$

The first term in parentheses above is in range T. Because the operator $TT^{\dagger}$ is the orthogonal projection of W onto range T [by 6.69(b)], the second term in parentheses above is in $(\operatorname{range} T)^{\perp}$ [see 6.57(f)].

Thus the Pythagorean theorem implies the desired inequality that the norm of the second term in parentheses above is less than or equal to $\|Tv - w\|$, with equality if and only if the first term in parentheses above equals 0. Hence we have equality if and only if $v - T^{\dagger}w \in \operatorname{null} T$, which is equivalent to the statement that $v \in T^{\dagger}w + \operatorname{null} T$, completing the proof of (a).

(b) Suppose $v \in T^{\dagger}w + \operatorname{null} T$. Hence $v - T^{\dagger}w \in \operatorname{null} T$. Now

$$v = (v - T^{\dagger}w) + T^{\dagger}w.$$

The definition of $T^{\dagger}$ implies that $T^{\dagger}w \in (\operatorname{null} T)^{\perp}$. Thus the Pythagorean theorem implies that $\|T^{\dagger}w\| \leq \|v\|$, with equality if and only if $v = T^{\dagger}w$. ■

A formula for $T^{\dagger}$ will be given in the next chapter (see 7.78).

6.71 example: *pseudoinverse of a linear map from* $\mathbf{F}^4$ *to* $\mathbf{F}^3$

Suppose $T \in \mathcal{L}(\mathbf{F}^4, \mathbf{F}^3)$ is defined by

$$T(a,b,c,d) = (a+b+c, 2c+d, 0).$$

This linear map is neither injective nor surjective, but we can compute its pseudoinverse. To do this, first note that $\operatorname{range} T = \{(x,y,0) : x, y \in \mathbf{F}\}$. Thus

$$P_{\operatorname{range} T}(x,y,z) = (x,y,0)$$

for each $(x,y,z) \in \mathbf{F}^3$. Also,

$$\operatorname{null} T = \{(a,b,c,d) \in \mathbf{F}^4 : a+b+c = 0 \text{ and } 2c+d = 0\}.$$

The list $(-1,1,0,0), (-1,0,1,-2)$ of two vectors in $\operatorname{null} T$ spans $\operatorname{null} T$ because if $(a,b,c,d) \in \operatorname{null} T$ then

$$(a,b,c,d) = b(-1,1,0,0) + c(-1,0,1,-2).$$

Because the list $(-1,1,0,0), (-1,0,1,-2)$ is linearly independent, this list is a basis of $\operatorname{null} T$.

Now suppose $(x,y,z) \in \mathbf{F}^3$. Then

6.72 $$T^\dagger(x,y,z) = \big(T|_{(\operatorname{null} T)^\perp}\big)^{-1} P_{\operatorname{range} T}(x,y,z) = \big(T|_{(\operatorname{null} T)^\perp}\big)^{-1}(x,y,0).$$

The right side of the equation above is the vector $(a,b,c,d) \in \mathbf{F}^4$ such that $T(a,b,c,d) = (x,y,0)$ and $(a,b,c,d) \in (\operatorname{null} T)^\perp$. In other words, a, b, c, d must satisfy the following equations:

$$\begin{aligned} a+b+c &= x \\ 2c+d &= y \\ -a+b &= 0 \\ -a+c-2d &= 0, \end{aligned}$$

where the first two equations are equivalent to the equation $T(a,b,c,d) = (x,y,0)$ and the last two equations come from the condition for (a,b,c,d) to be orthogonal to each of the basis vectors $(-1,1,0,0), (-1,0,1,-2)$ in this basis of $\operatorname{null} T$. Thinking of x and y as constants and a, b, c, d as unknowns, we can solve the system above of four equations in four unknowns, getting

$$a = \tfrac{1}{11}(5x-2y),\ b = \tfrac{1}{11}(5x-2y),\ c = \tfrac{1}{11}(x+4y),\ d = \tfrac{1}{11}(-2x+3y).$$

Hence 6.72 tells us that

$$T^\dagger(x,y,z) = \tfrac{1}{11}(5x-2y, 5x-2y, x+4y, -2x+3y).$$

The formula above for $T^\dagger$ shows that $TT^\dagger(x,y,z) = (x,y,0)$ for all $(x,y,z) \in \mathbf{F}^3$, which illustrates the equation $TT^\dagger = P_{\operatorname{range} T}$ from 6.69(b).

Exercises 6C

1 Suppose $v_1, \dots, v_m \in V$. Prove that

$$\{v_1, \dots, v_m\}^\perp = \big(\operatorname{span}(v_1, \dots, v_m)\big)^\perp.$$

2 Suppose U is a subspace of V with basis $u_1, \dots, u_m$ and

$$u_1, \dots, u_m, v_1, \dots, v_n$$

is a basis of V. Prove that if the Gram–Schmidt procedure is applied to the basis of V above, producing a list $e_1, \dots, e_m, f_1, \dots, f_n$, then $e_1, \dots, e_m$ is an orthonormal basis of U and $f_1, \dots, f_n$ is an orthonormal basis of $U^\perp$.

3 Suppose U is the subspace of $\mathbf{R}^4$ defined by

$$U = \operatorname{span}\big((1, 2, 3, -4), (-5, 4, 3, 2)\big).$$

Find an orthonormal basis of U and an orthonormal basis of $U^\perp$.

4 Suppose $e_1, \dots, e_n$ is a list of vectors in V with $\|e_k\| = 1$ for each $k = 1, \dots, n$ and

$$\|v\|^2 = \big|\langle v, e_1\rangle\big|^2 + \cdots + \big|\langle v, e_n\rangle\big|^2$$

for all $v \in V$. Prove that $e_1, \dots, e_n$ is an orthonormal basis of V.

This exercise provides a converse to 6.30(b).

5 Suppose that V is finite-dimensional and U is a subspace of V. Show that $P_{U^\perp} = I - P_U$, where I is the identity operator on V.

6 Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Show that

$$T = TP_{(\operatorname{null} T)^\perp} = P_{\operatorname{range} T}T.$$

7 Suppose that X and Y are finite-dimensional subspaces of V. Prove that $P_X P_Y = 0$ if and only if $\langle x, y\rangle = 0$ for all $x \in X$ and all $y \in Y$.

8 Suppose U is a finite-dimensional subspace of V and $v \in V$. Define a linear functional $\varphi\colon U \to \mathbf{F}$ by

$$\varphi(u) = \langle u, v\rangle$$

for all $u \in U$. By the Riesz representation theorem (6.42) as applied to the inner product space U, there exists a unique vector $w \in U$ such that

$$\varphi(u) = \langle u, w\rangle$$

for all $u \in U$. Show that $w = P_U v$.

9 Suppose V is finite-dimensional. Suppose $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and every vector in $\operatorname{null} P$ is orthogonal to every vector in $\operatorname{range} P$. Prove that there exists a subspace U of V such that $P = P_U$.

10 Suppose V is finite-dimensional and $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and

$$\|Pv\| \leq \|v\|$$

for every $v \in V$. Prove that there exists a subspace U of V such that $P = P_U$.

11 Suppose $T \in \mathcal{L}(V)$ and U is a finite-dimensional subspace of V. Prove that

$$U \text{ is invariant under } T \iff P_U T P_U = T P_U.$$

12 Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and U is a subspace of V. Prove that

$$U \text{ and } U^\perp \text{ are both invariant under } T \iff P_U T = T P_U.$$

13 Suppose $\mathbf{F} = \mathbf{R}$ and V is finite-dimensional. For each $v \in V$, let φ_v denote the linear functional on V defined by

$$\varphi_v(u) = \langle u, v \rangle$$

for all $u \in V$.

(a) Show that $v \mapsto \varphi_v$ is an injective linear map from V to V'.

(b) Use (a) and a dimension-counting argument to show that $v \mapsto \varphi_v$ is an isomorphism from V onto V'.

The purpose of this exercise is to give an alternative proof of the Riesz representation theorem (6.42 and 6.58) when $\mathbf{F} = \mathbf{R}$. *Thus you should not use the Riesz representation theorem as a tool in your solution.*

14 Suppose that $e_1, \ldots, e_n$ is an orthonormal basis of V. Explain why the dual basis (see 3.112) of $e_1, \ldots, e_n$ is $e_1, \ldots, e_n$ under the identification of V' with V provided by the Riesz representation theorem (6.58).

15 In $\mathbf{R}^4$, let

$$U = \operatorname{span}\big((1, 1, 0, 0), (1, 1, 1, 2)\big).$$

Find $u \in U$ such that $\|u - (1, 2, 3, 4)\|$ is as small as possible.

16 Suppose $C[-1, 1]$ is the vector space of continuous real-valued functions on the interval $[-1, 1]$ with inner product given by

$$\langle f, g \rangle = \int_{-1}^{1} fg$$

for all $f, g \in C[-1, 1]$. Let U be the subspace of $C[-1, 1]$ defined by

$$U = \{f \in C[-1, 1] : f(0) = 0\}.$$

(a) Show that $U^\perp = \{0\}$.

(b) Show that 6.49 and 6.52 do not hold without the finite-dimensional hypothesis.

17 Find $p \in \mathcal{P}_3(\mathbf{R})$ such that $p(0) = 0$, $p'(0) = 0$, and $\int_0^1 \left|2 + 3x - p(x)\right|^2\,dx$ is as small as possible.

18 Find $p \in \mathcal{P}_5(\mathbf{R})$ that makes $\int_{-\pi}^{\pi} \left|\sin x - p(x)\right|^2\,dx$ as small as possible.

The polynomial 6.65 is an excellent approximation to the answer to this exercise, but here you are asked to find the exact solution, which involves powers of π. A computer that can perform symbolic integration should help.

19 Suppose V is finite-dimensional and $P \in \mathcal{L}(V)$ is an orthogonal projection of V onto some subspace of V. Prove that $P^\dagger = P$.

20 Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Show that

$$\operatorname{null} T^\dagger = (\operatorname{range} T)^\perp \quad \text{and} \quad \operatorname{range} T^\dagger = (\operatorname{null} T)^\perp.$$

21 Suppose $T \in \mathcal{L}\bigl(\mathbf{F}^3, \mathbf{F}^2\bigr)$ is defined by

$$T(a, b, c) = (a + b + c, 2b + 3c).$$

(a) For $(x, y) \in \mathbf{F}^2$, find a formula for $T^\dagger(x, y)$.

(b) Verify that the equation $TT^\dagger = P_{\operatorname{range} T}$ from 6.69(b) holds with the formula for $T^\dagger$ obtained in (a).

(c) Verify that the equation $T^\dagger T = P_{(\operatorname{null} T)^\perp}$ from 6.69(c) holds with the formula for $T^\dagger$ obtained in (a).

22 Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that

$$TT^\dagger T = T \quad \text{and} \quad T^\dagger T T^\dagger = T^\dagger.$$

Both formulas above clearly hold if T is invertible because in that case we can replace $T^\dagger$ with T^{-1}.

23 Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that

$$\left(T^\dagger\right)^\dagger = T.$$

The equation above is analogous to the equation $\left(T^{-1}\right)^{-1} = T$ that holds if T is invertible.