

13.13 Projections Revisited

Chapter 4 built projections out of nothing but a direct sum: given $V = U \oplus W$, the operator that keeps the U -piece and discards the W -piece satisfies $P^2 = P$, and every operator satisfying $P^2 = P$ arises this way. Chapter 11 built a second kind by taking $W = U^\perp$, the orthogonal projection P_U . The second is a special case of the first, and Chapter 4 left the comparison open, saying only that its projections “are often called **oblique** projections, to distinguish them from the orthogonal projections met later, in the chapters on inner product spaces”. This section makes the distinction exact, measures how far an oblique projection is from being orthogonal, and closes Part III.

Throughout, V is a finite-dimensional inner product space over $F = \mathbb{R}$ or $\mathbb{C}$, and $n = \dim V$.

13.13.1 Which projections are orthogonal?

Two recollections, because the section is about the gap between them. From Chapter 2: an operator $P \in \mathcal{L}(V)$ with $P^2 = P$ is a **projection**, also called an **idempotent** (Definition 4.5.1); it satisfies $V = \operatorname{im} P \oplus \ker P$, it is the identity on $\operatorname{im} P$ and zero on $\ker P$, and each splitting $V = U \oplus W$ has exactly one projection with image U and kernel W (Theorem 4.5.1). No inner product appears there, and the field was arbitrary.

From Chapter 11: the splitting $V = U \oplus U^\perp$ of Theorem 11.3.2 is one of the many available, and P_U is the projection belonging to it (Definition 11.3.2). Chapter 4 warned that a subspace is the image of infinitely many projections, one per complement. Which one does the inner product pick out, and how is it recognized without being told the kernel?

A projection is orthogonal exactly when it throws away perpendicularly — and that is the same as costing nothing to move across the inner product.

Theorem 13.13.1. Orthogonal Projections Among the Idempotents

Let V be a finite-dimensional inner product space over F and let $P \in \mathcal{L}(V)$ satisfy $P^2 = P$. The following are equivalent.

- (a) P is the orthogonal projection onto its image: $P = P_U$ with $U = \operatorname{im} P$.
- (b) $\ker P = (\operatorname{im} P)^\perp$.
- (c) $P^* = P$.
- (d) P is normal.
- (e) $\|Pv\| \leq \|v\|$ for every $v \in V$.

Idea. Conditions (a) and (b) say one thing twice, once about the operator and once about its two subspaces. Going from (b) to (c) is a computation Chapter 11 has already done for P_U . Coming back from (c) is the four-subspaces theorem: $\ker P^*$ is always $(\operatorname{im} P)^\perp$, so once $P^* = P$ the kernel is forced; normality is the same remark with $\ker P^*$ replaced by $\ker P$. The one direction with work in it is (e) $\Rightarrow$ (b), where a bound on lengths has to become a statement about angles. Feed P a vector $u + tw$ with $u \in \operatorname{im} P$ and $w \in \ker P$: its image is u whatever t is, so choosing t to make $\|u + tw\|$ as small as possible squeezes $\langle u, w \rangle$ to zero.

Proof. **(a) $\Leftrightarrow$ (b).** Write $U = \text{im } P$. ($\Rightarrow$) If $P = P_U$, then $\ker P = \ker P_U = U^\perp$, since P_U is the projection onto U along $U^\perp$ (Definition 11.3.2 and Theorem 4.5.1 (b)). ($\Leftarrow$) If $\ker P = U^\perp$, then P is a projection with image U and kernel $U^\perp$, and so is P_U ; by the uniqueness in Theorem 4.5.1 (b), applied to $V = U \oplus U^\perp$, the two are equal.

(b) $\Rightarrow$ (c). By the equivalence just proved, $P = P_U$ with $U = \text{im } P$. By Theorem 11.3.3 (b), $\langle Pu, v \rangle = \langle u, Pv \rangle$ for all $u, v \in V$. That is the defining identity of the adjoint, so $P^* = P$ by the uniqueness in Definition 11.6.1.

(c) $\Rightarrow$ (d). If $P^* = P$ then $P^*P = P^2 = PP^*$, which is Definition 11.6.3.

(d) $\Rightarrow$ (b). Since P is normal, $\ker P = \ker P^*$ by Corollary 12.3.2 (a). By Theorem 11.6.5 (a), $\ker P^* = (\text{im } P)^\perp$. Combining the two gives (b).

(b) $\Rightarrow$ (e). Let $v \in V$ and split it as $v = Pv + (v - Pv)$. The first piece lies in $\text{im } P$, and the second lies in $\ker P$, because $P(v - Pv) = Pv - P^2v = 0$. By (b) the two pieces are orthogonal, so Theorem 11.1.3 gives

$$\|v\|^2 = \|Pv\|^2 + \|v - Pv\|^2 \geq \|Pv\|^2,$$

and taking nonnegative square roots gives $\|Pv\| \leq \|v\|$.

(e) $\Rightarrow$ (b). We first show $\langle u, w \rangle = 0$ for every $u \in \text{im } P$ and every $w \in \ker P$. There is nothing to prove when $w = 0$, so assume $w \neq 0$, write $z = \langle u, w \rangle$, and put

$$t = -\frac{z}{\|w\|^2}, \quad v = u + tw,$$

which is legitimate because $\|w\| \neq 0$. Then $Pv = Pu + tPw = u$, since P is the identity on $\text{im } P$ and zero on $\ker P$ (Theorem 4.5.1 (a)). Expanding the norm squared and using conjugate symmetry,

$$\begin{aligned} \|v\|^2 &= \|u\|^2 + \bar{t}\langle u, w \rangle + t\langle w, u \rangle + |t|^2\|w\|^2 \\ &= \|u\|^2 - \frac{|z|^2}{\|w\|^2} - \frac{|z|^2}{\|w\|^2} + \frac{|z|^2}{\|w\|^2} = \|u\|^2 - \frac{|z|^2}{\|w\|^2}. \end{aligned}$$

Now (e) applied to this v says $\|u\|^2 = \|Pv\|^2 \leq \|v\|^2$, which forces $|z|^2 \leq 0$ and hence $z = 0$.

So $\ker P \subseteq (\text{im } P)^\perp$. The two subspaces have the same dimension: $\dim \ker P = n - \text{rank } P$ by Theorem 2.3.1, and $\dim(\text{im } P)^\perp = n - \dim \text{im } P = n - \text{rank } P$ by Theorem 11.3.2 (c). A subspace contained in another of the same finite dimension equals it (Theorem 1.6.8), so $\ker P = (\text{im } P)^\perp$. This closes the cycle and proves the theorem. ■

The theorem is the promised comparison. Every projection of Chapter 4 that is **not** self-adjoint is oblique, and now “oblique” has a precise meaning: its kernel is a complement of its image that is not the perpendicular one. Chapter 12 has already met a pair of such projections, in Example (Two Projections That Are Not Orthogonal), and the Quick check after Theorem 12.8.2 traced exactly what their obliqueness cost.

Clause (c) also settles Chapter 4’s warning that “projection onto U ” is incomplete information.

Corollary 13.13.2. One Orthogonal Projection per Subspace

Let U be a subspace of a finite-dimensional inner product space V . Among the projections of V with image U — one for each complement of U , by Theorem 4.5.1 — exactly one is self-adjoint, namely P_U .

Proof. P_U has image U and is self-adjoint by Theorem 13.13.1, (a) $\Rightarrow$ (c). Conversely, let P be a self-adjoint projection with $\text{im } P = U$. By the same theorem, (c) $\Rightarrow$ (b), $\ker P = U^\perp$, which is also $\ker P_U$. Two projections with the same image and the same kernel are equal, by the uniqueness in Theorem 4.5.1 (b). ■

Chapter 4's own example is now decided by inspection. On $\mathbb{R}^2$ with the dot product, $P(x, y) = (x, 0)$ and $P'(x, y) = (x - y, 0)$ are both projections onto the x -axis, with standard matrices

$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{P}' = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}.$$

The standard basis is orthonormal, so Theorem 11.6.3 says self-adjointness is symmetry of these matrices. The first is symmetric, so $P = P_U$ for the x -axis U . The second is not, so P' is oblique — as the picture in Chapter 4 showed, it slides along the line $y = x$ rather than straight down. Clause (e) fails visibly at $\mathbf{v} = (1, -1)$: there $P'\mathbf{v} = (2, 0)$, and $\|P'\mathbf{v}\| = 2$ while $\|\mathbf{v}\| = \sqrt{2}$.

Quick check. Does $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ act on $\mathbb{R}^2$ as a projection? If so, is it an orthogonal projection?

Solution. $\mathbf{A}^2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \mathbf{A}$, so yes, it is a projection. Its image is the span of its columns, $\text{span}((1, 1))$, and its kernel is $\text{span}(\mathbf{e}_2)$. Since $\langle (1, 1), \mathbf{e}_2 \rangle = 1 \neq 0$, the kernel is not $(\text{im } \mathbf{A})^\perp$, so the projection is oblique by Theorem 13.13.1 (b). The quickest route is clause (c): $\mathbf{A}^\top \neq \mathbf{A}$.

Warning. Clause (e) is a statement about every vector, and one vector proves nothing. For the oblique $\mathbf{P}'$ above, $\|\mathbf{P}'\mathbf{e}_1\| = 1 = \|\mathbf{e}_1\|$ and $\|\mathbf{P}'(1, 1)\| = 0 \leq \sqrt{2}$; both are consistent with the bound, and both are worthless as evidence. The bound fails only in the direction $(1, -1)$. If you want to certify a projection as orthogonal by hand, check $\mathbf{P}^* = \mathbf{P}$ in an orthonormal basis, not the length of one image.

Warning. An orthogonal projection is almost never an orthogonal matrix. If $\mathbf{P}$ is unitary and idempotent, it is invertible, so multiplying $\mathbf{P}^2 = \mathbf{P}$ by $\mathbf{P}^{-1}$ gives $\mathbf{P} = \mathbf{I}$. Apart from the identity, every orthogonal projection is singular. “Orthogonal” in P_U names the direction of the discarding, not membership in $O(n)$.

13.13.2 The block that measures obliqueness

Chapter 2 already knows the shape of a projection up to a change of basis: with a basis of $\text{im } P$ followed by a basis of $\ker P$, the matrix is $\text{diag}(1, \dots, 1, 0, \dots, 0)$ with $\text{rank } P$ ones (Corollary 4.5.2). That basis is built to fit P and is usually not orthonormal, so the normal form erases exactly the information we now want. Insisting on an orthonormal basis costs one extra block, and that block is the answer.

Theorem 13.13.3. Canonical Form of an Idempotent in an Orthonormal Basis

Let V be an n -dimensional inner product space over F and let $P \in \mathcal{L}(V)$ be a projection of rank r . Let $(f_1, \dots, f_r)$ be any orthonormal basis of $\text{im } P$ and $(f_{r+1}, \dots, f_n)$ any orthonormal basis of $(\text{im } P)^\perp$. Then $\mathcal{B} = (f_1, \dots, f_n)$ is an orthonormal basis of V , and

$$[P]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} \mathbf{I}_r & \mathbf{X} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

for some $\mathbf{X} \in M_{r \times (n-r)}(F)$. Moreover P is an orthogonal projection if and only if $\mathbf{X} = \mathbf{0}$.

Idea. Only two facts are used, and both are already proved. First, P fixes its image, so the first r columns are forced to be $\mathbf{e}_1, \dots, \mathbf{e}_r$. Second, every output of P lies in $\text{im } P$, so the last $n - r$ columns have nothing below row r . What sits **above** row r in those columns is unconstrained, and that is $\mathbf{X}$: it records where P sends the directions perpendicular to its image, which an orthogonal projection would send to 0.

Proof. Write $U = \text{im } P$, so $\dim U = r$ and $\dim U^\perp = n - r$ by Theorem 11.3.2 (c). Each f_i with $i \leq r$ is orthogonal to each f_j with $j > r$, since the first lies in U and the second in $U^\perp$; so $\mathcal{B}$ is an orthonormal list of n vectors, hence an orthonormal basis of V (Theorem 11.1.4).

For $i \leq r$ we have $f_i \in \text{im } P$, so $Pf_i = f_i$ by Theorem 4.5.1 (a), and column i of $[P]_{\mathcal{B}}^{\mathcal{B}}$ is $\mathbf{e}_i$. For $j > r$ we have $Pf_j \in \text{im } P = U = \text{span}(f_1, \dots, f_r)$, so $Pf_j = \sum_{i \leq r} x_{ij} f_i$ for scalars $x_{ij} \in F$, and column j has zeros in rows $r + 1, \dots, n$. Collecting the x_{ij} into $\mathbf{X}$ gives the displayed shape.

For the last claim, $\mathcal{B}$ is orthonormal, so $[P^*]_{\mathcal{B}}^{\mathcal{B}} = ([P]_{\mathcal{B}}^{\mathcal{B}})^*$ by Theorem 11.6.3, and

$$([P]_{\mathcal{B}}^{\mathcal{B}})^* = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{X}^* & \mathbf{0} \end{pmatrix}.$$

The two matrices agree if and only if $\mathbf{X} = \mathbf{0}$. Since an operator is determined by its matrix in a basis, $P^* = P$ if and only if $\mathbf{X} = \mathbf{0}$, and $P^* = P$ says P is an orthogonal projection by Theorem 13.13.1. ■

The block $\mathbf{X}$ is not unique, because the two orthonormal bases were not. Replacing them changes $\mathbf{X}$ into $\mathbf{A}^* \mathbf{X} \mathbf{B}$ for some unitary $\mathbf{A} \in M_r(F)$ and $\mathbf{B} \in M_{n-r}(F)$, and that leaves the singular values of $\mathbf{X}$ alone: since

$$(\mathbf{A}^* \mathbf{X} \mathbf{B})^* (\mathbf{A}^* \mathbf{X} \mathbf{B}) = \mathbf{B}^* (\mathbf{X}^* \mathbf{X}) \mathbf{B},$$

the two matrices $(\mathbf{A}^*\mathbf{X}\mathbf{B})^*(\mathbf{A}^*\mathbf{X}\mathbf{B})$ and $\mathbf{X}^*\mathbf{X}$ are unitarily similar, so they have the same eigenvalues, and Definition 13.9.1 reads the singular values off those. So the list $\sigma_1(\mathbf{X}) \geq \sigma_2(\mathbf{X}) \geq \dots$ belongs to P itself, and P is an orthogonal projection exactly when that list is all zeros.

13.13.2.1 The two-dimensional case, in full

Take $n = 2$ and $r = 1$, so that $\mathbf{X}$ is a single scalar $x \in F$, and

$$[P]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1 & x \\ 0 & 0 \end{pmatrix}, \quad \mathcal{B} = (f_1, f_2),$$

with f_1 spanning the line $\text{im } P$ and f_2 spanning the perpendicular line $(\text{im } P)^\perp$. Everything about P can be read off this one number.

The kernel. $P(af_1 + bf_2) = (a + xb)f_1$, which is 0 exactly when $a = -xb$. So

$$\ker P = \text{span}(-xf_1 + f_2),$$

a line that meets $(\text{im } P)^\perp = \text{span}(f_2)$ only at 0 unless $x = 0$.

The angle. Suppose $F = \mathbb{R}$. The two lines $\text{im } P$ and $\ker P$ are distinct, since their intersection is $\{0\}$; let $\theta \in (0, \pi/2]$ be the angle between them, meaning the angle of Definition 11.1.5 between f_1 and whichever of $\pm(-xf_1 + f_2)$ makes it acute. With $k = -xf_1 + f_2$ we have $\|k\| = \sqrt{1 + x^2}$ and $\langle f_1, k \rangle = -x$, so

$$\cos \theta = \frac{|x|}{\sqrt{1 + x^2}}, \quad \sin \theta = \frac{1}{\sqrt{1 + x^2}}, \quad |x| = \cot \theta.$$

The dictionary is complete: $x = 0$ is $\theta = \pi/2$, the orthogonal case; and as the kernel tips toward the image, $\theta \rightarrow 0$ and $|x| \rightarrow \infty$. The single entry x is the obliqueness, in the units of a cotangent.

The stretching. For a unit vector $v = af_1 + bf_2$, we have $\|Pv\| = |a + xb| = |\langle (a, b), (1, \bar{x}) \rangle|$ in F^2 with the standard inner product, so Theorem 11.1.5 gives

$$\|Pv\| \leq \sqrt{1 + |x|^2},$$

with equality exactly when (a, b) is a multiple of $(1, \bar{x})$, for instance $(a, b) = (1, \bar{x})/\sqrt{1 + |x|^2}$. So the largest value of $\|Pv\|$ over unit vectors v is $\sqrt{1 + |x|^2}$, which over $\mathbb{R}$ is $1/\sin \theta$. Chapter 16 will give this number a name, the operator norm $\|P\|$; here it is enough to say that an orthogonal projection never lengthens a vector, while an oblique one lengthens some vector by $1/\sin \theta$, a factor that is unbounded as the kernel closes in on the image.

Example. Two Oblique Projections, Measured

For each of

$$\mathbf{P}' = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 2 & -1 \\ 2 & -1 \end{pmatrix}$$

on $\mathbb{R}^2$, verify that it is a projection, put it in the form of Theorem 13.13.3, and find the angle between its image and its kernel and the largest length of an image of a unit vector.

Solution. The matrix $\mathbf{P}'$. $(\mathbf{P}')^2 = \mathbf{P}'$ by direct multiplication, and $\text{im } \mathbf{P}' = \text{span}(\mathbf{e}_1)$, $\ker \mathbf{P}' = \text{span}((1, 1))$. Here $\mathbf{f}_1 = \mathbf{e}_1$ already spans the image and $\mathbf{f}_2 = \mathbf{e}_2$ spans its perpendicular, and the standard basis is orthonormal, so the given matrix is the canonical form, with $x = -1$. Hence $\cot \theta = 1$, that is $\theta = \pi/4$, and the largest value of $\|\mathbf{P}'\mathbf{v}\|$ on the unit circle is $\sqrt{1 + 1} = \sqrt{2}$, attained at $\mathbf{v} = (1, -1)/\sqrt{2}$. Both readings match Chapter 4's figure, where

$\mathbf{P}'$ projects along the line $y = x$.

The matrix $\mathbf{Q}$. $\mathbf{Q}^2 = \begin{pmatrix} 4-2 & -2+1 \\ 4-2 & -2+1 \end{pmatrix} = \mathbf{Q}$, so $\mathbf{Q}$ is a projection, of rank 1. Its image is $\text{span}((1, 1))$ and its kernel is $\text{span}((1, 2))$, since $\mathbf{Q}(1, 2) = (0, 0)$. Take

$$\mathbf{f}_1 = \frac{1}{\sqrt{2}}(1, 1), \quad \mathbf{f}_2 = \frac{1}{\sqrt{2}}(1, -1),$$

an orthonormal basis of $\text{im } \mathbf{Q}$ and of $(\text{im } \mathbf{Q})^\perp$ respectively. Then $\mathbf{Q}\mathbf{f}_1 = \mathbf{f}_1$ and $\mathbf{Q}\mathbf{f}_2 = \frac{1}{\sqrt{2}}(3, 3) = 3\mathbf{f}_1$, so $x = 3$ and

$$[T_{\mathbf{Q}}]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1 & 3 \\ 0 & 0 \end{pmatrix}.$$

Hence $\cot \theta = 3$. Checking against the two lines directly: the angle between $(1, 1)$ and $(1, 2)$ has $\cos \theta = 3/\sqrt{10}$ and $\sin \theta = 1/\sqrt{10}$, so $\cot \theta = 3$, and the largest value of $\|\mathbf{Q}\mathbf{v}\|$ on the unit circle is $\sqrt{1+9} = \sqrt{10} = 1/\sin \theta$. The two routes agree.

In higher rank the same block $\mathbf{X}$ does the same job, and [Exercise \(C3: How far an oblique projection can stretch\)](#) turns its largest singular value into the largest value of $\|Pv\|$ over unit v , by the argument that just ran in the 2×2 case. The reason the singular values appear is the one this chapter has used throughout: $\mathbf{X}$ is a rectangular matrix, so its eigenvalues do not exist, and Theorem 13.9.1 is what measures it instead.

13.13.3 Looking back at Part III

Part III added one piece of structure to the vector spaces of Parts I and II, and spent three chapters finding out what it buys.

Chapter 11 added the inner product, and with it length, orthogonality, and angle. Orthonormal bases made coordinates free, orthogonal complements split every space canonically, and projection onto a subspace solved the nearest-point problem, hence least squares. Its deepest construction was the **adjoint**, the price of moving an operator from one slot of the inner product to the other; the adjoint is what turned the kernel and the image into two orthogonal pairs of subspaces.

Chapter 12 asked which operators the inner product can diagonalize, and the adjoint answered: over $\mathbb{C}$ the normal ones, over $\mathbb{R}$ the self-adjoint ones. Written without bases, the spectral theorem says $T = \sum_i \lambda_i P_i$ with the P_i orthogonal projections onto the eigenspaces — which is why the present section had to exist, and why Theorem 12.8.1 needed clause (b) at all. Once a matrix is diagonal in an orthonormal basis, anything can be done to its eigenvalues and undone afterwards, and that is the functional calculus.

Chapter 13 asked which self-adjoint operators behave like nonnegative numbers, collected the equivalent answers in Theorem 13.1.4, and used them twice: to build the square root and the absolute value $|\mathbf{A}| = (\mathbf{A}^* \mathbf{A})^{1/2}$ of Definition 13.2.1, and then to get past the restriction that had hung over the whole part. The spectral theorem needs a square operator, and most matrices are not square — but $\mathbf{A}^* \mathbf{A}$ always is square, self-adjoint, and positive semidefinite, whatever $\mathbf{A}$ is. Applying the spectral theorem to it and reading the result back through $\mathbf{A}$ gives Theorem 13.9.1, and with it orthonormal bases of all four fundamental subspaces (Corollary 13.9.4), the polar decomposition of Theorem 13.10.1, the best low-rank approximation, and the pseudoinverse. A rectangular matrix has no eigenvalues; it has singular values, and they do the work.

So the part is one chain, each link an application of the last: **inner product** $\rightarrow$ **adjoint** $\rightarrow$ **spectral theorem** $\rightarrow$ **singular value decomposition**. Its named moves travel. To show a vector is zero, pair it with itself. To move an operator across the inner product, pay with the adjoint. Work with the norm squared. Fix an orthonormal basis. Project, then subtract. Test positivity with the single scalar $\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle$, and factor a positive matrix as $\mathbf{B}^*\mathbf{B}$. Diagonalize, work on the eigenvalues, come back. When the matrix is not square, apply the spectral theorem to $\mathbf{A}^*\mathbf{A}$. And when the answer is a sum ordered by size, truncate it.

What comes next takes the quadratic form $\mathbf{x} \mapsto \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle$, which was only a test in this chapter, and makes it the object of study: Chapter 14 classifies forms up to congruence, by inertia, over $\mathbb{R}$ and over $\mathbb{C}$. Chapter 16 gives the operator norm promised twice above, and Chapter 17 characterizes the eigenvalues of a Hermitian matrix as maxima and minima of $\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle$ — at which point the test becomes the theory.

13.13.4 Exercises

13.13.4.1 A. Check your understanding

Exercise. A1

- Define what it means for $P \in \mathcal{L}(V)$ to be a projection, and what it means for it to be the orthogonal projection onto a subspace U . Which of the two definitions needs an inner product?
- State three conditions on a projection P , each equivalent to “ P is an orthogonal projection”.
- Determine whether the following is correct, and justify your answer: if $\|Pv\| \leq \|v\|$ for some nonzero v , then the projection P is orthogonal.
- Determine whether the following is correct, and justify your answer: every projection of rank r on an n -dimensional inner product space has matrix $\mathbf{I}_r \oplus \mathbf{O}$ in some orthonormal basis.

Solution.

- P is a projection when $P^2 = P$ (Definition 4.5.1); P is the orthogonal projection onto U when it is the projection onto U along $U^\perp$ (Definition 11.3.2). Only the second needs an inner product, since $U^\perp$ is undefined without one.
- By Theorem 13.13.1: $\ker P = (\operatorname{im} P)^\perp$; $P^* = P$; P is normal; $\|Pv\| \leq \|v\|$ for every v . Any three of these will do.
- Incorrect. Take $\mathbf{P}' = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$ on $\mathbb{R}^2$ and $\mathbf{v} = (1, 1)$. Then $\mathbf{P}'\mathbf{v} = \mathbf{0}$, so $\|\mathbf{P}'\mathbf{v}\| = 0 \leq \sqrt{2}$, yet $\mathbf{P}'$ is oblique because $(\mathbf{P}')^\top \neq \mathbf{P}'$. Clause (e) of Theorem 13.13.1 quantifies over every $\mathbf{v}$.
- Incorrect in general, and correct exactly for the orthogonal projections. By Theorem 13.13.3 the matrix in a suitable orthonormal basis is $\begin{pmatrix} \mathbf{I}_r & \mathbf{X} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}$, and it is $\mathbf{I}_r \oplus \mathbf{O}$ only when $\mathbf{X} = \mathbf{O}$. If some orthonormal basis gave $\mathbf{I}_r \oplus \mathbf{O}$, that matrix would be Hermitian, so P would be self-adjoint by Theorem 11.6.3. The matrix $\mathbf{P}'$ of part (c) is a counterexample. (In a basis that is merely a basis, $\mathbf{I}_r \oplus \mathbf{O}$ is always attainable: that is Corollary 4.5.2.)

13.13.4.2 B. Practice

Exercise. B1: Which are orthogonal projections?

Determine which of the following matrices are projections, and among those, which are orthogonal projections. Justify your answer. For the oblique ones over $\mathbb{R}$, find the angle between image and kernel.

(a) $\mathbf{A} = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ on $\mathbb{R}^2$.

(b) $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ -2 & 0 \end{pmatrix}$ on $\mathbb{R}^2$.

(c) $\mathbf{C} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ on $\mathbb{R}^2$.

(d) $\mathbf{D} = \begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix}$ on $\mathbb{C}^2$.

Solution.

- (a) $\mathbf{A}^2 = \frac{1}{25} \begin{pmatrix} 5 & 10 \\ 10 & 20 \end{pmatrix} = \mathbf{A}$, so $\mathbf{A}$ is a projection, and $\mathbf{A}^\top = \mathbf{A}$. By [Theorem 13.13.1](#) it is the orthogonal projection onto $\text{im } \mathbf{A} = \text{span}((1, 2))$.
- (b) $\mathbf{B}^2 = \mathbf{B}$, so $\mathbf{B}$ is a projection, but $\mathbf{B}^\top \neq \mathbf{B}$, so it is oblique. Its image is $\text{span}((1, -2))$ and its kernel is $\text{span}(\mathbf{e}_2)$. The angle between these lines has $\cos \theta = |\langle (1, -2), (0, 1) \rangle| / (\sqrt{5} \cdot 1) = 2/\sqrt{5}$, so $\sin \theta = 1/\sqrt{5}$ and $\cot \theta = 2$; that is $\theta = \arctan(1/2)$, about 26.6° .
- (c) $\mathbf{C}^2 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = 2\mathbf{C} \neq \mathbf{C}$, so $\mathbf{C}$ is not a projection at all, although it is symmetric. Symmetry alone is not idempotence; $\frac{1}{2}\mathbf{C}$ is the orthogonal projection onto $\text{span}((1, 1))$.
- (d) $\mathbf{D}^2 = \mathbf{D}$, so $\mathbf{D}$ is a projection, and $\mathbf{D}^* = \begin{pmatrix} 1 & 0 \\ -i & 0 \end{pmatrix} \neq \mathbf{D}$, so it is oblique. It is already in the canonical form of [Theorem 13.13.3](#), with $x = i$. There is no angle over $\mathbb{C}$ (Definition 11.1.5 is stated for real spaces), but $|x| = 1$, and the largest value of $\|\mathbf{D}\mathbf{v}\|$ over unit $\mathbf{v}$ is $\sqrt{1+1} = \sqrt{2}$, the same number the real matrix $\mathbf{P}'$ produced.

Exercise. B2: Putting a projection in canonical form

Let $\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ act on $\mathbb{R}^3$ with the dot product.

- (a) Verify that $\mathbf{A}$ is a projection and find $\text{im } \mathbf{A}$ and $\text{ker } \mathbf{A}$.
- (b) Find an orthonormal basis $\mathcal{B}$ of $\mathbb{R}^3$ as in [Theorem 13.13.3](#) and write down $[T_{\mathbf{A}}]_{\mathcal{B}}^{\mathcal{B}}$.
- (c) Hence determine whether $\mathbf{A}$ is an orthogonal projection, and find $\text{rank } \mathbf{A}$ without row reducing.

Solution.

- (a) $\mathbf{A}^2 = \mathbf{A}$, by multiplying out: the second column of $\mathbf{A}^2$ is $\mathbf{A}(1, 1, 0) = (1, 1, 0)$, and

the first and third columns are $\mathbf{0}$ and $\mathbf{e}_3$, unchanged. The columns of $\mathbf{A}$ span $\text{im } \mathbf{A} = \text{span}((1, 1, 0), \mathbf{e}_3)$, which is 2-dimensional. And $\mathbf{A}\mathbf{x} = (x_2, x_2, x_3)$ is $\mathbf{0}$ exactly when $x_2 = x_3 = 0$, so $\ker \mathbf{A} = \text{span}(\mathbf{e}_1)$.

- (b) $\text{im } \mathbf{A}$ has the orthonormal basis $\mathbf{f}_1 = \frac{1}{\sqrt{2}}(1, 1, 0)$, $\mathbf{f}_2 = \mathbf{e}_3$, and $(\text{im } \mathbf{A})^\perp = \text{span}((1, -1, 0))$ has the orthonormal basis $\mathbf{f}_3 = \frac{1}{\sqrt{2}}(1, -1, 0)$. Now $\mathbf{A}\mathbf{f}_1 = \mathbf{f}_1$ and $\mathbf{A}\mathbf{f}_2 = \mathbf{f}_2$, while

$$\mathbf{A}\mathbf{f}_3 = \frac{1}{\sqrt{2}}(-1, -1, 0) = -\mathbf{f}_1.$$

Hence

$$[T_{\mathbf{A}}]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

- (c) $\mathbf{X} \neq \mathbf{0}$, so $\mathbf{A}$ is oblique by [Theorem 13.13.3](#); indeed $\ker \mathbf{A} = \text{span}(\mathbf{e}_1)$ is not $(\text{im } \mathbf{A})^\perp = \text{span}((1, -1, 0))$. The rank is the number of ones in the canonical form, namely 2; equivalently $\text{rank } \mathbf{A} = \text{tr } \mathbf{A} = 0 + 1 + 1 = 2$ by [Theorem 4.5.4](#).

Exercise. B3: The complementary projection

Let $P \in \mathcal{L}(V)$ be a projection on a finite-dimensional inner product space, with $U = \text{im } P$.

- (a) Prove that P is an orthogonal projection if and only if $\text{id}_V - P$ is.
- (b) Prove that in that case $\text{id}_V - P = P_{U^\perp}$.

Hint: Exercise (B2: The complementary projection) identifies the image and the kernel of $\text{id}_V - P$.

Solution.

- (a) By Exercise (B2: The complementary projection) (a), $\text{id}_V - P$ is a projection. By [Theorem 11.6.2](#) (a), (b) and (d), $(\text{id}_V - P)^* = \text{id}_V - P^*$. Hence $(\text{id}_V - P)^* = \text{id}_V - P$ if and only if $P^* = P$, and [Theorem 13.13.1](#) converts each side into the corresponding statement about orthogonal projections.
- (b) Suppose P is an orthogonal projection. By Exercise (B2: The complementary projection) (b), $\text{im}(\text{id}_V - P) = \ker P$, which is $U^\perp$ by [Theorem 13.13.1](#) (b). By (a), $\text{id}_V - P$ is an orthogonal projection, so by clause (a) of that theorem it is the orthogonal projection onto its own image, that is $P_{U^\perp}$.

13.13.4.3 C. Going deeper

Exercise. C1: Products of orthogonal projections

Let P and Q be orthogonal projections on a finite-dimensional inner product space V , with images U and W .

- (a) Prove that PQ is an orthogonal projection if and only if $PQ = QP$.
- (b) Suppose $PQ = QP$. Prove that $PQ = P_{U \cap W}$.
- (c) Give an explicit example of two orthogonal projections on $\mathbb{R}^2$ whose product is not a

projection.

Solution.

- (a) ($\Leftarrow$) Suppose $PQ = QP$. Then $(PQ)^2 = P(QP)Q = P(PQ)Q = P^2Q^2 = PQ$, so PQ is a projection, and $(PQ)^* = Q^*P^* = QP = PQ$ by Theorem 11.6.2 (c) and self-adjointness of P and Q . By Theorem 13.13.1, PQ is an orthogonal projection.
 ($\Rightarrow$) Suppose PQ is an orthogonal projection. Then $PQ = (PQ)^* = Q^*P^* = QP$.
- (b) Assume $PQ = QP$, so PQ is an orthogonal projection by (a), and by clause (a) of Theorem 13.13.1 it equals $P_{\text{im}(PQ)}$. It remains to show $\text{im}(PQ) = U \cap W$. ($\subseteq$) Every value PQv lies in $\text{im } P = U$; and $PQv = QPv$ lies in $\text{im } Q = W$. ($\supseteq$) If $v \in U \cap W$, then $Qv = v$ and $Pv = v$, since a projection fixes its image (Theorem 4.5.1 (a)), so $PQv = v$ and $v \in \text{im}(PQ)$.
- (c) Project onto the x -axis and onto the line $y = x$, with standard matrices

$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{Q} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Both are symmetric and idempotent, hence orthogonal projections. But $\mathbf{PQ} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ and $(\mathbf{PQ})^2 = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \neq \mathbf{PQ}$, so the product is not even a projection. Consistently with (a), $\mathbf{QP} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \neq \mathbf{PQ}$.

Exercise. C2: The Frobenius norm detects obliqueness

Let $\mathbf{P} \in M_n(F)$ be idempotent, of rank r .

- (a) Prove that $\|\mathbf{P}\|_F^2 \geq r$.
 (b) Prove that equality holds if and only if $\mathbf{P}$ is an orthogonal projection.

Hint: the canonical form of Theorem 13.13.3, together with the invariance of the Frobenius norm under a unitary change of orthonormal basis.

Solution. Apply Theorem 13.13.3 to the operator $\mathbf{x} \mapsto \mathbf{P}\mathbf{x}$ on F^n , and let $\mathbf{U}$ be the matrix whose columns are the vectors of the orthonormal basis it produces. Then $\mathbf{U}$ is unitary (Theorem 11.7.3 (f)), and the change-of-basis formula gives

$$\mathbf{U}^*\mathbf{P}\mathbf{U} = \mathbf{U}^{-1}\mathbf{P}\mathbf{U} = \begin{pmatrix} \mathbf{I}_r & \mathbf{X} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}.$$

By Lemma 11.7.5, $\|\mathbf{P}\|_F = \|\mathbf{U}^*\mathbf{P}\mathbf{U}\|_F$.

- (a) Summing the squared moduli of the entries of the block matrix,

$$\|\mathbf{P}\|_F^2 = \|\mathbf{I}_r\|_F^2 + \|\mathbf{X}\|_F^2 = r + \|\mathbf{X}\|_F^2 \geq r,$$

since $\|\mathbf{X}\|_F^2 \geq 0$.

- (b) Equality holds if and only if $\|\mathbf{X}\|_F = 0$, that is $\mathbf{X} = \mathbf{O}$, which by Theorem 13.13.3 is exactly the condition that $\mathbf{P}$ be an orthogonal projection. (Note that $r = \text{rank } \mathbf{P} = \text{tr } \mathbf{P}$ by Theorem 4.5.4, so the criterion may also be written $\text{tr}(\mathbf{P}^*\mathbf{P}) = \text{tr } \mathbf{P}$.)

Exercise. C3: How far an oblique projection can stretch

Let P be a projection of rank r on an n -dimensional inner product space V over F , with $0 < r < n$, and let $\mathbf{X}$ be the block of [Theorem 13.13.3](#), with largest singular value $\sigma_1 = \sigma_1(\mathbf{X})$.

- (a) Prove that $\|\mathbf{X}\mathbf{z}\| \leq \sigma_1\|\mathbf{z}\|$ for every $\mathbf{z} \in F^{n-r}$, with equality for some $\mathbf{z} \neq \mathbf{0}$.
- (b) Prove that the largest value of $\|Pv\|$ over unit vectors $v \in V$ is $\sqrt{1 + \sigma_1^2}$.
- (c) Deduce the two-dimensional formula $\sqrt{1 + |x|^2}$ of the text, and state what the answer becomes when P is an orthogonal projection.

Solution.

- (a) Write $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ as in [Theorem 13.9.1](#), with $\mathbf{U}, \mathbf{V}$ unitary, and let p be the length of the list of singular values. For $\mathbf{z} \in F^{n-r}$ put $\mathbf{w} = \mathbf{V}^*\mathbf{z}$, so that $\|\mathbf{w}\| = \|\mathbf{z}\|$ and $\|\mathbf{X}\mathbf{z}\| = \|\mathbf{\Sigma}\mathbf{w}\|$, a unitary matrix preserving norms ([Theorem 11.7.3](#)). The vector $\mathbf{\Sigma}\mathbf{w}$ has entries $\sigma_i w_i$ for $i \leq p$ and 0 beyond, so

$$\|\mathbf{\Sigma}\mathbf{w}\|^2 = \sum_{i \leq p} \sigma_i^2 |w_i|^2 \leq \sigma_1^2 \sum_{i \leq p} |w_i|^2 \leq \sigma_1^2 \|\mathbf{w}\|^2 = \sigma_1^2 \|\mathbf{z}\|^2,$$

since σ_1 is the largest of the σ_i . Taking $\mathbf{z} = \mathbf{V}\mathbf{e}_1$, a unit vector, gives $\mathbf{w} = \mathbf{e}_1$ and $\|\mathbf{X}\mathbf{z}\| = \sigma_1$, so the bound is attained.

- (b) Use the orthonormal basis $\mathcal{B}$ of [Theorem 13.13.3](#) and write a vector of V by its coordinate vector $(\mathbf{y}, \mathbf{z})$ with $\mathbf{y} \in F^r$ and $\mathbf{z} \in F^{n-r}$; since $\mathcal{B}$ is orthonormal, [Theorem 11.2.1](#) (c) gives $\|v\|^2 = \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2$, and Pv has coordinates $(\mathbf{y} + \mathbf{X}\mathbf{z}, \mathbf{0})$. By [Corollary 11.1.6](#), then (a), then [Theorem 11.1.5](#) in $\mathbb{R}^2$ applied to $(\|\mathbf{y}\|, \|\mathbf{z}\|)$ and $(1, \sigma_1)$,

$$\begin{aligned} \|Pv\| &= \|\mathbf{y} + \mathbf{X}\mathbf{z}\| \leq \|\mathbf{y}\| + \sigma_1\|\mathbf{z}\| \\ &\leq \sqrt{1 + \sigma_1^2} \sqrt{\|\mathbf{y}\|^2 + \|\mathbf{z}\|^2} = \sqrt{1 + \sigma_1^2} \|v\|. \end{aligned}$$

For equality, let $\mathbf{z}_0$ be the unit vector of (a), so $\mathbf{X}\mathbf{z}_0 = \sigma_1\mathbf{u}_0$ with $\|\mathbf{u}_0\| = 1$ (take $\mathbf{u}_0 = \mathbf{X}\mathbf{z}_0/\sigma_1$ if $\sigma_1 > 0$; if $\sigma_1 = 0$ take any unit $\mathbf{u}_0$). Put

$$\mathbf{y} = \frac{\mathbf{u}_0}{\sqrt{1 + \sigma_1^2}}, \quad \mathbf{z} = \frac{\sigma_1\mathbf{z}_0}{\sqrt{1 + \sigma_1^2}}.$$

Then $\|\mathbf{y}\|^2 + \|\mathbf{z}\|^2 = 1$ and $\mathbf{y} + \mathbf{X}\mathbf{z} = (1 + \sigma_1^2)\mathbf{u}_0/\sqrt{1 + \sigma_1^2}$, whose norm is $\sqrt{1 + \sigma_1^2}$.

- (c) When $n = 2$ and $r = 1$, $\mathbf{X}$ is the 1×1 matrix (x) , whose only singular value is $|x|$; the formula becomes $\sqrt{1 + |x|^2}$. When P is an orthogonal projection, $\mathbf{X} = \mathbf{O}$, so $\sigma_1 = 0$ and the largest value is 1 — an orthogonal projection never lengthens a vector, which is clause (e) of [Theorem 13.13.1](#) with the sharpest constant.