

20.8 Principal Angles

The sections so far measured how far eigenvalues move. The next section asks how far **eigenvectors** move, and more generally invariant subspaces, and before it can answer it needs a way to say how far apart two subspaces are. For two lines there is an obvious answer, the angle between them. This section shows that for subspaces of dimension $k \geq 2$ one angle is not enough, and that the right replacement is a list of k angles read off from a singular value decomposition. It then proves the two facts Section 9 will use: the sines of these angles are the singular values of a matrix built from the orthogonal complement, and the largest sine is the distance between the two orthogonal projections.

Throughout, F is $\mathbb{R}$ or $\mathbb{C}$, and F^n carries its standard inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x}$, with $\mathbf{x}^* = \mathbf{x}^\top$ when $F = \mathbb{R}$. For a subspace U of F^n , P_U is the orthogonal projection onto U (Definition 11.3.2), identified with its matrix in the standard basis. Singular values of a matrix with k columns follow the convention of Chapter 17 §09: $\sigma_i := 0$ for $\min(\text{rows}, k) < i \leq k$, so that σ_i^2 is the i -th largest eigenvalue of $\mathbf{A}^* \mathbf{A}$ for every $i \leq k$ (Eq. 17.9.1).

20.8.1 One angle is not enough

For two lines $\text{span}(\mathbf{u})$ and $\text{span}(\mathbf{w})$ there is one number to compute, the angle between them. For planes the first idea is to copy it: call the angle between U and W the **smallest** angle between a line in U and a line in W .

That idea fails at once in $\mathbb{R}^3$. Two distinct planes $U, W \subseteq \mathbb{R}^3$ always meet in a line, since

$$\dim(U \cap W) = 2 + 2 - \dim(U + W) \geq 4 - 3 = 1$$

by Theorem 1.7.3. So the smallest angle is 0 for **every** pair of planes: it cannot tell a plane tilted by one degree from a plane at right angles.

Taking the **largest** such angle instead does no better in $\mathbb{R}^4$. Let $U = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ and compare

$$W_1 = \text{span}(\mathbf{e}_1, \mathbf{e}_3), \quad W_2 = \text{span}(\mathbf{e}_3, \mathbf{e}_4).$$

Each contains a vector perpendicular to all of U (namely $\mathbf{e}_3$), so each has a line at angle $\pi/2$ from U . But W_1 shares the whole line $\text{span}(\mathbf{e}_1)$ with U , while W_2 is perpendicular to U in every direction. One number cannot record both facts. A list can: $(0, \pi/2)$ for W_1 and $(\pi/2, \pi/2)$ for W_2 . The rest of this section explains where such a list comes from in general, when the subspaces are not lined up with the coordinate axes.

Two subspaces can be given orthonormal bases that face each other in pairs; the principal angles are the angles between the partners.

20.8.2 Principal angles

Everything will be computed from orthonormal bases written as the columns of a matrix, so we first record how such matrices behave.

Let $U \subseteq F^n$ be a subspace with $\dim U = k \geq 1$. An **orthonormal basis matrix** of U is a matrix $\mathbf{Q} \in M_{n \times k}(F)$ whose columns form an orthonormal basis of U . One exists: apply Gram-Schmidt (Theorem 11.2.2) to any basis of U .

Lemma 20.8.1. Orthonormal Basis Matrices

Let $U \subseteq F^n$ have dimension $k \geq 1$ and let $\mathbf{Q} \in M_{n \times k}(F)$ be an orthonormal basis matrix of U .

- (a) $\mathbf{Q}^* \mathbf{Q} = \mathbf{I}_k$ and $P_U = \mathbf{Q} \mathbf{Q}^*$.
 (b) $\|\mathbf{Q} \mathbf{y}\| = \|\mathbf{y}\|$ for every $\mathbf{y} \in F^k$. For every matrix $\mathbf{Y}$ with k rows and every matrix $\mathbf{Z}$ with k columns,

$$\begin{aligned} \|\mathbf{Q} \mathbf{Y}\|_2 &= \|\mathbf{Y}\|_2, & \|\mathbf{Q} \mathbf{Y}\|_F &= \|\mathbf{Y}\|_F, \\ \|\mathbf{Z} \mathbf{Q}^*\|_2 &= \|\mathbf{Z}\|_2, & \|\mathbf{Z} \mathbf{Q}^*\|_F &= \|\mathbf{Z}\|_F. \end{aligned}$$

- (c) Every orthonormal basis matrix of U is $\mathbf{Q} \mathbf{R}$ for some $\mathbf{R} \in U(k)$ (orthogonal when $F = \mathbb{R}$).
 (d) $\mathbf{I}_n - P_U = P_{U^\perp}$. If moreover $k < n$ and $\mathbf{Q}_\perp \in M_{n \times (n-k)}(F)$ is an orthonormal basis matrix of $U^\perp$, then the square matrix $(\mathbf{Q} \mid \mathbf{Q}_\perp)$ is unitary, and

$$\mathbf{Q} \mathbf{Q}^* + \mathbf{Q}_\perp \mathbf{Q}_\perp^* = \mathbf{I}_n.$$

Proof.

- (a) The (i, j) entry of $\mathbf{Q}^* \mathbf{Q}$ is $\mathbf{q}_i^* \mathbf{q}_j = \langle \mathbf{q}_j, \mathbf{q}_i \rangle = \delta_{ij}$, because the columns $\mathbf{q}_1, \dots, \mathbf{q}_k$ are orthonormal. By Theorem 11.3.3 (a), for every $\mathbf{v} \in F^n$,

$$P_U \mathbf{v} = \sum_{i=1}^k \langle \mathbf{v}, \mathbf{q}_i \rangle \mathbf{q}_i = \sum_{i=1}^k \mathbf{q}_i (\mathbf{q}_i^* \mathbf{v}) = \mathbf{Q} \mathbf{Q}^* \mathbf{v},$$

the last step being column-times-row expansion of $\mathbf{Q} \mathbf{Q}^*$.

- (b) By (a), $\|\mathbf{Q} \mathbf{y}\|^2 = \mathbf{y}^* \mathbf{Q}^* \mathbf{Q} \mathbf{y} = \mathbf{y}^* \mathbf{y} = \|\mathbf{y}\|^2$. Hence $\|\mathbf{Q} \mathbf{Y} \mathbf{x}\| = \|\mathbf{Y} \mathbf{x}\|$ for every $\mathbf{x}$, and taking the maximum over unit $\mathbf{x}$ (Definition 16.3.2) gives $\|\mathbf{Q} \mathbf{Y}\|_2 = \|\mathbf{Y}\|_2$. For the Frobenius norm, $\|\mathbf{Q} \mathbf{Y}\|_F^2 = \text{tr}(\mathbf{Y}^* \mathbf{Q}^* \mathbf{Q} \mathbf{Y}) = \text{tr}(\mathbf{Y}^* \mathbf{Y}) = \|\mathbf{Y}\|_F^2$. Next, $\|\mathbf{Z} \mathbf{Q}^*\|_F^2 = \text{tr}(\mathbf{Q} \mathbf{Z}^* \mathbf{Z} \mathbf{Q}^*) = \text{tr}(\mathbf{Z}^* \mathbf{Z} \mathbf{Q}^* \mathbf{Q}) = \|\mathbf{Z}\|_F^2$, by Theorem 0.9.6 (3). For the spectral norm, let $\mathbf{x} \in F^n$. By the first sentence and (a), $\|\mathbf{Q}^* \mathbf{x}\| = \|\mathbf{Q} \mathbf{Q}^* \mathbf{x}\| = \|P_U \mathbf{x}\| \leq \|\mathbf{x}\|$ (Theorem 11.3.3 (c)). So $\|\mathbf{Z} \mathbf{Q}^* \mathbf{x}\| \leq \|\mathbf{Z}\|_2 \|\mathbf{Q}^* \mathbf{x}\| \leq \|\mathbf{Z}\|_2 \|\mathbf{x}\|$ by Theorem 16.3.2 (a), and $\|\mathbf{Z} \mathbf{Q}^*\|_2 \leq \|\mathbf{Z}\|_2$. Conversely, choose a unit $\mathbf{y} \in F^k$ with $\|\mathbf{Z} \mathbf{y}\| = \|\mathbf{Z}\|_2$ (Lemma 16.3.1). Then $\mathbf{x} = \mathbf{Q} \mathbf{y}$ is a unit vector with $\mathbf{Z} \mathbf{Q}^* \mathbf{x} = \mathbf{Z} \mathbf{y}$, so $\|\mathbf{Z} \mathbf{Q}^*\|_2 \geq \|\mathbf{Z}\|_2$.
 (c) Let $\mathbf{Q}'$ be another orthonormal basis matrix of U , and put $\mathbf{R} = \mathbf{Q}^* \mathbf{Q}' \in M_k(F)$. Each column of $\mathbf{Q}'$ lies in U , where P_U is the identity, so $\mathbf{Q}' = P_U \mathbf{Q}' = \mathbf{Q} \mathbf{Q}^* \mathbf{Q}' = \mathbf{Q} \mathbf{R}$ by (a). Then $\mathbf{R}^* \mathbf{R} = \mathbf{Q}'^* \mathbf{Q} \mathbf{Q}^* \mathbf{Q}' = \mathbf{Q}'^* P_U \mathbf{Q}' = \mathbf{Q}'^* \mathbf{Q}' = \mathbf{I}_k$, so the square matrix $\mathbf{R}$ is unitary (Definition 11.7.2).
 (d) For the first identity, let $\mathbf{v} \in F^n$ and write $\mathbf{v} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U$ and $\mathbf{w} \in U^\perp$, which is possible in exactly one way by Theorem 11.3.2 (a). Then $P_U \mathbf{v} = \mathbf{u}$ by Definition 11.3.2. Since $(U^\perp)^\perp = U$ by Theorem 11.3.2 (b), the splitting $\mathbf{v} = \mathbf{w} + \mathbf{u}$ is also the one that defines $P_{U^\perp}$, so $P_{U^\perp} \mathbf{v} = \mathbf{w} = \mathbf{v} - P_U \mathbf{v}$. As $\mathbf{v}$ was arbitrary, $\mathbf{I}_n - P_U = P_{U^\perp}$. Nothing here restricts k : when $k = n$, $U^\perp = \{\mathbf{0}\}$ and both sides are $\mathbf{0}$.

Now let $k < n$. The columns of $\mathbf{Q}$ and of $\mathbf{Q}_\perp$ are orthonormal, and each column of $\mathbf{Q}$ is orthogonal to each column of $\mathbf{Q}_\perp$ because $U \perp U^\perp$. Their number is $k + (n - k) = n$, by Theorem 11.3.2 (c). So the square matrix $\mathbf{W} = (\mathbf{Q} \mid \mathbf{Q}_\perp) \in M_n(F)$ satisfies $\mathbf{W}^* \mathbf{W} = \mathbf{I}_n$, hence is unitary, and then $\mathbf{W} \mathbf{W}^* = \mathbf{I}_n$ as well, since $\mathbf{W}^* = \mathbf{W}^{-1}$. By block multiplication (Theorem 8.1.1), $\mathbf{W} \mathbf{W}^* = \mathbf{Q} \mathbf{Q}^* + \mathbf{Q}_\perp \mathbf{Q}_\perp^*$. This also re-proves the first identity here, since

$$\mathbf{Q}_\perp \mathbf{Q}_\perp^* = P_{U^\perp} \text{ by (a) applied to } U^\perp.$$

■

Now the definition. The question is how much of U survives when it is projected into W . By parts (a) and (b) of the lemma, for a unit vector $\mathbf{u} = \mathbf{Q}_U \mathbf{y}$ of U the surviving length is $\|P_W \mathbf{u}\| = \|\mathbf{Q}_W \mathbf{Q}_W^* \mathbf{u}\| = \|\mathbf{Q}_W^* \mathbf{u}\| = \|(\mathbf{Q}_W^* \mathbf{Q}_U) \mathbf{y}\|$. So the matrix $\mathbf{Q}_W^* \mathbf{Q}_U$ records the answer for every direction of U at once, and its singular values are the stretch factors along the best directions.

Definition 20.8.1. Principal Angles

Let $U, W \subseteq F^n$ be subspaces with

$$1 \leq k = \dim U \leq l = \dim W,$$

and let $\mathbf{Q}_U \in M_{n \times k}(F)$ and $\mathbf{Q}_W \in M_{n \times l}(F)$ be orthonormal basis matrices of U and W . The **principal angles** between U and W are the k numbers

$$0 \leq \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k \leq \frac{\pi}{2}$$

determined by

$$\cos \theta_i = \sigma_i(\mathbf{Q}_W^* \mathbf{Q}_U) \quad (i = 1, \dots, k).$$

We write $\Theta(U, W) = \text{diag}(\theta_1, \dots, \theta_k)$ and $\sin \Theta(U, W) = \text{diag}(\sin \theta_1, \dots, \sin \theta_k)$.

In words: fix orthonormal bases of the two subspaces, form the $l \times k$ matrix whose (i, j) entry is the inner product of the j -th basis vector of U with the i -th basis vector of W , and take its k singular values. These lie in $[0, 1]$, as we check next, so each is the cosine of exactly one angle in $[0, \pi/2]$. The largest cosine gives the **smallest** angle, so the decreasing singular values produce increasing angles. The number of angles is the **smaller** dimension, k , and the definition asks for $k \leq l$; for a pair with $\dim U > \dim W$, swap the names. The two notations at the end are for use in norms: $\|\sin \Theta(U, W)\|_2 = \sin \theta_k$, the largest entry of a nonnegative diagonal matrix (Lemma 17.3.5), and $\|\sin \Theta(U, W)\|_F = (\sum_i \sin^2 \theta_i)^{1/2}$.

Well-definedness. Three things need checking. The answer must not depend on the orthonormal bases chosen. The singular values must be at most 1, or there is no angle. And when $k = l$ the definition treats U and W asymmetrically, so we should check that swapping them changes nothing.

Proposition 20.8.2. Principal Angles Are Well Defined

In the setting of Definition 20.8.1, put $\mathbf{M} = \mathbf{Q}_W^* \mathbf{Q}_U \in M_{l \times k}(F)$.

- (a) The singular values of $\mathbf{M}$ do not depend on the choice of the orthonormal basis matrices $\mathbf{Q}_U$ and $\mathbf{Q}_W$.
- (b) $1 \geq \sigma_1(\mathbf{M}) \geq \cdots \geq \sigma_k(\mathbf{M}) \geq 0$, so $\theta_1, \dots, \theta_k$ exist, are unique, and are nondecreasing.
- (c) $\sigma_i(\mathbf{Q}_U^* \mathbf{Q}_W) = \sigma_i(\mathbf{M})$ for $i = 1, \dots, k$. In particular, if $k = l$, the principal angles between W and U are those between U and W .
- (d) If $k = 1$ and $U = \text{span}(\mathbf{u})$ with $\|\mathbf{u}\| = 1$, then $\cos \theta_1 = \|P_W \mathbf{u}\|$. If also $l = 1$ and $W = \text{span}(\mathbf{w})$ with $\|\mathbf{w}\| = 1$, then $\cos \theta_1 = |\langle \mathbf{u}, \mathbf{w} \rangle|$.

Proof.

- (a) By Lemma 20.8.1 (c), any other choices are $\mathbf{Q}_U \mathbf{R}$ and $\mathbf{Q}_W \mathbf{R}'$ with $\mathbf{R} \in U(k)$ and $\mathbf{R}' \in U(l)$. The new matrix is $\mathbf{M}' = \mathbf{R}'^* \mathbf{M} \mathbf{R}$, and

$$\mathbf{M}'^* \mathbf{M}' = \mathbf{R}^* \mathbf{M}^* \mathbf{R}' \mathbf{R}'^* \mathbf{M} \mathbf{R} = \mathbf{R}^* (\mathbf{M}^* \mathbf{M}) \mathbf{R},$$

using $\mathbf{R}' \mathbf{R}'^* = \mathbf{I}_l$. Since $\mathbf{R}^* = \mathbf{R}^{-1}$, this is similar to $\mathbf{M}^* \mathbf{M}$, so the two have the same characteristic polynomial (Theorem 7.10.3) and the same eigenvalues with multiplicity. By Definition 13.9.1, $\mathbf{M}'$ and $\mathbf{M}$ have the same singular values.

- (b) By Theorem 16.3.2 (d) and Lemma 20.8.1 (b), $\|\mathbf{M}\|_2 \leq \|\mathbf{Q}_W^*\|_2 \|\mathbf{Q}_U\|_2$, and $\|\mathbf{Q}_U\|_2 = \|\mathbf{Q}_U \mathbf{I}_k\|_2 = \|\mathbf{I}_k\|_2 = 1$ by Theorem 16.3.2 (c). Also $\|\mathbf{Q}_W^* \mathbf{x}\| = \|P_W \mathbf{x}\| \leq \|\mathbf{x}\|$, as in the proof of that lemma, so $\|\mathbf{Q}_W^*\|_2 \leq 1$. Hence $\sigma_1(\mathbf{M}) = \|\mathbf{M}\|_2 \leq 1$ by Theorem 16.3.3 (c), and all singular values lie in $[0, 1]$. The cosine is a strictly decreasing bijection from $[0, \pi/2]$ onto $[0, 1]$, so each $\sigma_i(\mathbf{M})$ is $\cos \theta_i$ for exactly one $\theta_i \in [0, \pi/2]$, and $\sigma_1 \geq \dots \geq \sigma_k$ gives $\theta_1 \leq \dots \leq \theta_k$.
- (c) Let $\mathbf{M} = \mathbf{Y} \mathbf{\Sigma} \mathbf{Z}^*$ be a singular value decomposition (Theorem 13.9.1). Then $\mathbf{M}^* = \mathbf{Z} \mathbf{\Sigma}^T \mathbf{Y}^*$, where $\mathbf{Z}, \mathbf{Y}$ are unitary and $\mathbf{\Sigma}^T \in M_{k \times l}(\mathbb{R})$ carries the same diagonal entries $\sigma_1(\mathbf{M}) \geq \dots \geq \sigma_k(\mathbf{M})$ and zeros elsewhere. So this is a factorization of $\mathbf{M}^*$ of the shape in Theorem 13.9.1, and Theorem 13.9.2 gives $\sigma_i(\mathbf{M}^*) = \sigma_i(\mathbf{M})$. Since $\mathbf{M}^* = \mathbf{Q}_U^* \mathbf{Q}_W$, this is the first claim. When $k = l$, the principal angles between W and U are defined by $\sigma_i(\mathbf{Q}_U^* \mathbf{Q}_W)$, so they coincide with those between U and W .
- (d) With $\mathbf{Q}_U = \mathbf{u}$, the matrix $\mathbf{M} = \mathbf{Q}_W^* \mathbf{u}$ is a single column, and its one singular value is $(\mathbf{M}^* \mathbf{M})^{1/2} = \|\mathbf{Q}_W^* \mathbf{u}\|$. By Lemma 20.8.1 (b) and (a), $\|\mathbf{Q}_W^* \mathbf{u}\| = \|\mathbf{Q}_W \mathbf{Q}_W^* \mathbf{u}\| = \|P_W \mathbf{u}\|$. If $l = 1$ and $\mathbf{Q}_W = \mathbf{w}$, then $\mathbf{M} = \mathbf{w}^* \mathbf{u} = \langle \mathbf{u}, \mathbf{w} \rangle$, whose singular value is $|\langle \mathbf{u}, \mathbf{w} \rangle|$. ■

Examples.

- Two lines in $\mathbb{R}^2$.** For $\mathbf{u} = \mathbf{e}_1$ and $\mathbf{w} = (1, 1)/\sqrt{2}$, part (d) gives $\cos \theta_1 = 1/\sqrt{2}$, so $\theta_1 = \pi/4$. Replacing $\mathbf{w}$ by $-\mathbf{w}$ changes neither the line nor the answer.
- Nested subspaces.** If $U \subseteq W$, then every column of $\mathbf{Q}_U$ is fixed by P_W , so $\mathbf{M}^* \mathbf{M} = \mathbf{Q}_U^* \mathbf{Q}_W \mathbf{Q}_W^* \mathbf{Q}_U = \mathbf{Q}_U^* P_W \mathbf{Q}_U = \mathbf{Q}_U^* \mathbf{Q}_U = \mathbf{I}_k$. All singular values are 1 and all principal angles are 0. The degenerate case $U = W$ is included, and it is the reason the angles start at 0.
- Perpendicular subspaces.** If $U \subseteq W^\perp$, every entry of $\mathbf{M}$ is the inner product of a vector of U with a vector of W , hence 0, so all principal angles are $\pi/2$.
- The planes of the opening.** For $U = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $W_1 = \text{span}(\mathbf{e}_1, \mathbf{e}_3)$, take the standard vectors as orthonormal bases. Then $\mathbf{M} = \mathbf{Q}_{W_1}^T \mathbf{Q}_U = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, with singular values 1, 0, so the principal angles are $(0, \pi/2)$. For $W_2 = \text{span}(\mathbf{e}_3, \mathbf{e}_4)$ we are in Example 3, with angles $(\pi/2, \pi/2)$. These are the two lists that the single-number attempts could not tell apart.

Nonexample by minimal change. In Example 4, replace the orthonormal basis $(\mathbf{e}_1, \mathbf{e}_3)$ of W_1 by the basis $(\mathbf{e}_1, (\mathbf{e}_1 + \mathbf{e}_3)/\sqrt{2})$, whose vectors still have length 1 but are no longer orthogonal. The matrix of inner products becomes $\begin{pmatrix} 1 & 0 \\ 1/\sqrt{2} & 0 \end{pmatrix}$, with largest singular value $\sqrt{3/2} > 1$, and no angle has that cosine. What fails is the orthogonality of the columns of $\mathbf{Q}_W$: without it, $\mathbf{Q}_W \mathbf{Q}_W^*$ is not the projection P_W , and the matrix no longer measures how much of U survives in W .

Why this definition: the principal vectors. The singular value decomposition of $\mathbf{M}$ does

more than produce numbers. It produces bases of U and W in which the whole configuration is visible, and this is the precise meaning of the slogan.

Proposition 20.8.3. Principal Vectors

In the setting of Definition 20.8.1, let $\mathbf{Q}_W^* \mathbf{Q}_U = \mathbf{Y} \mathbf{\Sigma} \mathbf{Z}^*$ be a singular value decomposition (Theorem 13.9.1), with $\mathbf{Y} \in M_l(F)$ and $\mathbf{Z} \in M_k(F)$ unitary, columns $\mathbf{y}_1, \dots, \mathbf{y}_l$ and $\mathbf{z}_1, \dots, \mathbf{z}_k$. Put

$$\mathbf{u}_i = \mathbf{Q}_U \mathbf{z}_i \quad (1 \leq i \leq k), \quad \mathbf{w}_j = \mathbf{Q}_W \mathbf{y}_j \quad (1 \leq j \leq l).$$

Then $(\mathbf{u}_1, \dots, \mathbf{u}_k)$ is an orthonormal basis of U , $(\mathbf{w}_1, \dots, \mathbf{w}_l)$ is an orthonormal basis of W , and

$$\langle \mathbf{u}_i, \mathbf{w}_j \rangle = \begin{cases} \cos \theta_i & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Proof. The matrix $\mathbf{Q}_U \mathbf{Z}$ has columns $\mathbf{u}_1, \dots, \mathbf{u}_k$, all in U , and $(\mathbf{Q}_U \mathbf{Z})^* (\mathbf{Q}_U \mathbf{Z}) = \mathbf{Z}^* \mathbf{Z} = \mathbf{I}_k$ by Lemma 20.8.1 (a). So the $\mathbf{u}_i$ are k orthonormal vectors of U . They are linearly independent (Theorem 11.1.4), and $\dim U = k$, so they form a basis. The same argument applies to the $\mathbf{w}_j$ in W . Finally,

$$\langle \mathbf{u}_i, \mathbf{w}_j \rangle = \mathbf{w}_j^* \mathbf{u}_i = \mathbf{y}_j^* (\mathbf{Q}_W^* \mathbf{Q}_U) \mathbf{z}_i = \mathbf{y}_j^* \mathbf{Y} \mathbf{\Sigma} \mathbf{Z}^* \mathbf{z}_i = \mathbf{e}_j^* \mathbf{\Sigma} \mathbf{e}_i,$$

which is the (j, i) entry of $\mathbf{\Sigma}$: it equals $\sigma_i = \cos \theta_i$ when $j = i$, and 0 otherwise. ■

So U and W have orthonormal bases that face each other in pairs. Each $\mathbf{u}_i$ makes the angle θ_i with its partner $\mathbf{w}_i$, in the sense of part (d) of the proposition, and is perpendicular to every other $\mathbf{w}_j$. The last $l - k$ vectors of W have no partner, and they are perpendicular to all of U . The vectors $\mathbf{u}_i, \mathbf{w}_i$ are called **principal vectors**. Two subspaces, however they sit in F^n , look like a list of k independent pairs of lines, and the principal angles are the angles within the pairs.

The principal vectors are usually **not** the basis vectors one started with. For $U = W = \mathbb{R}^2$ with bases $(\mathbf{e}_1, \mathbf{e}_2)$ and $((\mathbf{e}_1 + \mathbf{e}_2)/\sqrt{2}, (\mathbf{e}_1 - \mathbf{e}_2)/\sqrt{2})$, each given basis vector of U is at angle $\pi/4$ from each given basis vector of W . The principal angles are nevertheless 0, 0, by Example 2, since the subspaces are equal. Pairing the given columns one by one measures the bases, not the subspaces.

Warning. Lines have no orientation, so principal angles use $|\langle \mathbf{u}, \mathbf{w} \rangle|$ and lie in $[0, \pi/2]$. The angle of Definition 11.1.5 is between **vectors** in a **real** space, and it can reach π . For $\mathbf{u} = \mathbf{e}_1$ and $\mathbf{w} = -\mathbf{e}_1$ in $\mathbb{R}^2$ that angle is π , yet $\text{span}(\mathbf{u}) = \text{span}(\mathbf{w})$ and the principal angle is 0. Over $\mathbb{C}$ the vector angle is not even defined: $\mathbf{u} = \mathbf{e}_1$ and $\mathbf{w} = i\mathbf{e}_1$ span the same complex line, $\langle \mathbf{u}, \mathbf{w} \rangle = i = -i$ is not real, and the principal angle is again 0 because $|-i| = 1$.

Quick check. Find the principal angle between the line $U = \text{span}((1, 2, 2))$ and the plane $W = \{\mathbf{x} \in \mathbb{R}^3 : x_3 = 0\}$.

Solution. Here $k = 1 \leq 2 = l$. The unit vector $\mathbf{u} = \frac{1}{3}(1, 2, 2)$ spans U , and $P_W \mathbf{u} = \frac{1}{3}(1, 2, 0)$, since $(\mathbf{e}_1, \mathbf{e}_2)$ is an orthonormal basis of W (Theorem 11.3.3 (a)). By Proposition 20.8.2 (d), $\cos \theta_1 = \|P_W \mathbf{u}\| = \sqrt{5}/3$. So $\theta_1 = \arccos(\sqrt{5}/3) \approx 41.8^\circ$, and $\sin \theta_1 = 2/3$, which is the distance from $\mathbf{u}$ to the plane.

20.8.3 Sines and the orthogonal complement

The cosines measure how much of U survives projection into W . What Section 9 can actually bound is the part that does **not** survive: the component of U in $W^\perp$. Pythagoras says the two parts are complementary, and the theorem makes this exact.

Theorem 20.8.4. Sines of Principal Angles

Let $U, W \subseteq F^n$ with $1 \leq k = \dim U \leq l = \dim W$, principal angles $\theta_1 \leq \dots \leq \theta_k$, and orthonormal basis matrices $\mathbf{Q}_U, \mathbf{Q}_W$.

- (a) Suppose $l < n$, and let $\mathbf{Q}_{W^\perp} \in M_{n \times (n-l)}(F)$ be an orthonormal basis matrix of $W^\perp$. Put $\mathbf{M} = \mathbf{Q}_W^* \mathbf{Q}_U$ and $\mathbf{N} = \mathbf{Q}_{W^\perp}^* \mathbf{Q}_U \in M_{(n-l) \times k}(F)$. Then

$$\mathbf{M}^* \mathbf{M} + \mathbf{N}^* \mathbf{N} = \mathbf{I}_k,$$

and the eigenvalues of $\mathbf{N}^* \mathbf{N}$, with multiplicity, are $\sin^2 \theta_1, \dots, \sin^2 \theta_k$. Equivalently, $\sigma_i(\mathbf{N}) = \sin \theta_{k+1-i}$ for $i = 1, \dots, k$, with the convention for σ_i stated at the start of the section.

- (b) For every $l \leq n$,

$$\|(\mathbf{I} - P_W)P_U\|_2 = \|(\mathbf{I} - P_W)\mathbf{Q}_U\|_2 = \sin \theta_k = \|\sin \Theta(U, W)\|_2,$$

and

$$\|(\mathbf{I} - P_W)P_U\|_F = \|(\mathbf{I} - P_W)\mathbf{Q}_U\|_F = \|\sin \Theta(U, W)\|_F.$$

Idea. Split the identity of F^n along $W \oplus W^\perp$ and sandwich it between $\mathbf{Q}_U^*$ and $\mathbf{Q}_U$. That gives two Hermitian $k \times k$ matrices adding up to $\mathbf{I}_k$, and such matrices share their eigenvectors, with eigenvalues c and $1 - c$. The first has eigenvalues $\cos^2 \theta_i$ by definition, so the second has eigenvalues $\sin^2 \theta_i$. Part (b) then removes the orthonormal factors, which change no norm.

Proof.

- (a) By Lemma 20.8.1 (d) applied to W , $\mathbf{I}_n = \mathbf{Q}_W \mathbf{Q}_W^* + \mathbf{Q}_{W^\perp} \mathbf{Q}_{W^\perp}^*$. Multiplying on the left by $\mathbf{Q}_U^*$ and on the right by $\mathbf{Q}_U$, and using $\mathbf{Q}_U^* \mathbf{Q}_U = \mathbf{I}_k$ (part (a) of the lemma), gives $\mathbf{I}_k = \mathbf{M}^* \mathbf{M} + \mathbf{N}^* \mathbf{N}$.

The matrix $\mathbf{M}^* \mathbf{M} \in M_k(F)$ is Hermitian (real symmetric when $F = \mathbb{R}$). By Corollary 12.4.3 (or Corollary 12.5.2 when $F = \mathbb{R}$) there is an orthonormal basis $\mathbf{z}_1, \dots, \mathbf{z}_k$ of F^k consisting of eigenvectors of $\mathbf{M}^* \mathbf{M}$. Since $\mathbf{M}$ has $l \geq k$ rows, all k eigenvalues of $\mathbf{M}^* \mathbf{M}$ are squared singular values (Definition 13.9.1), so we may number the $\mathbf{z}_i$ so that $\mathbf{M}^* \mathbf{M} \mathbf{z}_i = \sigma_i(\mathbf{M})^2 \mathbf{z}_i = \cos^2 \theta_i \mathbf{z}_i$. Then

$$\mathbf{N}^* \mathbf{N} \mathbf{z}_i = (\mathbf{I}_k - \mathbf{M}^* \mathbf{M}) \mathbf{z}_i = (1 - \cos^2 \theta_i) \mathbf{z}_i = \sin^2 \theta_i \mathbf{z}_i.$$

So the unitary matrix with columns $\mathbf{z}_1, \dots, \mathbf{z}_k$ diagonalizes $\mathbf{N}^*\mathbf{N}$ with diagonal $\sin^2 \theta_1, \dots, \sin^2 \theta_k$, and these are its eigenvalues with multiplicity (Theorem 7.10.3). The sine is increasing on $[0, \pi/2]$, so in decreasing order they read $\sin^2 \theta_k \geq \dots \geq \sin^2 \theta_1 \geq 0$. By Eq. 17.9.1, the i -th largest eigenvalue of $\mathbf{N}^*\mathbf{N}$ is $\sigma_i(\mathbf{N})^2$ for every $i \leq k$, with the stated convention. Taking nonnegative square roots gives $\sigma_i(\mathbf{N}) = \sin \theta_{k+1-i}$.

- (b) If $l = n$, then $W = F^n$, so $U \subseteq W$ and every $\theta_i = 0$ (Example 2 above), while $\mathbf{I} - P_W = \mathbf{O}$; all four norms are 0. Suppose $l < n$. By Lemma 20.8.1 (d) and (a),

$$(\mathbf{I} - P_W)\mathbf{Q}_U = \mathbf{Q}_{W^\perp}\mathbf{Q}_{W^\perp}^*\mathbf{Q}_U = \mathbf{Q}_{W^\perp}\mathbf{N},$$

and $(\mathbf{I} - P_W)P_U = \mathbf{Q}_{W^\perp}\mathbf{N}\mathbf{Q}_U^*$. By part (b) of the lemma, applied to $\mathbf{Q}_{W^\perp}$ and then to $\mathbf{Q}_U$, both matrices have the spectral and Frobenius norms of $\mathbf{N}$. Now $\|\mathbf{N}\|_2 = \sigma_1(\mathbf{N}) = \sin \theta_k$ by Theorem 16.3.3 (c) and part (a). Also $\|\mathbf{N}\|_F^2 = \text{tr}(\mathbf{N}^*\mathbf{N})$, which is the sum of the eigenvalues of $\mathbf{N}^*\mathbf{N}$ (Theorem 9.3.3), that is $\sum_i \sin^2 \theta_i = \|\sin \Theta(U, W)\|_F^2$. This proves the theorem. ■

Part (b) has a plain geometric reading. For a unit vector $\mathbf{u} \in U$, the vector $(\mathbf{I} - P_W)\mathbf{u}$ is the error in approximating $\mathbf{u}$ from W , and its length is the distance from $\mathbf{u}$ to W (Theorem 11.3.5). So $\sin \theta_k$ is the **largest distance from a unit vector of U to W** , and θ_k is the worst angle any direction of U makes with W . That is the number to bound when one asks whether a computed subspace is close to the true one, and part (a) turns it into the norm of the concrete matrix $\mathbf{Q}_{W^\perp}^*\mathbf{Q}_U$. In Section 9, U will be an invariant subspace of $\mathbf{A}$ and W the corresponding one for $\mathbf{A} + \mathbf{E}$. The columns of $\mathbf{Q}_{W^\perp}$ will be the remaining eigenvectors of $\mathbf{A} + \mathbf{E}$, and $\mathbf{N}$ will turn out to satisfy a Sylvester equation, so the bounds of Section 7 will apply to it.

20.8.4 The distance between two subspaces

A subspace is determined by its orthogonal projection, since $U = \text{im } P_U$ by Theorem 4.5.1 (b), applied to $F^n = U \oplus U^\perp$. So a second natural way to compare U and W is the size of $P_U - P_W$. When the dimensions agree, this measures exactly the largest principal angle.

Theorem 20.8.5. The Distance Between Projections

Let $U, W \subseteq F^n$ with $\dim U = \dim W = k \geq 1$, and principal angles $\theta_1 \leq \dots \leq \theta_k$. Then

- (a) $\|P_U - P_W\|_2 = \sin \theta_k$;
- (b) $\|P_U - P_W\|_F = \sqrt{2} \|\sin \Theta(U, W)\|_F$;
- (c) $\|P_U - P_W\|_2 \leq 1$, with equality if and only if $U \cap W^\perp \neq \{\mathbf{0}\}$.

Idea. For (b), the **squared** Frobenius norm of the Hermitian matrix $\mathbf{D} = P_U - P_W$ is $\text{tr}(\mathbf{D}^2)$, which expands into traces of projections: $k + k - 2 \text{tr}(P_U P_W)$, and $\text{tr}(P_U P_W) = \sum \cos^2 \theta_i$. For (a), split $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ along $U \oplus U^\perp$. Then $\mathbf{D}$ sends $\mathbf{x}_1$ to $(\mathbf{I} - P_W)\mathbf{x}_1 \in W^\perp$ and $\mathbf{x}_2$ to $-P_W\mathbf{x}_2 \in W$, so Pythagoras separates the two contributions. The first is controlled by the sines for (U, W) , the second by the sines for (W, U) , and equal dimensions make these the same list.

Proof. Put $\mathbf{D} = P_U - P_W$, and let $\mathbf{Q}_U, \mathbf{Q}_W$ be orthonormal basis matrices and $\mathbf{M} = \mathbf{Q}_W^* \mathbf{Q}_U \in M_k(F)$.

- (b) By [Lemma 20.8.1](#) (a), $P_U = \mathbf{Q}_U \mathbf{Q}_U^*$ and $P_W = \mathbf{Q}_W \mathbf{Q}_W^*$ are Hermitian and satisfy $P_U^2 = P_U$, $P_W^2 = P_W$. Hence $\mathbf{D}$ is Hermitian and

$$\begin{aligned} \|\mathbf{D}\|_F^2 &= \text{tr}(\mathbf{D}^* \mathbf{D}) = \text{tr}(\mathbf{D}^2) \\ &= \text{tr}(P_U) + \text{tr}(P_W) - \text{tr}(P_U P_W) - \text{tr}(P_W P_U). \end{aligned}$$

By Theorem 0.9.6 (3), $\text{tr}(P_U) = \text{tr}(\mathbf{Q}_U^* \mathbf{Q}_U) = \text{tr}(\mathbf{I}_k) = k$, likewise $\text{tr}(P_W) = k$, and $\text{tr}(P_W P_U) = \text{tr}(P_U P_W)$. Also

$$\begin{aligned} \text{tr}(P_U P_W) &= \text{tr}(\mathbf{Q}_U \mathbf{Q}_U^* \mathbf{Q}_W \mathbf{Q}_W^*) \\ &= \text{tr}(\mathbf{Q}_W^* \mathbf{Q}_U \mathbf{Q}_U^* \mathbf{Q}_W) = \text{tr}(\mathbf{M} \mathbf{M}^*), \end{aligned}$$

and $\text{tr}(\mathbf{M} \mathbf{M}^*) = \|\mathbf{M}^*\|_F^2 = \|\mathbf{M}\|_F^2 = \text{tr}(\mathbf{M}^* \mathbf{M}) = \sum_i \cos^2 \theta_i$, since the eigenvalues of $\mathbf{M}^* \mathbf{M}$ are $\cos^2 \theta_1, \dots, \cos^2 \theta_k$ (Definition 13.9.1, Theorem 9.3.3). Therefore

$$\|\mathbf{D}\|_F^2 = 2k - 2 \sum_{i=1}^k \cos^2 \theta_i = 2 \sum_{i=1}^k \sin^2 \theta_i,$$

which is (b).

- (c) Put $s = \sin \theta_k$. By [Theorem 20.8.4](#) (b), $\|(\mathbf{I} - P_W)P_U\|_2 = s$. The same theorem with the roles of U and W exchanged, which is allowed because $\dim W = \dim U$, gives $\|(\mathbf{I} - P_U)P_W\|_2 = s$ too, because by [Proposition 20.8.2](#) (c) the principal angles between W and U are those between U and W .

Let $\mathbf{x} \in F^n$, and write $\mathbf{x}_1 = P_U \mathbf{x} \in U$ and $\mathbf{x}_2 = \mathbf{x} - \mathbf{x}_1 \in U^\perp$ (Definition 11.3.2). Then $P_U \mathbf{x} = \mathbf{x}_1$ and $P_W \mathbf{x} = P_W \mathbf{x}_1 + P_W \mathbf{x}_2$, so

$$\mathbf{D} \mathbf{x} = (\mathbf{I} - P_W) \mathbf{x}_1 - P_W \mathbf{x}_2.$$

The first term lies in $W^\perp$ and the second in W , by Theorem 11.3.2, so they are orthogonal, and by Theorem 11.1.3

$$\|\mathbf{D} \mathbf{x}\|^2 = \|(\mathbf{I} - P_W) \mathbf{x}_1\|^2 + \|P_W \mathbf{x}_2\|^2. \quad (20.8.1)$$

We bound the two terms. Since $\mathbf{x}_1 = P_U \mathbf{x}_1$, $\|(\mathbf{I} - P_W) \mathbf{x}_1\| = \|(\mathbf{I} - P_W) P_U \mathbf{x}_1\| \leq s \|\mathbf{x}_1\|$ by Theorem 16.3.2 (a). For the second term, P_W is Hermitian and idempotent and $\mathbf{x}_2 = (\mathbf{I} - P_U) \mathbf{x}_2$, so

$$\begin{aligned} \|P_W \mathbf{x}_2\|^2 &= \langle P_W \mathbf{x}_2, \mathbf{x}_2 \rangle = \langle P_W \mathbf{x}_2, (\mathbf{I} - P_U) \mathbf{x}_2 \rangle \\ &= \langle (\mathbf{I} - P_U) P_W (P_W \mathbf{x}_2), \mathbf{x}_2 \rangle \leq s \|P_W \mathbf{x}_2\| \|\mathbf{x}_2\|, \end{aligned}$$

using Theorem 11.3.3 (b) for $\mathbf{I} - P_U = P_{U^\perp}$ ([Lemma 20.8.1](#) (d)), then $P_W = P_W^2$, then Cauchy-Schwarz (Theorem 11.1.5) with $\|(\mathbf{I} - P_U) P_W\|_2 = s$. Hence $\|P_W \mathbf{x}_2\| \leq s \|\mathbf{x}_2\|$, the case $P_W \mathbf{x}_2 = \mathbf{0}$ being trivial. Adding, and using $\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 = \|\mathbf{x}\|^2$ (Theorem 11.1.3),

$$\|\mathbf{D} \mathbf{x}\|^2 \leq s^2 (\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2) = s^2 \|\mathbf{x}\|^2.$$

So $\|\mathbf{D}\|_2 \leq s$. For the reverse, [Theorem 20.8.4](#) (b) gives $\|(\mathbf{I} - P_W)\mathbf{Q}_U\|_2 = s$, and by [Lemma 16.3.1](#) there is a unit $\mathbf{y} \in F^k$ with $\|(\mathbf{I} - P_W)\mathbf{Q}_U\mathbf{y}\| = s$. The vector $\mathbf{x} = \mathbf{Q}_U\mathbf{y} \in U$ is a unit vector ([Lemma 20.8.1](#) (b)) with $P_U\mathbf{x} = \mathbf{x}$, so $\mathbf{D}\mathbf{x} = (\mathbf{I} - P_W)\mathbf{x}$ and $\|\mathbf{D}\mathbf{x}\| = s$. Hence $\|\mathbf{D}\|_2 = s$.

- (c) By (a) and [Proposition 20.8.2](#) (b), $\|\mathbf{D}\|_2 = \sin \theta_k \leq 1$, with equality if and only if $\cos \theta_k = \sigma_k(\mathbf{M}) = 0$. The square matrix $\mathbf{M}$ has $\sigma_k(\mathbf{M}) = 0$ if and only if $\text{rank } \mathbf{M} < k$ ([Theorem 13.9.1](#)), that is, if and only if $\mathbf{M}\mathbf{y} = \mathbf{0}$ for some $\mathbf{y} \neq \mathbf{0}$. Put $\mathbf{u} = \mathbf{Q}_U\mathbf{y}$, which is a nonzero vector of U exactly when $\mathbf{y} \neq \mathbf{0}$, by [Lemma 20.8.1](#) (b). Then $\mathbf{M}\mathbf{y} = \mathbf{Q}_W^*\mathbf{u}$, and $\mathbf{Q}_W^*\mathbf{u} = \mathbf{0}$ if and only if $P_W\mathbf{u} = \mathbf{Q}_W\mathbf{Q}_W^*\mathbf{u} = \mathbf{0}$, again by parts (b) and (a) of the lemma. Finally $P_W\mathbf{u} = \mathbf{0}$ means that $\mathbf{u} \in \ker P_W = W^\perp$, by [Theorem 4.5.1](#) (b) applied to $F^n = W \oplus W^\perp$. So equality holds if and only if U contains a nonzero vector of $W^\perp$. ■

The factor $\sqrt{2}$ in (b) has a clean explanation. When $k = l$, the principal vectors of [Proposition 20.8.3](#) give $P_U = \sum_i \mathbf{u}_i \mathbf{u}_i^*$ and $P_W = \sum_i \mathbf{w}_i \mathbf{w}_i^*$ ([Lemma 20.8.1](#) (a)), so $P_U - P_W$ is a sum of the k terms $\mathbf{u}_i \mathbf{u}_i^* - \mathbf{w}_i \mathbf{w}_i^*$. These act in the mutually orthogonal planes $\text{span}(\mathbf{u}_i, \mathbf{w}_i)$. Each term is Hermitian with trace 0 and rank at most 2, so its nonzero eigenvalues are a pair $\pm t_i$, and being Hermitian it is normal, so $\|\mathbf{u}_i \mathbf{u}_i^* - \mathbf{w}_i \mathbf{w}_i^*\|_F^2 = t_i^2 + t_i^2$ by [Theorem 12.11.1](#). On the other hand, expanding the trace with $\mathbf{u}_i^* \mathbf{u}_i = \mathbf{w}_i^* \mathbf{w}_i = 1$ and $|\langle \mathbf{u}_i, \mathbf{w}_i \rangle| = \cos \theta_i$ ([Proposition 20.8.3](#)),

$$\|\mathbf{u}_i \mathbf{u}_i^* - \mathbf{w}_i \mathbf{w}_i^*\|_F^2 = 2 - 2|\langle \mathbf{u}_i, \mathbf{w}_i \rangle|^2 = 2 \sin^2 \theta_i,$$

so $t_i = \sin \theta_i$. Each angle is therefore counted twice in the Frobenius norm and once in the spectral norm. The worked example below shows the $\pm$ pairs explicitly.

Example. Two Planes in Four Dimensions

Let $U = \text{span}(\mathbf{e}_1, \mathbf{e}_2) \subseteq \mathbb{R}^4$ and $W = \text{span}(\mathbf{a}, \mathbf{b})$, where

$$\mathbf{a} = (3, 4, 4, 3), \quad \mathbf{b} = (3, -4, 4, -3).$$

Find the principal angles, a pair of principal vectors for each, the sines from [Theorem 20.8.4](#), and $\|P_U - P_W\|$ in both norms. Then check the last two numbers directly.

Solution. Orthonormal bases. $\langle \mathbf{a}, \mathbf{b} \rangle = 9 - 16 + 16 - 9 = 0$ and $\|\mathbf{a}\|^2 = \|\mathbf{b}\|^2 = 50$, so $\mathbf{Q}_W = \frac{1}{5\sqrt{2}}(\mathbf{a} \mid \mathbf{b})$ is an orthonormal basis matrix of W , and $\mathbf{Q}_U = (\mathbf{e}_1 \mid \mathbf{e}_2)$. The matrix $\mathbf{Q}_W^\top \mathbf{Q}_U$ holds the first two entries of $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbf{M} = \mathbf{Q}_W^\top \mathbf{Q}_U = \frac{1}{5\sqrt{2}} \begin{pmatrix} 3 & 4 \\ 3 & -4 \end{pmatrix}, \quad \mathbf{M}^\top \mathbf{M} = \begin{pmatrix} 9/25 & 0 \\ 0 & 16/25 \end{pmatrix}.$$

So the singular values are $4/5$ and $3/5$, and

$$\theta_1 = \arccos \frac{4}{5} \approx 36.87^\circ, \quad \theta_2 = \arccos \frac{3}{5} \approx 53.13^\circ.$$

Note $\sin \theta_1 = 3/5 = \cos \theta_2$, so $\theta_1 + \theta_2 = \pi/2$ here.

Principal vectors. The eigenvector of $\mathbf{M}^\top \mathbf{M}$ for $16/25$ is $\mathbf{z}_1 = \mathbf{e}_2$, and $\mathbf{y}_1 = \mathbf{M}\mathbf{z}_1/\sigma_1 =$

$\frac{1}{5\sqrt{2}}(4, -4) \cdot \frac{5}{4} = \frac{1}{\sqrt{2}}(1, -1)$. By [Proposition 20.8.3](#),

$$\mathbf{u}_1 = \mathbf{e}_2, \quad \mathbf{w}_1 = \mathbf{Q}_W \mathbf{y}_1 = \frac{1}{10}(\mathbf{a} - \mathbf{b}) = (0, \frac{4}{5}, 0, \frac{3}{5}),$$

and $\langle \mathbf{u}_1, \mathbf{w}_1 \rangle = 4/5$. Likewise $\mathbf{z}_2 = \mathbf{e}_1$, $\mathbf{y}_2 = \frac{1}{\sqrt{2}}(1, 1)$, $\mathbf{u}_2 = \mathbf{e}_1$, and $\mathbf{w}_2 = \frac{1}{10}(\mathbf{a} + \mathbf{b}) = (\frac{3}{5}, 0, \frac{4}{5}, 0)$, with $\langle \mathbf{u}_2, \mathbf{w}_2 \rangle = 3/5$, while $\langle \mathbf{u}_1, \mathbf{w}_2 \rangle = \langle \mathbf{u}_2, \mathbf{w}_1 \rangle = 0$. So W is obtained from U by tilting $\mathbf{e}_2$ towards $\mathbf{e}_4$ through θ_1 and $\mathbf{e}_1$ towards $\mathbf{e}_3$ through θ_2 , in two perpendicular planes.

Sines. The vectors $\mathbf{c} = (4, -3, -3, 4)$ and $\mathbf{d} = (4, 3, -3, -4)$ are orthogonal to $\mathbf{a}$, to $\mathbf{b}$, and to each other:

$$\begin{aligned} \langle \mathbf{c}, \mathbf{a} \rangle &= 12 - 12 - 12 + 12 = 0, & \langle \mathbf{c}, \mathbf{b} \rangle &= 12 + 12 - 12 - 12 = 0, \\ \langle \mathbf{d}, \mathbf{a} \rangle &= 12 + 12 - 12 - 12 = 0, & \langle \mathbf{d}, \mathbf{b} \rangle &= 12 - 12 - 12 + 12 = 0, \end{aligned}$$

and $\langle \mathbf{c}, \mathbf{d} \rangle = 16 - 9 + 9 - 16 = 0$. Also $\|\mathbf{c}\|^2 = \|\mathbf{d}\|^2 = 50$. Since $\dim W^\perp = 2$ (Theorem 11.3.2 (c)), $\mathbf{Q}_{W^\perp} = \frac{1}{5\sqrt{2}}(\mathbf{c} \mid \mathbf{d})$ is an orthonormal basis matrix of $W^\perp$, and

$$\mathbf{N} = \mathbf{Q}_{W^\perp}^\top \mathbf{Q}_U = \frac{1}{5\sqrt{2}} \begin{pmatrix} 4 & -3 \\ 4 & 3 \end{pmatrix}, \quad \mathbf{N}^\top \mathbf{N} = \begin{pmatrix} 16/25 & 0 \\ 0 & 9/25 \end{pmatrix}.$$

Its eigenvalues are $16/25 = \sin^2 \theta_2$ and $9/25 = \sin^2 \theta_1$, and $\mathbf{M}^\top \mathbf{M} + \mathbf{N}^\top \mathbf{N} = \mathbf{I}_2$, as [Theorem 20.8.4](#) (a) says.

Distances. By [Theorem 20.8.5](#), $\|P_U - P_W\|_2 = \sin \theta_2 = 4/5$ and $\|P_U - P_W\|_F = \sqrt{2}(\frac{9}{25} + \frac{16}{25})^{1/2} = \sqrt{2}$.

Direct check. By [Lemma 20.8.1](#) (a), $P_W = \mathbf{Q}_W \mathbf{Q}_W^\top = \frac{1}{50}(\mathbf{a}\mathbf{a}^\top + \mathbf{b}\mathbf{b}^\top)$, and

$$P_W = \frac{1}{50} \begin{pmatrix} 18 & 0 & 24 & 0 \\ 0 & 32 & 0 & 24 \\ 24 & 0 & 32 & 0 \\ 0 & 24 & 0 & 18 \end{pmatrix}, \quad P_U = \text{diag}(1, 1, 0, 0).$$

In $\mathbf{D} = P_U - P_W$, coordinates 1, 3 do not interact with coordinates 2, 4, and the two 2×2 blocks are

$$\frac{1}{50} \begin{pmatrix} 32 & -24 \\ -24 & -32 \end{pmatrix}, \quad \frac{1}{50} \begin{pmatrix} 18 & -24 \\ -24 & -18 \end{pmatrix}.$$

Each is symmetric with trace 0, so its eigenvalues are $\pm\sqrt{-\det}$: $\pm\frac{1}{50}\sqrt{32^2 + 24^2} = \pm\frac{4}{5}$ and $\pm\frac{1}{50}\sqrt{18^2 + 24^2} = \pm\frac{3}{5}$. So $\mathbf{D}$ has eigenvalues $\pm\frac{4}{5}, \pm\frac{3}{5}$. By [Lemma 17.3.5](#), $\|\mathbf{D}\|_2 = 4/5$, and $\|\mathbf{D}\|_F^2 = \text{tr}(\mathbf{D}^2) = 2(\frac{16}{25} + \frac{9}{25}) = 2$. Both agree with the theorem, and the eigenvalues come in the $\pm \sin \theta_i$ pairs described before the example.

Remark. A metric, and why the dimensions must match. On the set of k -dimensional subspaces of F^n , the number $d(U, W) = \|P_U - P_W\|_2 = \sin \theta_k$ is a metric. It is symmetric. It is zero only when $P_U = P_W$, that is, when $U = \text{im } P_U = \text{im } P_W = W$. And it satisfies the triangle inequality because $\|\cdot\|_2$ is a norm (Theorem 16.3.2 (b)). Between subspaces of **different** dimensions it is useless. If $k < l$, then $\dim(W \cap U^\perp) \geq l + (n - k) - n > 0$ by Theorem 1.7.3, so some unit $\mathbf{w} \in W$ is perpendicular to U . For it $(P_U - P_W)\mathbf{w} = -\mathbf{w}$, so $\|P_U - P_W\|_2 \geq 1$. On the other hand, [Eq. 20.8.1](#), which did not use the dimensions, together with $\|P\mathbf{v}\| \leq \|\mathbf{v}\|$ for orthogonal projections (Theorem 11.3.3 (c)), gives $\|P_U - P_W\|_2 \leq 1$.

So the distance is exactly 1, whatever the principal angles are. This is why Section 9 always compares invariant subspaces of the same dimension.

20.8.5 Exercises

20.8.5.1 A. Check your understanding

Exercise. A1

- (a) Define the principal angles between subspaces $U, W \subseteq F^n$ with $1 \leq \dim U \leq \dim W$.
- (b) True or false: if $U \subseteq W$, then every principal angle between U and W is 0. Justify your answer.
- (c) True or false: the principal angle between $\text{span}(\mathbf{u})$ and $\text{span}(-\mathbf{u})$, for $\mathbf{u} \neq \mathbf{0}$, is π . Justify your answer.
- (d) True or false: if $U \cap W \neq \{\mathbf{0}\}$, then $\theta_1 = 0$. Justify your answer.
- (e) State the formula for $\|P_U - P_W\|_2$ in terms of principal angles, with its hypothesis.
- (f) True or false: for two distinct lines U, W in $\mathbb{R}^2$, $\|P_U - P_W\|_2 < 1$. Justify your answer.

Solution.

- (a) Choose orthonormal basis matrices $\mathbf{Q}_U \in M_{n \times k}(F)$ and $\mathbf{Q}_W \in M_{n \times l}(F)$. The principal angles $0 \leq \theta_1 \leq \dots \leq \theta_k \leq \pi/2$ are defined by $\cos \theta_i = \sigma_i(\mathbf{Q}_W^* \mathbf{Q}_U)$, $i = 1, \dots, k$ (Definition 20.8.1). They do not depend on the bases (Proposition 20.8.2).
- (b) True. As in Example 2 after Proposition 20.8.2, $\mathbf{M}^* \mathbf{M} = \mathbf{Q}_U^* P_W \mathbf{Q}_U = \mathbf{Q}_U^* \mathbf{Q}_U = \mathbf{I}_k$, so every $\cos \theta_i = 1$.
- (c) False. The two lines are equal, so the angle is 0: with $\mathbf{u}_0 = \mathbf{u}/\|\mathbf{u}\|$, $|\langle \mathbf{u}_0, -\mathbf{u}_0 \rangle| = 1$ (Proposition 20.8.2 (d)). The value π is the angle between the vectors $\mathbf{u}$ and $-\mathbf{u}$ of Definition 11.1.5, which is a different notion.
- (d) True. Take a unit $\mathbf{u} \in U \cap W$ and write $\mathbf{u} = \mathbf{Q}_U \mathbf{y}$ with $\|\mathbf{y}\| = 1$ (Lemma 20.8.1 (b)). Then $\|\mathbf{M} \mathbf{y}\| = \|\mathbf{Q}_W^* \mathbf{u}\| = \|P_W \mathbf{u}\| = \|\mathbf{u}\| = 1$, so $\sigma_1(\mathbf{M}) = \|\mathbf{M}\|_2 \geq 1$. With Proposition 20.8.2 (b), $\cos \theta_1 = 1$.
- (e) If $\dim U = \dim W = k \geq 1$, then $\|P_U - P_W\|_2 = \sin \theta_k$, the sine of the largest principal angle (Theorem 20.8.5 (a)).
- (f) False. The two axes $U = \text{span}(\mathbf{e}_1)$ and $W = \text{span}(\mathbf{e}_2)$ have $U \cap W^\perp = U \neq \{\mathbf{0}\}$, so $\|P_U - P_W\|_2 = 1$ by Theorem 20.8.5 (c). Directly, $P_U - P_W = \text{diag}(1, -1)$.

20.8.5.2 B. Practice

Exercise. B1: Two planes in three dimensions

Let $U = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $W = \text{span}((1, 0, 0), (1, 3, 4))$ in $\mathbb{R}^3$.

- (a) Find an orthonormal basis matrix of W .
- (b) Compute the principal angles between U and W .

- (c) Hence compute $\|P_U - P_W\|_2$ and $\|P_U - P_W\|_F$.
 (d) Explain why $\theta_1 = 0$ could have been predicted without computation.

Solution.

- (a) Gram–Schmidt: $\mathbf{q}_1 = \mathbf{e}_1$, and $(1, 3, 4) - \langle (1, 3, 4), \mathbf{e}_1 \rangle \mathbf{e}_1 = (0, 3, 4)$, of length 5, so $\mathbf{q}_2 = (0, \frac{3}{5}, \frac{4}{5})$ and $\mathbf{Q}_W = (\mathbf{q}_1 \mid \mathbf{q}_2)$.
 (b) With $\mathbf{Q}_U = (\mathbf{e}_1 \mid \mathbf{e}_2)$, the matrix $\mathbf{M} = \mathbf{Q}_W^\top \mathbf{Q}_U$ holds the first two entries of $\mathbf{q}_1$ and $\mathbf{q}_2$:

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ 0 & 3/5 \end{pmatrix}.$$

It is diagonal with nonnegative entries, so $\mathbf{M}^\top \mathbf{M} = \text{diag}(1, \frac{9}{25})$ and the singular values are 1 and $\frac{3}{5}$. So $\theta_1 = 0$ and $\theta_2 = \arccos \frac{3}{5} \approx 53.13^\circ$, with $\sin \theta_2 = \frac{4}{5}$.

- (c) By [Theorem 20.8.5](#), $\|P_U - P_W\|_2 = \sin \theta_2 = \frac{4}{5}$ and $\|P_U - P_W\|_F = \sqrt{2} (0 + \frac{16}{25})^{1/2} = \frac{4\sqrt{2}}{5}$.
 (d) Two planes in $\mathbb{R}^3$ meet in a line (Theorem 1.7.3, as in the opening of the section), and a common unit vector forces $\cos \theta_1 = 1$ (Exercise A1 (d)). Here the common line is $\text{span}(\mathbf{e}_1)$.

Exercise. B2: Which projections are at distance one?

Determine, for each pair, whether $\|P_U - P_W\|_2 = 1$. Justify your answer.

- (a) $U = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $W = \text{span}(\mathbf{e}_2, \mathbf{e}_3)$ in $\mathbb{R}^3$.
 (b) $U = \text{span}((1, 1, 0))$ and $W = \text{span}((1, -1, 1))$ in $\mathbb{R}^3$.
 (c) $U = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $W = \text{span}((1, 0, 1), (0, 1, 1))$ in $\mathbb{R}^3$.
 (d) $U = \text{span}((1, 2, 3))$ and $W = \text{span}((1, 0, 0), (0, 1, 0))$ in $\mathbb{R}^3$.

Solution.

- (a) Yes. $W^\perp = \text{span}(\mathbf{e}_1) \subseteq U$, so $U \cap W^\perp \neq \{\mathbf{0}\}$, and [Theorem 20.8.5](#) (c) applies since both dimensions are 2.
 (b) Yes. $\langle (1, 1, 0), (1, -1, 1) \rangle = 0$, so $U \subseteq W^\perp$, and (c) of the theorem applies with $k = 1$.
 (c) No. Both are planes, and $W^\perp = \text{span}((1, 1, -1))$, since $(1, 1, -1)$ is orthogonal to both spanning vectors of W and $\dim W^\perp = 1$. This vector has nonzero third entry, so it is not in U , and $U \cap W^\perp = \{\mathbf{0}\}$. By the theorem, $\|P_U - P_W\|_2 < 1$. (In fact the principal angles are 0 and $\arccos(1/\sqrt{3})$, so the distance is $\sqrt{2/3}$.)
 (d) Yes, for a different reason: the dimensions are 1 and 2, and the remark after [Example \(Two Planes in Four Dimensions\)](#) shows that $\|P_U - P_W\|_2 = 1$ whenever the dimensions differ. The theorem itself does not apply. Its principal angle $\theta_1 = \arccos(\sqrt{5}/\sqrt{14})$ is far from $\pi/2$, which illustrates why the projection distance is only used for equal dimensions.

Exercise. B3: Two lines

Let $\mathbf{u}, \mathbf{w} \in F^n$ be unit vectors and $c = |\langle \mathbf{u}, \mathbf{w} \rangle|$. Prove that

$$\|\mathbf{u}\mathbf{u}^* - \mathbf{w}\mathbf{w}^*\|_2 = \sqrt{1 - c^2}, \quad \|\mathbf{u}\mathbf{u}^* - \mathbf{w}\mathbf{w}^*\|_F = \sqrt{2} \sqrt{1 - c^2}.$$

Solution. Let $U = \text{span}(\mathbf{u})$ and $W = \text{span}(\mathbf{w})$, both of dimension 1. By Lemma 20.8.1 (a) with $\mathbf{Q}_U = \mathbf{u}$ and $\mathbf{Q}_W = \mathbf{w}$, $P_U = \mathbf{u}\mathbf{u}^*$ and $P_W = \mathbf{w}\mathbf{w}^*$. By Proposition 20.8.2 (d), the single principal angle satisfies $\cos \theta_1 = c$, so $\sin \theta_1 = \sqrt{1 - c^2}$, since $\sin \theta_1 \geq 0$ on $[0, \pi/2]$. Now Theorem 20.8.5 (a) and (b) with $k = 1$ give $\|P_U - P_W\|_2 = \sin \theta_1$ and $\|P_U - P_W\|_F = \sqrt{2} \sin \theta_1$, which are the two claims.

20.8.5.3 C. Going deeper

Exercise. C1: Counting zero and right angles

Let $U, W \subseteq F^n$ with $1 \leq k = \dim U \leq l = \dim W$, orthonormal basis matrices $\mathbf{Q}_U, \mathbf{Q}_W$, and $\mathbf{M} = \mathbf{Q}_W^* \mathbf{Q}_U$.

- (a) Prove that $\mathbf{y}^*(\mathbf{I}_k - \mathbf{M}^* \mathbf{M})\mathbf{y} = \|(\mathbf{I} - P_W)\mathbf{Q}_U \mathbf{y}\|^2$ for every $\mathbf{y} \in F^k$.
- (b) Deduce that the number of principal angles equal to 0 is $\dim(U \cap W)$.
- (c) Prove that the number of principal angles equal to $\pi/2$ is $\dim(U \cap W^\perp)$.

Hint for (b): a positive semidefinite matrix $\mathbf{B}$ with $\mathbf{y}^* \mathbf{B} \mathbf{y} = 0$ has $\mathbf{B} \mathbf{y} = \mathbf{0}$.

Solution.

- (a) Let $\mathbf{u} = \mathbf{Q}_U \mathbf{y}$. By Lemma 20.8.1 (a) and (b), $\mathbf{y}^* \mathbf{y} = \|\mathbf{u}\|^2$ and $\mathbf{y}^* \mathbf{M}^* \mathbf{M} \mathbf{y} = \|\mathbf{Q}_W^* \mathbf{u}\|^2 = \|P_W \mathbf{u}\|^2$. Since $\mathbf{u} = P_W \mathbf{u} + (\mathbf{I} - P_W)\mathbf{u}$ with the two parts orthogonal (Theorem 11.3.2), Theorem 11.1.3 gives $\|\mathbf{u}\|^2 - \|P_W \mathbf{u}\|^2 = \|(\mathbf{I} - P_W)\mathbf{u}\|^2$.
- (b) Put $\mathbf{B} = \mathbf{I}_k - \mathbf{M}^* \mathbf{M}$. It is Hermitian, and by (a) it is positive semidefinite (Definition 13.1.1). Its eigenvalues are $1 - \cos^2 \theta_i = \sin^2 \theta_i$, with an orthonormal eigenbasis $\mathbf{z}_1, \dots, \mathbf{z}_k$ (as in the proof of Theorem 20.8.4). So the number of i with $\theta_i = 0$ is the number of zero eigenvalues, which is $\dim \ker \mathbf{B}$.
We claim $\ker \mathbf{B} = \{\mathbf{y} : \mathbf{Q}_U \mathbf{y} \in W\}$. If $\mathbf{B} \mathbf{y} = \mathbf{0}$, then $\mathbf{y}^* \mathbf{B} \mathbf{y} = 0$, so $(\mathbf{I} - P_W)\mathbf{Q}_U \mathbf{y} = \mathbf{0}$ by (a), and $\mathbf{Q}_U \mathbf{y} = P_W \mathbf{Q}_U \mathbf{y} \in W$. Conversely, if $\mathbf{Q}_U \mathbf{y} \in W$, then $\mathbf{y}^* \mathbf{B} \mathbf{y} = 0$ by (a). Writing $\mathbf{y} = \sum_i c_i \mathbf{z}_i$, this says $\sum_i \sin^2 \theta_i |c_i|^2 = 0$. So $c_i = 0$ whenever $\sin \theta_i \neq 0$, and then $\mathbf{B} \mathbf{y} = \sum_i c_i \sin^2 \theta_i \mathbf{z}_i = \mathbf{0}$. This proves the claim; the second half of it is the fact in the hint, proved here for this $\mathbf{B}$.
The map $\mathbf{y} \mapsto \mathbf{Q}_U \mathbf{y}$ is linear and injective (Lemma 20.8.1 (b)). It carries $\ker \mathbf{B}$ onto $\{\mathbf{u} \in U : \mathbf{u} \in W\} = U \cap W$, because every vector of U is $\mathbf{Q}_U \mathbf{y}$ for some $\mathbf{y}$. So $\dim \ker \mathbf{B} = \dim(U \cap W)$.
- (c) $\theta_i = \pi/2$ exactly when $\sigma_i(\mathbf{M}) = 0$. The number of nonzero singular values of $\mathbf{M}$ is $\text{rank } \mathbf{M}$ (Theorem 13.9.1), so the number of zero ones among the k is $k - \text{rank } \mathbf{M} = \dim \ker \mathbf{M}$, by the Rank–Nullity Theorem (Theorem 2.3.1) for $\mathbf{y} \mapsto \mathbf{M} \mathbf{y}$ on F^k . Next, $\mathbf{M} \mathbf{y} = \mathbf{Q}_W^* \mathbf{Q}_U \mathbf{y} = \mathbf{0}$ if and only if $P_W \mathbf{Q}_U \mathbf{y} = \mathbf{Q}_W (\mathbf{Q}_W^* \mathbf{Q}_U \mathbf{y}) = \mathbf{0}$, because $\mathbf{Q}_W$ is injective. That holds if and only if $\mathbf{Q}_U \mathbf{y} \in W^\perp$. As in (b), $\mathbf{y} \mapsto \mathbf{Q}_U \mathbf{y}$ carries $\ker \mathbf{M}$ injectively onto $U \cap W^\perp$, so the count is $\dim(U \cap W^\perp)$.

Exercise. C2: The smallest angle as a maximum

Let $U, W \subseteq F^n$ with $1 \leq k = \dim U \leq l = \dim W$ and principal angles $\theta_1 \leq \dots \leq \theta_k$.

(a) Prove that

$$\cos \theta_1 = \max\{|\langle \mathbf{u}, \mathbf{w} \rangle| : \mathbf{u} \in U, \mathbf{w} \in W, \|\mathbf{u}\| = \|\mathbf{w}\| = 1\},$$

and that the maximum is attained at the principal vectors $\mathbf{u}_1, \mathbf{w}_1$.

- (b) Deduce that θ_1 is the smallest principal angle between a line of U and a line of W , which is the notion tried and rejected at the start of the section.
- (c) Hence explain, in one sentence, why θ_1 alone cannot distinguish the pairs (U, W_1) and (U, U) of the opening, but the full list θ_1, θ_2 can.

Solution.

- (a) Let $\mathbf{u} \in U$ and $\mathbf{w} \in W$ be unit vectors. By [Lemma 20.8.1](#) (a) and (b), $\mathbf{u} = \mathbf{Q}_U \mathbf{y}$ and $\mathbf{w} = \mathbf{Q}_W \mathbf{z}$ with $\mathbf{y} = \mathbf{Q}_U^* \mathbf{u}$ and $\mathbf{z} = \mathbf{Q}_W^* \mathbf{w}$ unit vectors. Then $\langle \mathbf{u}, \mathbf{w} \rangle = \mathbf{z}^* \mathbf{Q}_W^* \mathbf{Q}_U \mathbf{y} = \mathbf{z}^* \mathbf{M} \mathbf{y}$, and by Cauchy–Schwarz and Theorem 16.3.2 (a),

$$|\langle \mathbf{u}, \mathbf{w} \rangle| \leq \|\mathbf{z}\| \|\mathbf{M} \mathbf{y}\| \leq \|\mathbf{M}\|_2 = \sigma_1(\mathbf{M}) = \cos \theta_1,$$

using Theorem 16.3.3 (c). By [Proposition 20.8.3](#), the unit vectors $\mathbf{u}_1 \in U$ and $\mathbf{w}_1 \in W$ satisfy $\langle \mathbf{u}_1, \mathbf{w}_1 \rangle = \cos \theta_1$, so the maximum exists and equals $\cos \theta_1$.

- (b) By [Proposition 20.8.2](#) (d), the principal angle between the lines $\text{span}(\mathbf{u})$ and $\text{span}(\mathbf{w})$ is $\arccos|\langle \mathbf{u}, \mathbf{w} \rangle|$. Since $\arccos$ is decreasing, the smallest such angle corresponds to the largest $|\langle \mathbf{u}, \mathbf{w} \rangle|$, which by (a) is $\cos \theta_1$.
- (c) Both pairs share a line, so both have $\theta_1 = 0$, and the first number cannot separate them. The second angle is $\pi/2$ for (U, W_1) and 0 for (U, U) , by Examples 4 and 2 after [Proposition 20.8.2](#).