UNIT 6

ORTHOGONALITY

DEFINITION Two subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ of a vector space are ***orthogonal*** if every vector $\boldsymbol{v}$ in $\boldsymbol{V}$ is perpendicular to every vector $\boldsymbol{w}$ in $\boldsymbol{W}$:

$$\boldsymbol{v} \cdot \boldsymbol{w} = 0 \quad \textit{or} \quad \boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = 0 \quad \textit{for all } \boldsymbol{v} \textit{ in } \mathbf{V} \textit{ and all } \boldsymbol{w} \textit{ in } \mathbf{W}.$$

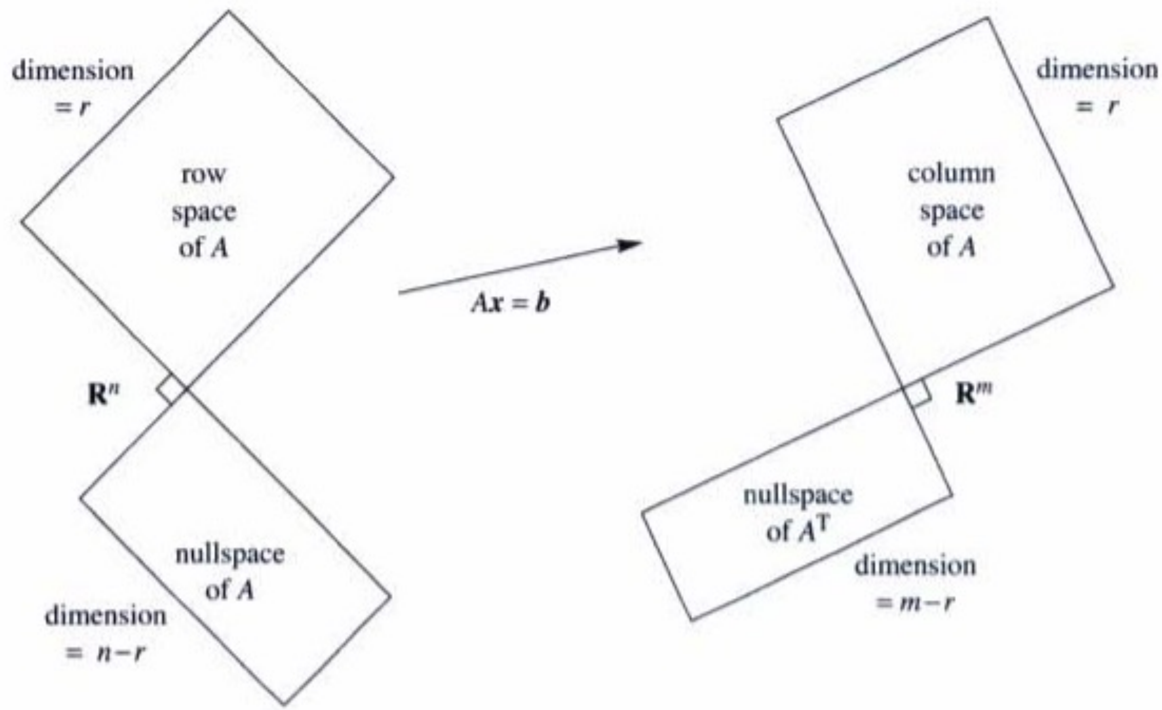

Figure 4.1 Two pairs of orthogonal subspaces. Dimensions add to n and add to m.

Example 1 The floor of your room (extended to infinity) is a subspace $\boldsymbol{V}$. The line where two walls meet is a subspace $\boldsymbol{W}$ (one-dimensional). Those subspaces are orthogonal. Every vector up the meeting line is perpendicular to every vector in the floor. The origin $(0, 0, 0)$ is in the corner. We assume you don't live in a tent.

Example 2 Suppose $\boldsymbol{V}$ is still the floor but $\boldsymbol{W}$ is a wall (a two-dimensional space). The wall and floor look like orthogonal subspaces but they are not! You can find vectors in $\boldsymbol{V}$ and $\boldsymbol{W}$ that are not perpendicular. In fact a vector running along the bottom of the wall is also in the floor. This vector is in both $\boldsymbol{V}$ and $\boldsymbol{W}$—and it is not perpendicular to itself.

When a vector is in two orthogonal subspaces, it *must* be zero. It is perpendicular to itself. It is $\boldsymbol{v}$ and it is $\boldsymbol{w}$, so $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = 0$. This has to be the zero vector.

Note 1

4A Every vector x in the nullspace of A is perpendicular to every row of A, because $Ax = \mathbf{0}$. ***The nullspace and row space are orthogonal subspaces.***

Example 3 The rows of A are perpendicular to $x = (1, 1, -1)$ in the nullspace:

$$Ax = \begin{bmatrix} 1 & 3 & 4 \\ 5 & 2 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{gives the dot products} \quad \begin{matrix} 1 + 3 - 4 = 0 \\ 5 + 2 - 7 = 0 \end{matrix}$$

Now we turn to the other two subspaces. In this example, the column space is all of $\mathbf{R}^2$. The nullspace of A^{T} is only the zero vector. Those two subspaces are also orthogonal.

Note 2

4B Every vector y in the nullspace of A^T is perpendicular to every column of A. *The left nullspace and the column space are orthogonal in* $\mathbf{R}^m$.

$$A^T y = \begin{bmatrix} (\text{column } 1)^T \\ \cdots \\ (\text{column } n)^T \end{bmatrix} \begin{bmatrix} \\ y \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \cdots \\ 0 \end{bmatrix}. \tag{3}$$

The dot product of y with every column of A is zero. Then y in the left nullspace is perpendicular to each column—and to the whole column space.

DEFINITION The ***orthogonal complement*** of V contains *every* vector that is perpendicular to V. This orthogonal subspace is denoted by $V^{\perp}$ (pronounced "V perp").

The nullspace is the orthogonal complement of the row space (in $\mathbf{R}^n$*).*
The left nullspace is the orthogonal complement of the column space (in $\mathbf{R}^m$*).*

Projections

When $\boldsymbol{b}$ is projected onto a line, ***its projection p is the part of b along that line***. If $\boldsymbol{b}$ is projected onto a plane, $\boldsymbol{p}$ is the part in that plane. ***The projection p is*** $P\boldsymbol{b}$**.** There is a projection matrix P that multiplies $\boldsymbol{b}$ to give $\boldsymbol{p}$.

- Example-Let b=(2,3,4)
- The projection of b onto z axis we call p_1.
- The projection of b onto xy plane we call p_2 .

The projection matrices P_1 and P_2 are 3 by 3. They multiply $\boldsymbol{b}$ with 3 components to produce $\boldsymbol{p}$ with 3 components. Projection onto a line comes from a rank one matrix. Projection onto a plane comes from a rank two matrix:

$$\text{Onto the } z \text{ axis:} \quad P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{Onto the } xy \text{ plane:} \quad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

P_1 picks out the z component of every vector. P_2 picks out the x and y components. To find $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, multiply $\boldsymbol{b}$ by P_1 and P_2 (small $\boldsymbol{p}$ for the vector, capital P for the matrix that produces it):

$$\boldsymbol{p}_1 = P_1\boldsymbol{b} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix} \qquad \boldsymbol{p}_2 = P_2\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}.$$

In this case the projections P_1 and P_2 are perpendicular. The xy plane and the z axis are ***orthogonal subspaces***, like the floor of a room and the line between two walls.

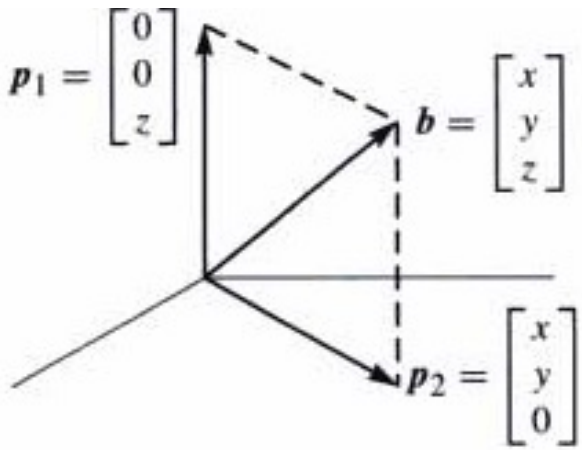

Figure 4.3 The projections of b onto the z axis and the xy plane.

There fore projection of b onto z axis gives $p_1 = (0,0,4)$

Projection of b onto xy plane gives p_2=(2,3,0)

Projection onto line

4E The projection of $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ is the vector $\boldsymbol{p} = \widehat{x}\boldsymbol{a} = \dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a}.$

Special case 1: If $\boldsymbol{b} = \boldsymbol{a}$ then $\widehat{x} = 1$. The projection of $\boldsymbol{a}$ onto $\boldsymbol{a}$ is itself.

Special case 2: If $\boldsymbol{b}$ is perpendicular to $\boldsymbol{a}$ then $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = 0$. The projection is $\boldsymbol{p} = \boldsymbol{0}$.

Now comes the ***projection matrix***. In the formula for $\boldsymbol{p}$, what matrix is multiplying $\boldsymbol{b}$? You can see it better if the number $\widehat{x}$ is on the right side of $\boldsymbol{a}$:

$$p = a\widehat{x} = a\frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a} = Pb \quad \textit{when the matrix is} \quad P = \frac{aa^{\mathrm{T}}}{a^{\mathrm{T}}a}.$$

Note

$$\boldsymbol{p} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a} \quad \text{so its length is} \quad \|\boldsymbol{p}\| = \frac{\|\boldsymbol{a}\|\,\|\boldsymbol{b}\|\cos\theta}{\|\boldsymbol{a}\|^2}\|\boldsymbol{a}\| = \|\boldsymbol{b}\|\cos\theta.$$

We are given a line through the origin, in the direction of $\boldsymbol{a} = (a_1, \ldots, a_m)$. Along that line, we want the point $\boldsymbol{p}$ closest to $\boldsymbol{b} = (b_1, \ldots, b_m)$. The key to projection is orthogonality: ***The line from b to p is perpendicular to the vector a***. This is the dotted line marked $\boldsymbol{e}$ in Figure 4.4—which we now compute by algebra.

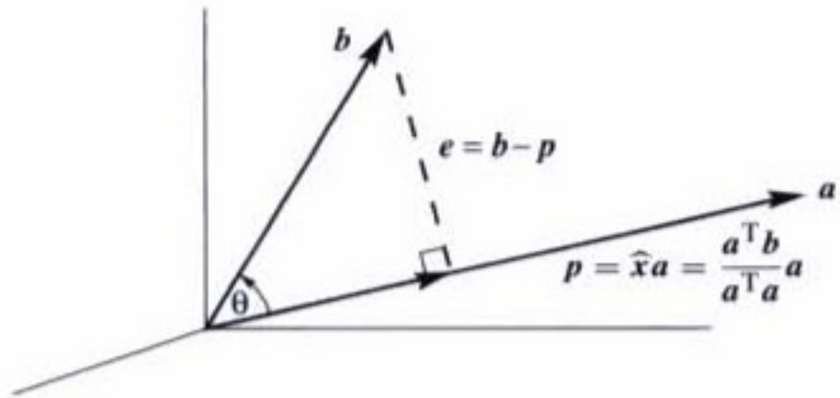

Figure 4.4 The projection $\boldsymbol{p}$, perpendicular to $\boldsymbol{e}$, has length $\|\boldsymbol{b}\| \cos\theta$.

The projection $\boldsymbol{p}$ is some multiple of $\boldsymbol{a}$. Call it $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$ = "x hat" times $\boldsymbol{a}$. Our first step is to compute this unknown number $\widehat{x}$. That will give the vector $\boldsymbol{p}$. Then from the formula for $\boldsymbol{p}$, we read off the projection matrix P. These three steps will lead to all projection matrices: *find* $\widehat{x}$, *then find* ***the vector p***, *then find* ***the matrix*** P.

The dotted line $\boldsymbol{b} - \boldsymbol{p}$ is $\boldsymbol{b} - \widehat{x}\boldsymbol{a}$. It is perpendicular to $\boldsymbol{a}$—this will determine $\widehat{x}$. Use the fact that two vectors are perpendicular when their dot product is zero:

$$\boldsymbol{a} \cdot (\boldsymbol{b} - \widehat{x}\boldsymbol{a}) = 0 \quad \text{or} \quad \boldsymbol{a} \cdot \boldsymbol{b} - \widehat{x}\boldsymbol{a} \cdot \boldsymbol{a} = 0 \quad \text{or} \quad \widehat{x} = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\boldsymbol{a} \cdot \boldsymbol{a}} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}. \tag{2}$$

Example 1 Project $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ to find $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$ in Figure 4.4.

Solution The number $\widehat{x}$ is the ratio of $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = 5$ to $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = 9$. So the projection is $\boldsymbol{p} = \frac{5}{9}\boldsymbol{a}$. The error vector between $\boldsymbol{b}$ and $\boldsymbol{p}$ is $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$. Those vectors $\boldsymbol{p}$ and $\boldsymbol{e}$ will add to $\boldsymbol{b}$:

$$\boldsymbol{p} = \frac{5}{9}\boldsymbol{a} = \left(\frac{5}{9}, \frac{10}{9}, \frac{10}{9}\right) \text{ and } \boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \left(\frac{4}{9}, -\frac{1}{9}, -\frac{1}{9}\right).$$

The error $\boldsymbol{e}$ should be perpendicular to $\boldsymbol{a} = (1, 2, 2)$ and it is: $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{a} = \frac{4}{9} - \frac{2}{9} - \frac{2}{9} = 0$.

Projection onto subspace

4F The combination $\widehat{x}_1 \boldsymbol{a}_1 + \cdots + \widehat{x}_n \boldsymbol{a}_n = A\widehat{x}$ that is closest to $\boldsymbol{b}$ comes from

$$A^{\mathrm{T}}(\boldsymbol{b} - A\widehat{x}) = \mathbf{0} \quad \text{or} \quad A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}. \tag{5}$$

The symmetric matrix $A^{\mathrm{T}}A$ is n by n. It is invertible if the $\boldsymbol{a}$'s are independent. The solution is $\widehat{x} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}$. The ***projection*** of $\boldsymbol{b}$ onto the subspace is the vector

$$\boldsymbol{p} = A\widehat{x} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}. \tag{6}$$

This formula shows the n by n ***projection matrix*** that produces $\boldsymbol{p} = P\boldsymbol{b}$:

$$P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}. \tag{7}$$

Start with n vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in $\mathbf{R}^m$. Assume that these $\boldsymbol{a}$'s are linearly independent. ***Problem: Find the combination*** $\widehat{x}_1 \boldsymbol{a}_1 + \cdots + \widehat{x}_n \boldsymbol{a}_n$ ***that is closest to a given vector*** $\boldsymbol{b}$. We are projecting each $\boldsymbol{b}$ in $\mathbf{R}^m$ onto the subspace spanned by the $\boldsymbol{a}$'s.

We solve this problem for an n-dimensional subspace in three steps: *Find the vector $\widehat{x}$, find the projection $p = A\widehat{x}$, find the matrix P.*

The key step was $A^{\mathrm{T}}(b - A\widehat{x}) = \mathbf{0}$. We used geometry ($e$ is perpendicular to all the a's). Linear algebra gives this "normal equation" too, in a very quick way:

1. Our subspace is the column space of A.
2. The error vector $b - A\widehat{x}$ is perpendicular to that column space.
3. Therefore $b - A\widehat{x}$ is in the left nullspace. This means $A^{\mathrm{T}}(b - A\widehat{x}) = \mathbf{0}$.

The left nullspace is important in projections. This nullspace of A^{T} contains the error vector $e = b - A\widehat{x}$. The vector b is being split into the projection p and the error $e = b - p$. Figure 4.5 shows the right triangle with sides p, e, and b.

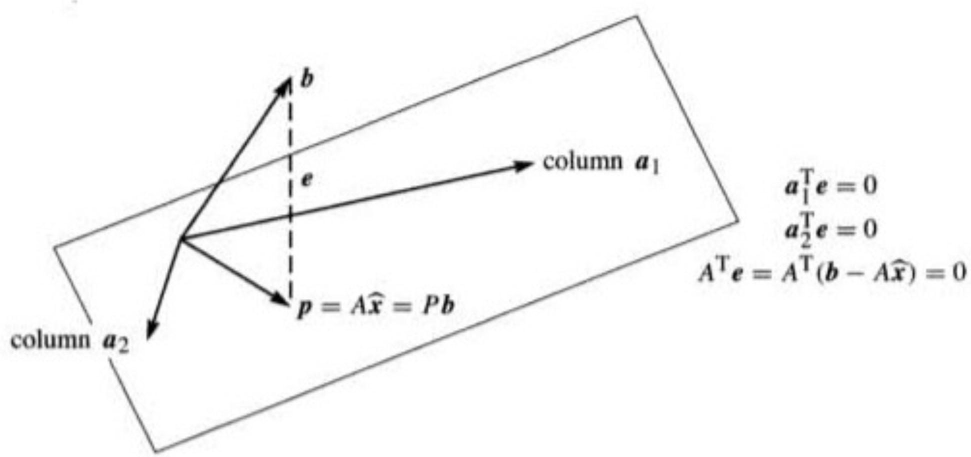

Figure 4.5 The projection p is the nearest point to b in the column space of A. The perpendicular error e must be in the nullspace of A^{T}.

Example 3 If $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ find $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p}$ and P.

Solution Compute the square matrix $A^{\mathrm{T}}A$ and also the vector $A^{\mathrm{T}}\boldsymbol{b}$:

$$A^{\mathrm{T}}A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \quad \text{and} \quad A^{\mathrm{T}}b = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}.$$

Now solve the normal equation $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ to find $\widehat{\boldsymbol{x}}$:

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \quad \text{gives} \quad \widehat{\boldsymbol{x}} = \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}. \tag{8}$$

The combination $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$ is the projection of $\boldsymbol{b}$ onto the column space of A:

$$\boldsymbol{p} = 5\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 3\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}. \quad \text{The error is} \quad \boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}. \tag{9}$$

That solves the problem for one particular $\boldsymbol{b}$. To solve it for every $\boldsymbol{b}$, compute the matrix $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. The determinant of $A^{\mathrm{T}}A$ is $15 - 9 = 6$; $(A^{\mathrm{T}}A)^{-1}$ is easy. Then multiply A times $(A^{\mathrm{T}}A)^{-1}$ times A^{T} to reach P:

$$(A^{\mathrm{T}}A)^{-1} = \frac{1}{6}\begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \quad \text{and} \quad P = \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}. \tag{10}$$

Orthogonal Bases and Gram Schmidt

The vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are ***orthogonal*** when their dot products $\boldsymbol{q}_i \cdot \boldsymbol{q}_j$ are zero. More exactly $\boldsymbol{q}_i^{\mathrm{T}} \boldsymbol{q}_j = 0$ whenever $i \neq j$. With one more step—just divide each vector by its length—the vectors become ***orthogonal unit vectors***. Their lengths are all 1. Then the basis is called ***orthonormal***.

DEFINITION The vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are ***orthonormal*** if

$$q_i^{\mathrm{T}} q_j = \begin{cases} 0 & \text{when } i \neq j \quad (\textbf{\textit{orthogonal}} \text{ vectors}) \\ 1 & \text{when } i = j \quad (\textbf{\textit{unit}} \text{ vectors: } \|\boldsymbol{q}_i\| = 1) \end{cases}$$

A matrix with orthonormal columns is assigned the special letter Q.

The matrix Q is easy to work with because $Q^TQ = I$. This repeats in matrix language that the columns $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are orthonormal. It is equation (1) below, and Q is not required to be square.

When Q is square, $Q^TQ = I$ means that $Q^T = Q^{-1}$: ***transpose = inverse.***

4I *A matrix Q with orthonormal columns satisfies $Q^TQ = I$:*

$$\boxed{Q^TQ} = \begin{bmatrix} - & q_1^T & - \\ - & q_2^T & - \\ & & \\ - & q_n^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & & q_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \boxed{I}. \qquad (1)$$

When row i of Q^T multiplies column j of Q, the dot product is $\boldsymbol{q}_i^T\boldsymbol{q}_j$. Off the diagonal $(i \neq j)$ that dot product is zero by orthogonality. On the diagonal $(i = j)$ the unit vectors give $\boldsymbol{q}_i^T\boldsymbol{q}_i = \|\boldsymbol{q}_i\|^2 = 1$.

If the columns are only orthogonal (not unit vectors), then $Q^{\mathrm{T}}Q$ is a diagonal matrix (not the identity matrix). We wouldn't use the letter Q. But this matrix is almost as good. The important thing is orthogonality—then it is easy to produce unit vectors.

To repeat: $Q^{\mathrm{T}}Q = I$ even when Q is rectangular. In that case Q^{T} is only an inverse from the left. For square matrices we also have $QQ^{\mathrm{T}} = I$, so Q^{T} is the two-sided inverse of Q. The rows of a square Q are orthonormal like the columns. ***The inverse is the transpose***. In this square case we call Q an ***orthogonal matrix***.[2]

Here are three examples of orthogonal matrices—rotation and permutation and reflection. The quickest test is to check $Q^{\mathrm{T}}Q = I$.

Example 1 **(Rotation)** Q rotates every vector in the plane through the angle θ:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{and} \quad Q^{\mathrm{T}} = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

The columns of Q are orthogonal (take their dot product). They are unit vectors because $\sin^2\theta + \cos^2\theta = 1$. Those columns give an ***orthonormal basis*** for the plane $\mathbf{R}^2$. The standard basis vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ are rotated through θ (see Figure 4.10a).

Q^{-1} rotates vectors back through $-\theta$. It agrees with Q^{T}, because the cosine of $-\theta$ is the cosine of θ, and $\sin(-\theta) = -\sin\theta$. We have $Q^{\mathrm{T}}Q = I$ and $QQ^{\mathrm{T}} = I$.

Example 2 **(Permutation)** These matrices change the order to (y, z, x) and (y, x):

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

All columns of these Q's are unit vectors (their lengths are obviously 1). They are also orthogonal (the 1's appear in different places). *The inverse of a permutation matrix is its transpose.* The inverse puts the components back into their original order:

$$\textbf{Inverse = transpose:} \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Every permutation matrix is an orthogonal matrix.

4) *If Q has orthonormal columns ($Q^{\mathrm{T}}Q = I$), it leaves lengths unchanged:*

$$\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\| \text{ for every vector } \boldsymbol{x}. \tag{3}$$

Q also preserves dot products: $(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$. Just use $Q^{\mathrm{T}}Q = I$!

Projection with orthonormal bases

This chapter is about projections onto subspaces. We developed the equations for $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p}$ and P. When the columns of A were a basis for the subspace, all formulas involved $A^{\mathrm{T}}A$. The entries of $A^{\mathrm{T}}A$ are the dot products $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{a}_j$.

Suppose the basis vectors are actually orthonormal. The $\boldsymbol{a}$'s become $\boldsymbol{q}$'s. Then $A^{\mathrm{T}}A$ simplifies to $Q^{\mathrm{T}}Q = I$. Look at the improvements in $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p}$ and P. Instead of $Q^{\mathrm{T}}Q$ we print a blank for the identity matrix:

$$\underline{\quad\quad}\,\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b} \quad \text{and} \quad \boldsymbol{p} = Q\widehat{\boldsymbol{x}} \quad \text{and} \quad P = Q\,\underline{\quad\quad}\,Q^{\mathrm{T}}. \tag{4}$$

The least squares solution of $Qx = b$ is $\widehat{x} = Q^{\mathrm{T}}b$. The projection matrix is $P = QQ^{\mathrm{T}}$.

There are no matrices to invert. This is the point of an orthonormal basis. The best $\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$ just has dot products of $\boldsymbol{b}$ with the rows of Q^{T}, which are the $\boldsymbol{q}$'s:

$$\widehat{\boldsymbol{x}} = \begin{bmatrix} - & \boldsymbol{q}_1^{\mathrm{T}} & - \\ & \vdots & \\ - & \boldsymbol{q}_n^{\mathrm{T}} & - \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{b} \\ \\ \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b} \end{bmatrix} \qquad \text{(dot products)}$$

We have n separate 1-dimensional projections. The "coupling matrix" or "correlation matrix" $A^{\mathrm{T}}A$ is now $Q^{\mathrm{T}}Q = I$. There is no coupling. Here is $\boldsymbol{p} = Q\widehat{\boldsymbol{x}}$:

$$\textbf{\textit{Projection}} \qquad p = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix} \begin{bmatrix} q_1^{\mathrm{T}}b \\ \vdots \\ q_n^{\mathrm{T}}b \end{bmatrix} = q_1(q_1^{\mathrm{T}}b) + \cdots + q_n(q_n^{\mathrm{T}}b). \qquad (5)$$

Important case: When Q is square and $m = n$, the subspace is the whole space. Then $Q^{\mathrm{T}} = Q^{-1}$ and $\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$ is the same as $\boldsymbol{x} = Q^{-1}\boldsymbol{b}$. The solution is exact! The projection of $\boldsymbol{b}$ onto the whole space is $\boldsymbol{b}$ itself. In this case $P = QQ^{\mathrm{T}} = I$.

You may think that projection onto the whole space is not worth mentioning. But when $\boldsymbol{p} = \boldsymbol{b}$, our formula assembles $\boldsymbol{b}$ out of its 1-dimensional projections. If $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ is an orthonormal basis for the whole space, so Q is square, then every $\boldsymbol{b}$ is the sum of its components along the $\boldsymbol{q}$'s:

$$\boldsymbol{b} = \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) + \boldsymbol{q}_2(\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b}) + \cdots + \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}). \qquad (6)$$