Linear Algebra

Orthogonal Bases and Gram-Schmidt

Source: Introduction to Linear Algebra by Gilbert Strang

- **How orthogonality and orthonormality helps**
- **How to construct orthonormal vectors from original vectors**

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

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

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

The matrix Q is easy to work with because $Q^{\mathrm{T}}Q = I$. 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, $Q^{\mathrm{T}}Q = I$ means that $Q^{\mathrm{T}} = Q^{-1}$: ***transpose = inverse***.

4I *A matrix Q with orthonormal columns satisfies $Q^{\mathrm{T}}Q = I$:*

$$Q^{\mathrm{T}}Q = \begin{bmatrix} - \ \boldsymbol{q}_1^{\mathrm{T}} \ - \\ - \ \boldsymbol{q}_2^{\mathrm{T}} \ - \\ \\ - \ \boldsymbol{q}_n^{\mathrm{T}} \ - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & & \boldsymbol{q}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I. \qquad (1)$$

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]

We have $Q^{\mathrm{T}}Q = I$ and $QQ^{\mathrm{T}} = I$.

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

Every permutation matrix is an orthogonal matrix.

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

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 $\mathbf{R}^2$.

$$Q = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

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

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

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

This chapter is about projections onto subspaces. We developed the equations for $\widehat{x}$ and 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 $a_i^{\mathrm{T}}a_j$.

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

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

The point of this section is that "orthogonal is good." Projections and least squares always involve $A^{\mathrm{T}}A$. When this matrix becomes $Q^{\mathrm{T}}Q = I$, the inverse is no problem.

For this to be true, we had to say "*If* the vectors are orthonormal." ***Now we find a way to create orthonormal vectors.***

Start with three independent vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$. We intend to construct three orthogonal vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Then (at the end is easiest) we divide $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ by their lengths. That produces three orthonormal vectors $\boldsymbol{q}_1 = \boldsymbol{A}/\|\boldsymbol{A}\|$, $\boldsymbol{q}_2 = \boldsymbol{B}/\|\boldsymbol{B}\|$, $\boldsymbol{q}_3 = \boldsymbol{C}/\|\boldsymbol{C}\|$.

Gram-Schmidt Begin by choosing $\boldsymbol{A} = \boldsymbol{a}$. This first direction is accepted. The next direction $\boldsymbol{B}$ must be perpendicular to $\boldsymbol{A}$. ***Start with b and subtract its projection along***

$\boldsymbol{A}$. This leaves the perpendicular part, which is the orthogonal vector $\boldsymbol{B}$:

Gram-Schmidt idea $$\boxed{B = b - \frac{A^{\mathrm{T}}b}{A^{\mathrm{T}}A}A.}$$ (7

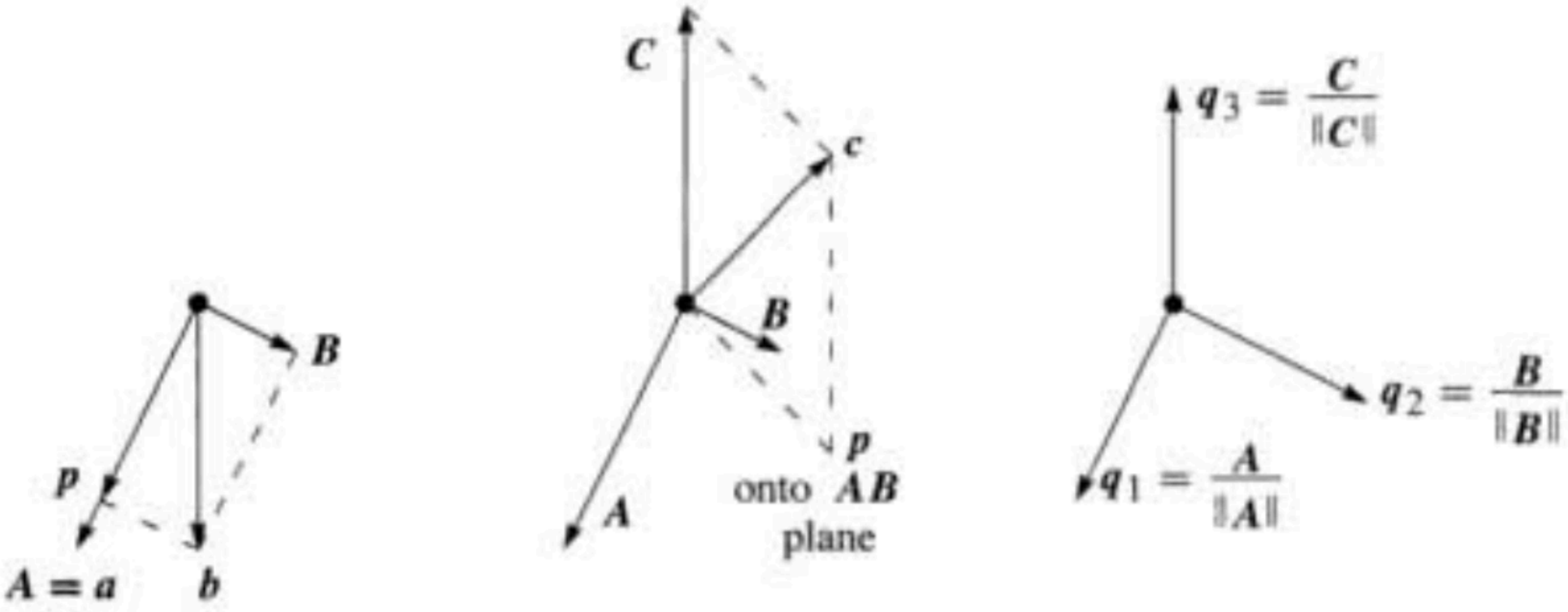

Figure 4.11 First project $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ and find $\boldsymbol{B}$ as $\boldsymbol{b}-\boldsymbol{p}$. Then project $\boldsymbol{c}$ onto the $\boldsymbol{AB}$ plane and find $\boldsymbol{C}$ as $\boldsymbol{c}-\boldsymbol{p}$. Then divide by $\|\boldsymbol{A}\|$, $\|\boldsymbol{B}\|$, and $\|\boldsymbol{C}\|$.

$\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal in Figure 4.11. Take the dot product with $\boldsymbol{A}$ to verify that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = 0$. This vector $\boldsymbol{B}$ is what we have called the error vector $\boldsymbol{e}$, perpendicular to $\boldsymbol{A}$. Notice that $\boldsymbol{B}$ in equation (7) is not zero (otherwise $\boldsymbol{a}$ and $\boldsymbol{b}$ would be dependent). The directions $\boldsymbol{A}$ and $\boldsymbol{B}$ are now set.

The third direction starts with $\boldsymbol{c}$. This is not a combination of $\boldsymbol{A}$ and $\boldsymbol{B}$ (because $\boldsymbol{c}$ is not a combination of $\boldsymbol{a}$ and $\boldsymbol{b}$). But most likely $\boldsymbol{c}$ is not perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$. So subtract off its components in those two directions to get $\boldsymbol{C}$:

$$C = c - \frac{A^{\mathrm{T}} c}{A^{\mathrm{T}} A} A - \frac{B^{\mathrm{T}} c}{B^{\mathrm{T}} B} B. \tag{8}$$

This is the one and only idea of the Gram-Schmidt process. ***Subtract from every new vector its projections in the directions already set.*** That idea is repeated at every step.[2] If we also had a fourth vector $\boldsymbol{d}$, we would subtract its projections onto $\boldsymbol{A}$, $\boldsymbol{B}, \boldsymbol{C}$ to get $\boldsymbol{D}$. At the end, divide the orthogonal vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ by their lengths. The resulting vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \boldsymbol{q}_4$ are orthonormal.

Example 5 Suppose the independent non-orthogonal vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{c} = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}.$$

Then $\boldsymbol{A} = \boldsymbol{a}$ has $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = 2$. Subtract from $\boldsymbol{b}$ its projection along $\boldsymbol{A} = (1, -1, 0)$:

$$\boldsymbol{B} = \boldsymbol{b} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A} = \boldsymbol{b} - \tfrac{2}{2}\boldsymbol{A} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

Check: $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} = 0$ as required. Now subtract two projections from $\boldsymbol{c}$ to get $\boldsymbol{C}$:

$$\boldsymbol{C} = \boldsymbol{c} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A} - \frac{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}}\boldsymbol{B} = \boldsymbol{c} - \tfrac{6}{2}\boldsymbol{A} + \tfrac{6}{6}\boldsymbol{B} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Check: $\boldsymbol{C} = (1, 1, 1)$ is perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$. Finally convert $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ to unit vectors (length 1, orthonormal). The lengths of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are $\sqrt{2}$ and $\sqrt{6}$ and $\sqrt{3}$. Divide by those lengths, for an orthonormal basis:

$$\boldsymbol{q}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{q}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{q}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

$$\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^{\mathrm{T}} a & q_1^{\mathrm{T}} b & q_1^{\mathrm{T}} c \\ & q_2^{\mathrm{T}} b & q_2^{\mathrm{T}} c \\ & & q_3^{\mathrm{T}} c \end{bmatrix} \quad \text{or} \quad \boxed{A = QR.} \tag{9}$$

$A = QR$ is Gram-Schmidt in a nutshell. Multiply by Q^{T} to see why $R = Q^{\mathrm{T}} A$.

4K (Gram-Schmidt) From independent vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, Gram-Schmidt constructs orthonormal vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$. The matrices with these columns satisfy $A = QR$. Then $R = Q^{\mathrm{T}} A$ is ***triangular*** because later $\boldsymbol{q}$'s are orthogonal to earlier $\boldsymbol{a}$'s.

Here are the $\boldsymbol{a}$'s and $\boldsymbol{q}$'s from the example. The i, j entry of $R = Q^{\mathrm{T}} A$ is row i of Q^{T} times column j of A. This is the dot product of $\boldsymbol{q}_i$ with $\boldsymbol{a}_j$:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & \sqrt{3} \end{bmatrix} = QR.$$

The lengths of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are the numbers $\sqrt{2}, \sqrt{6}, \sqrt{3}$ on the diagonal of R. Because of the square roots, QR looks less beautiful than LU. Both factorizations are absolutely central to calculations in linear algebra.

Any m by n matrix A with independent columns can be factored into QR. The m by n matrix Q has orthonormal columns, and the square matrix R is upper triangular with positive diagonal.