# Orthogonal matrices and Gram-Schmidt

In this lecture we finish introducing orthogonality. Using an orthonormal basis or a matrix with orthonormal columns makes calculations much easier. The Gram-Schmidt process starts with any basis and produces an orthonormal basis that spans the same space as the original basis.

## Orthonormal vectors

The vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_n$ are *orthonormal* if:

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

In other words, they all have (normal) length 1 and are perpendicular (ortho) to each other. Orthonormal vectors are always independent.

## Orthonormal matrix

If the columns of $Q = \begin{bmatrix} \mathbf{q}_1 & \ldots & \mathbf{q}_n \end{bmatrix}$ are orthonormal, then $Q^TQ = I$ is the identity.

Matrices with orthonormal columns are a new class of important matrices to add to those on our list: triangular, diagonal, permutation, symmetric, reduced row echelon, and projection matrices. We'll call them "orthonormal matrices".

A square orthonormal matrix $Q$ is called an *orthogonal matrix*. If $Q$ is square, then $Q^TQ = I$ tells us that $Q^T = Q^{-1}$.

For example, if $Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ then $Q^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$. Both $Q$ and $Q^T$ are orthogonal matrices, and their product is the identity.

The matrix $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal. The matrix $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ is not, but we can adjust that matrix to get the orthogonal matrix $Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. We can use the same tactic to find some larger orthogonal matrices called *Hadamard matrices*:

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

An example of a rectangular matrix with orthonormal columns is:

$$Q = \frac{1}{3} \begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}.$$

We can extend this to a (square) orthogonal matrix:

$$\frac{1}{3}\begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix}.$$

These examples are particularly nice because they don't include complicated square roots.

## Orthonormal columns are good

Suppose $Q$ has orthonormal columns. The matrix that projects onto the column space of $Q$ is:

$$P = Q^T(Q^TQ)^{-1}Q^T.$$

If the columns of $Q$ are orthonormal, then $Q^TQ = I$ and $P = QQ^T$. If $Q$ is square, then $P = I$ because the columns of $Q$ span the entire space.

Many equations become trivial when using a matrix with orthonormal columns. If our basis is orthonormal, the projection component $\hat{x}_i$ is just $\mathbf{q}_i^T\mathbf{b}$ because $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$ becomes $\hat{\mathbf{x}} = Q^T\mathbf{b}$.

## Gram-Schmidt

With elimination, our goal was "make the matrix triangular". Now our goal is "make the matrix orthonormal".

We start with two independent vectors **a** and **b** and want to find orthonormal vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ that span the same plane. We start by finding orthogonal vectors **A** and **B** that span the same space as **a** and **b**. Then the unit vectors $\mathbf{q}_1 = \frac{\mathbf{A}}{||\mathbf{A}||}$ and $\mathbf{q}_2 = \frac{\mathbf{B}}{||\mathbf{B}||}$ form the desired orthonormal basis.

Let $\mathbf{A} = \mathbf{a}$. We get a vector orthogonal to **A** in the space spanned by **a** and **b** by projecting **b** onto **a** and letting $\mathbf{B} = \mathbf{b} - \mathbf{p}$. (**B** is what we previously called **e**.)

$$\mathbf{B} = \mathbf{b} - \frac{\mathbf{A}^T\mathbf{b}}{\mathbf{A}^T\mathbf{A}}\mathbf{A}.$$

If we multiply both sides of this equation by $\mathbf{A}^T$, we see that $\mathbf{A}^T\mathbf{B} = 0$.

What if we had started with three independent vectors, **a**, **b** and **c**? Then we'd find a vector **C** orthogonal to both **A** and **B** by subtracting from **c** its components in the **A** and **B** directions:

A、B are orthogonal to each other

$$\mathbf{C} = \mathbf{c} - \frac{\mathbf{A}^T\mathbf{c}}{\mathbf{A}^T\mathbf{A}}\mathbf{A} - \frac{\mathbf{B}^T\mathbf{c}}{\mathbf{B}^T\mathbf{B}}\mathbf{B}. = c\left(I - \left(\frac{A}{A^TA}A^T + \frac{B}{B^TB}B^T\right)\right)$$

$$O = \begin{bmatrix} | & | \\ A & B \\ | & | \end{bmatrix}$$

$$P = O(O^TO)^{-1}O^T$$
$$= O\begin{bmatrix} A^TA & \\ & B^TB \end{bmatrix}^{-1} O^T$$
$$= O\begin{bmatrix} (A^TA)^{-1} & \\ & (B^TB)^{-1} \end{bmatrix} O^T$$
$$= \begin{bmatrix} | & | \\ (A^TA)^{-1}A & (B^TB)^{-1}B \\ | & | \end{bmatrix} O^T$$
$$= \begin{bmatrix} | & | \\ (A^TA)^{-1}A & (B^TB)^{-1}B \\ | & | \end{bmatrix}\begin{bmatrix} - A^T - \\ - B^T - \end{bmatrix}$$

$$P_c = \begin{bmatrix} \cdots \end{bmatrix}\begin{bmatrix} A^Tc \\ B^Tc \end{bmatrix}$$
$$= \begin{bmatrix} | & | \\ (A^TA)^{-1}A & (B^TB)^{-1}B \\ | & | \end{bmatrix}\begin{bmatrix} A^Tc \\ B^Tc \end{bmatrix}$$
$$= \begin{bmatrix} | & | \\ (A^TA)^{-1}A\,A^Tc + (B^TB)^{-1}B\,B^Tc \\ | & | \end{bmatrix}$$
$$= \underbrace{(A^TA)^{-1}}_{\text{Sclar}}A\underbrace{A^Tc}_{\text{Sclar}} + \underbrace{(B^TB)^{-1}}_{\text{Sclar}}B\underbrace{B^Tc}_{\text{Sclar}}$$
$$= \frac{A^Tc}{(A^TA)^{-1}}A + \frac{B^Tc}{(B^TB)^{-1}}B$$

For example, suppose $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$. Then $\mathbf{A} = \mathbf{a}$ and:

$$\begin{aligned} \mathbf{B} &= \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{\mathbf{A}^T\mathbf{b}}{\mathbf{A}^T\mathbf{A}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}. \end{aligned}$$

Normalizing, we get:

$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{2} \end{bmatrix}.$$

The column space of $Q$ is the plane spanned by $\mathbf{a}$ and $\mathbf{b}$.

When we studied elimination, we wrote the process in terms of matrices and found $A = LU$. A similar equation $A = QR$ relates our starting matrix $A$ to the result $Q$ of the Gram-Schmidt process. Where $L$ was lower triangular, $R$ is upper triangular.

Suppose $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$. Then:

$$\overset{A}{\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}} = \overset{Q}{\begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix}} \overset{R}{\begin{bmatrix} \mathbf{a}_1^T\mathbf{q}_1 & \mathbf{a}_2^T\mathbf{q}_1 \\ \mathbf{a}_1^T\mathbf{q}_2 & \mathbf{a}_2^T\mathbf{q}_2 \end{bmatrix}}.$$

If $R$ is upper triangular, then it should be true that $\mathbf{a}_1^T\mathbf{q}_2 = 0$. This must be true because we chose $\mathbf{q}_1$ to be a unit vector in the direction of $\mathbf{a}_1$. All the later $\mathbf{q}_i$ were chosen to be perpendicular to the earlier ones.

Notice that $R = Q^T A$. This makes sense; $Q^T Q = I$.