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 \( q_1, q_2, \ldots, q_n \) are orthonormal if:

\[
q_i^T q_j = \begin{cases} 
0 & \text{if } i \neq j \\
1 & \text{if } i = j.
\end{cases}
\]

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 \( Q = [ \quad q_1 \ldots q_n \quad ] \) are orthonormal, then \( Q^T Q = I \) 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 \( Q^T Q = I \) 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:

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

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:

\[
\begin{bmatrix}
1 & -2 & 2 \\
-2 & -2 & 1 \\
2 & 2 & 1
\end{bmatrix}.
\]

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 \( q_i^Tb \) because \( A^TA\hat{x} = A^Tb \) becomes \( \hat{x} = Q^Tb \).

**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 \( q_1 \) and \( q_2 \) 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 \( q_1 = \frac{A}{||A||} \) and \( q_2 = \frac{B}{||B||} \) form the desired orthonormal basis.

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

\[
B = b - \frac{A^Tb}{A^TA}A.
\]

If we multiply both sides of this equation by \( A^T \), we see that \( A^TB = 0 \).

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:

\[
C = c - \frac{A^Tc}{A^TA}A - \frac{B^Tc}{B^TB}B.
\]
For example, suppose $a = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$. Then $A = a$ and:

$$B = \begin{bmatrix}
1 \\
0 \\
2
\end{bmatrix} - \frac{A^T b}{A^T A} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix}
1 \\
0 \\
2
\end{bmatrix} - \frac{3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix}
0 \\
-1 \\
1
\end{bmatrix}.$$  

Normalizing, we get:

$$Q = \begin{bmatrix} q_1 & q_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{2} \end{bmatrix}.$$  

The column space of $Q$ is the plane spanned by $a$ and $b$.

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 $A = \begin{bmatrix} a_1 & a_2 \end{bmatrix}$. Then:

$$A \begin{bmatrix} Q & R \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \end{bmatrix} \begin{bmatrix} a_1^T q_1 & a_2^T q_1 \\ a_1^T q_2 & a_2^T q_2 \end{bmatrix}.$$  

If $R$ is upper triangular, then it should be true that $a_1^T q_2 = 0$. This must be true because we chose $q_1$ to be a unit vector in the direction of $a_1$. All the later $q_i$ were chosen to be perpendicular to the earlier ones.

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

3