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 $\mathbf{q}_1, \mathbf{q}_2, ...\mathbf{q}_n$ 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 $Q = \begin{bmatrix} \mathbf{q}_1 & ... & \mathbf{q}_n \end{bmatrix}$ 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*:

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

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:

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

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:

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

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; $Q^T Q = I$.

Exercises on orthogonal matrices and Gram-Schmidt

Problem 17.1: (4.4 #10.b *Introduction to Linear Algebra:* Strang)

Orthonormal vectors are automatically linearly independent.

Matrix Proof: Show that $Q\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$. Since Q may be rectangular, you can use Q^T but not Q^{-1}.

Problem 17.2: (4.4 #18) Given the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ listed below, use the Gram-Schmidt process to find orthogonal vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ that span the same space.

$$\mathbf{a} = (1, -1, 0, 0), \mathbf{b} = (0, 1, -1, 0), \mathbf{c} = (0, 0, 1, -1).$$

Show that $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ and $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ are bases for the space of vectors perpendicular to $\mathbf{d} = (1, 1, 1, 1)$.

Exercises on orthogonal matrices and Gram-Schmidt

Problem 17.1: (4.4 #10.b *Introduction to Linear Algebra:* Strang) Orthonormal vectors are automatically linearly independent.

Matrix Proof: Show that $Q\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$. Since Q may be rectangular, you can use Q^T but not Q^{-1}.

Solution: By definition, Q is a matrix whose columns are orthonormal, and so we know that $Q^TQ = I$ (where Q may be rectangular). Then:

$$Q\mathbf{x} = \mathbf{0} \Longrightarrow Q^TQ\mathbf{x} = Q^T\mathbf{0} \Longrightarrow I\mathbf{x} = \mathbf{0} \Longrightarrow \mathbf{x} = \mathbf{0}.$$

Thus the nullspace of Q is the zero vector, and so the columns of Q are linearly independent. There are no non-zero linear combinations of the columns that equal the zero vector. Thus, orthonormal vectors are automatically linearly independent.

Problem 17.2: (4.4 #18) Given the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ listed below, use the Gram-Schmidt process to find orthogonal vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ that span the same space.

$$\mathbf{a} = (1,-1,0,0), \mathbf{b} = (0,1,-1,0), \mathbf{c} = (0,0,1,-1).$$

Show that $\{\mathbf{A},\mathbf{B},\mathbf{C}\}$ and $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$ are bases for the space of vectors perpendicular to $\mathbf{d} = (1,1,1,1)$.

Solution: We apply Gram-Schmidt to $\mathbf{a},\mathbf{b},\mathbf{c}$. First, we set

$$\mathbf{A} = \mathbf{a} = (\mathbf{1},-\mathbf{1},\mathbf{0},\mathbf{0}).$$

Next we find $\mathbf{B}$:

$$\mathbf{B} = \mathbf{b} - \frac{\mathbf{A^Tb}}{\mathbf{A^TA}}\mathbf{A} = (0,1,-1,0) + \frac{1}{2}(1,-1,0,0) = \left(\frac{\mathbf{1}}{\mathbf{2}},\frac{\mathbf{1}}{\mathbf{2}},-\mathbf{1},\mathbf{0}\right).$$

And then we find $\mathbf{C}$:

$$\mathbf{C} = \mathbf{c} - \frac{\mathbf{A^Tc}}{\mathbf{A^TA}}\mathbf{A} - \frac{\mathbf{B^Tc}}{\mathbf{B^TB}}\mathbf{B} = (0,0,1,-1) + \frac{2}{3}\left(\frac{1}{2},\frac{1}{2},-1,0\right) = \left(\frac{\mathbf{1}}{\mathbf{3}},\frac{\mathbf{1}}{\mathbf{3}},\frac{\mathbf{1}}{\mathbf{3}},-\mathbf{1}\right).$$

We know from the first problem that the elements of the set $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ are linearly independent, and each vector is orthogonal to (1,1,1,1). The space of vectors perpendicular to $\mathbf{d}$ is three dimensional (since the row space of $(1, 1, 1, 1)$ is one-dimensional, and the number of dimensions of the row space added to the number of dimensions of the nullspace add to 4). Therefore $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ forms a basis for the space of vectors perpendicular to $\mathbf{d}$.

Similarly, $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is a basis for the space of vectors perpendicular to $\mathbf{d}$ because the vectors are linearly independent, orthogonal to (1,1,1,1), and because there are three of them.