

Lecture 18

We've now spent a good deal of time talking about projections, and what they can do for us. In particular, projections are at the heart of solving the **least-squares problem** which is used for data fitting. It turns out that projections are super easy if you have an **orthogonal basis** for the subspace you're projecting on to.

Suppose $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ form a basis for some subspace S . If some vector \mathbf{b} in S is given by $\mathbf{b} = 1\mathbf{a}_1 + 2\mathbf{a}_2 - 3\mathbf{a}_3$ we say that $(1, 2, -3)$ are the **coordinates** of \mathbf{b} with respect to the basis.

OK, so given a vector \mathbf{b} in S and the basis vectors, how do we find the coordinates of \mathbf{b} ?

One way we could do this is to stack the basis vectors in a matrix, $A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3]$ and solve the linear system

$$\mathbf{b} = A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

OK, so sometimes this is easy, but sometimes it's hard. It turns out that things get way easier if your basis vectors are orthogonal. Things get even *nicer* if the basis vectors are orthogonal and have unit length.

Consider the way we generally think of vectors in \mathbb{R}^3 . The vector $[1, 2, -3]^T$ has coordinates $(1, 2, -3)$ with respect to the standard basis vectors \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 . The standard basis vectors are orthogonal and have unit length!

But there can be other bases with this same property. Suppose that \mathbf{q}_1 , \mathbf{q}_2 , and \mathbf{q}_3 are a basis for some subspace S that satisfy

$$\mathbf{q}_i^T \mathbf{q}_j = 0 \quad \text{for } i \neq j \quad \text{and also} \quad \|\mathbf{q}_j\| = 1 \quad \text{for all } j$$

Then we say that the \mathbf{q} 's form an **orthonormal** basis for S .

OK, so how do we find x_1 , x_2 , and x_3 such that $\mathbf{b} = x_1 \mathbf{q}_1 + x_2 \mathbf{q}_2 + x_3 \mathbf{q}_3$?

Notice that if we take the dot product of \mathbf{b} with \mathbf{q}_1 we get

$$\mathbf{q}_1^T \mathbf{b} = x_1 \mathbf{q}_1^T \mathbf{q}_1 + x_2 \mathbf{q}_1^T \mathbf{q}_2 + x_3 \mathbf{q}_1^T \mathbf{q}_3 = x_1$$

where the last equality follows from the fact that the dot product of \mathbf{q}_1 with \mathbf{q}_2 and \mathbf{q}_3 are both zero, because they're orthogonal. Also, the dot product of \mathbf{q}_1 with itself is one.

Similarly, we have

$$\mathbf{q}_2^T \mathbf{b} = x_1 \mathbf{q}_2^T \mathbf{q}_1 + x_2 \mathbf{q}_2^T \mathbf{q}_2 + x_3 \mathbf{q}_2^T \mathbf{q}_3 = x_2$$

$$\mathbf{q}_3^T \mathbf{b} = x_1 \mathbf{q}_3^T \mathbf{q}_1 + x_2 \mathbf{q}_3^T \mathbf{q}_2 + x_3 \mathbf{q}_1^T \mathbf{q}_3 = x_3$$

So, if we have an orthonormal basis, finding the coordinates of \mathbf{b} in that basis are as simple as taking the dot product of the basis vectors with \mathbf{b} . That's **way easier** than solving a linear system.

Example 1: Let $\mathbf{q}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}$, $\mathbf{q}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$, and $\mathbf{q}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$. Find the coords of $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ in this basis.

We have

$$x_1 = \mathbf{q}_1^T \mathbf{b} = 1/3, \quad x_2 = \mathbf{q}_2^T \mathbf{b} = -2/3, \quad x_3 = \mathbf{q}_3^T \mathbf{b} = 2/3$$

So

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{3} \mathbf{q}_1 - \frac{2}{3} \mathbf{q}_2 + \frac{2}{3} \mathbf{q}_3 = \begin{bmatrix} 1/9 \\ 2/9 \\ 2/9 \end{bmatrix} + \begin{bmatrix} 4/9 \\ 2/9 \\ -4/9 \end{bmatrix} + \begin{bmatrix} 4/9 \\ -4/9 \\ 2/9 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \checkmark$$

Nice things happen when we put orthonormal vectors into a matrix. Consider $Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3]$ where the \mathbf{q}'_s are length- m vectors. Notice that

$$Q^T Q = \begin{bmatrix} \mathbf{q}_1^T \mathbf{q}_1 & \mathbf{q}_1^T \mathbf{q}_2 & \mathbf{q}_1^T \mathbf{q}_3 \\ \mathbf{q}_2^T \mathbf{q}_1 & \mathbf{q}_2^T \mathbf{q}_2 & \mathbf{q}_2^T \mathbf{q}_3 \\ \mathbf{q}_3^T \mathbf{q}_1 & \mathbf{q}_3^T \mathbf{q}_2 & \mathbf{q}_3^T \mathbf{q}_3 \end{bmatrix} = I$$

Example 2: Let $Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{3} \end{bmatrix}$. Then

$$Q^T Q = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{3} \end{bmatrix} = \begin{bmatrix} 1/2 + 1/2 & 1/\sqrt{6} - 1/\sqrt{6} \\ 1/\sqrt{6} - 1/\sqrt{6} & 1/3 + 1/3 + 1/3 \end{bmatrix} = I$$

If the matrix Q with orthonormal columns is a square matrix then $QQ^T = I$ as well. But if Q is not square then $QQ^T \neq I$.

If Q is square and $Q^T Q = QQ^T = I$ then we call Q an **orthogonal matrix**.

Notice that a nice property of orthogonal matrices is that $Q^T = Q^{-1}$!

Orthogonal matrices exhibit nice properties when they act on vectors. In particular, multiplication of a vector by an orthogonal matrix does not change the length of the vector. Furthermore, if you apply an orthogonal matrix to two vectors, it does not change the angle between the two vectors.

Proof: If Q is orthogonal then $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.

We'll show the equivalent fact that $\|Q\mathbf{x}\|^2 = \|\mathbf{x}\|^2$.

$$\|Q\mathbf{x}\|^2 = (Q\mathbf{x})^T (Q\mathbf{x}) = \mathbf{x}^T Q^T Q \mathbf{x} = \mathbf{x}^T I \mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|^2$$

Proof: If Q is orthogonal then $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.

$$(Q\mathbf{x}) \cdot (Q\mathbf{y}) = (Q\mathbf{x})^T (Q\mathbf{y}) = \mathbf{x}^T Q^T Q \mathbf{y} = \mathbf{x}^T I \mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$$

Having an orthonormal basis for a subspace makes projecting onto that subspace extraordinarily easy. Suppose that $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ is a basis for S and let $Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3]$. Then the projection of \mathbf{b} onto S is given by $\mathbf{p} = Q\hat{\mathbf{x}}$ where $\hat{\mathbf{x}}$ solves the normal equations $Q^T Q \mathbf{x} = Q^T \mathbf{b}$. But $Q^T Q = I$ so we have immediately $\hat{\mathbf{x}} = Q^T \mathbf{b}$. Plugging this into the projection equation we have $\mathbf{p} = QQ^T \mathbf{b}$, which means that the usually complicated to compute projection matrix is given by $P = QQ^T$.

OK, so an orthonormal basis for a subspace makes projects easier. But, given any old basis for a subspace $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$, can we come up with an orthonormal basis for that same subspace?

The answer is YES, and it's done using an algorithm called the **Gram-Schmidt** process.

Let's starting with an easy case where we have two linearly independent vectors \mathbf{w}_1 and \mathbf{w}_2 and we want to find orthonormal vectors \mathbf{q}_1 and \mathbf{q}_2 such that $\text{span}\{\mathbf{q}_1, \mathbf{q}_2\} = \text{span}\{\mathbf{w}_1, \mathbf{w}_2\}$.

We're going to start by finding vectors \mathbf{v}_1 and \mathbf{v}_2 which are orthogonal and span the same space as \mathbf{w}_1 and \mathbf{w}_2 . Then we'll scale them to have unit length.

The first step is to simply set $\mathbf{v}_1 = \mathbf{w}_1$.

Next we define \mathbf{v}_2 as follows

$$\mathbf{v}_2 = \mathbf{w}_2 - r\mathbf{v}_1$$

where r is a yet-to-be determined constant. Notice that necessarily the \mathbf{v} 's span the same space as the \mathbf{w} 's since the \mathbf{v} 's are simple linear combinations of the \mathbf{w} 's. To make it so that \mathbf{w}_2 is orthogonal to \mathbf{w}_1 we simply choose r to make it so. We do this by dotting \mathbf{w}_1 and \mathbf{w}_2 together and choosing r so that the result is zero.

$$\mathbf{v}_2^T \mathbf{v}_1 = (\mathbf{w}_2 - r\mathbf{v}_1)^T \mathbf{v}_1 = 0$$

Distributing the dot product, we have

$$\mathbf{w}_2^T \mathbf{v}_1 - r\mathbf{v}_1^T \mathbf{v}_1 = 0 \quad \Rightarrow \quad \mathbf{w}_2^T \mathbf{v}_1 = r\mathbf{v}_1^T \mathbf{v}_1 \quad \Rightarrow \quad r = \frac{\mathbf{w}_2^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}$$

$$\text{Then } \mathbf{v}_2 = \mathbf{w}_2 - \frac{\mathbf{w}_2^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1} \mathbf{v}_1$$

The last step is to normalize the \mathbf{v} 's:

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} \quad \text{and} \quad \mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$$

Example 2: Let $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$.

First we set $\mathbf{v}_1 = \mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Next we compute $r = \frac{\mathbf{w}_2^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1} = \frac{3}{3}$. Then we form

$$\mathbf{v}_2 = \mathbf{w}_2 - r\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} - \frac{3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Normalizing, we have

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

Now, what if we add a third vector into the mix?

Example 3: Find an orthonormal basis for the space spanned by $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, and $\mathbf{w}_3 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$.

The good news is that the work we did in Example 2 is still good. We already have vectors \mathbf{v}_1 and \mathbf{v}_2 that are orthogonal and in the desired space. Next we need to choose \mathbf{v}_3 that is orthogonal to \mathbf{v}_1 and \mathbf{v}_2 and is in the span of the \mathbf{w} 's. We do this by writing \mathbf{v}_3 as a linear combination of \mathbf{v}_1 and \mathbf{v}_2 :

$$\mathbf{v}_3 = \mathbf{w}_3 - r_1 \mathbf{v}_1 - r_2 \mathbf{v}_2$$

The coefficients r_1 and r_2 will be chosen to enforce the orthogonality. We have

$$0 = \mathbf{v}_3^T \mathbf{v}_1 = (\mathbf{w}_3 - r_1 \mathbf{v}_1 - r_2 \mathbf{v}_2)^T \mathbf{v}_1 \quad \Rightarrow \quad 0 = \mathbf{w}_3^T \mathbf{v}_1 - r_1 \mathbf{v}_1^T \mathbf{v}_1 \quad \Rightarrow \quad r_1 = \frac{\mathbf{w}_3^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}$$

$$0 = \mathbf{v}_3^T \mathbf{v}_2 = (\mathbf{w}_3 - r_1 \mathbf{v}_1 - r_2 \mathbf{v}_2)^T \mathbf{v}_2 \quad \Rightarrow \quad 0 = \mathbf{w}_3^T \mathbf{v}_2 - r_2 \mathbf{v}_2^T \mathbf{v}_2 \quad \Rightarrow \quad r_2 = \frac{\mathbf{w}_3^T \mathbf{v}_2}{\mathbf{v}_2^T \mathbf{v}_2}$$

For this example, we have

$$r_1 = \frac{3}{3} = 1 \quad \text{and} \quad r_2 = \frac{1}{2}$$

Then

$$\mathbf{v}_3 = \mathbf{w}_3 - \mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1 \\ -1/2 \end{bmatrix}$$

Normalizing then gives $\mathbf{q}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}$.

It turns out that the Gram-Schmidt process gives us another way to factor a matrix. We can write $A = QR$ where Q has orthonormal columns and R is upper triangular.

Example 4: Let $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$. We found in Example 2 that an orthonormal basis for the column space of A is given by

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

To get the R in the QR factorization we note that since Q has orthonormal columns

$$A = QR \quad \Rightarrow \quad Q^T A = Q^T QR \quad \Rightarrow \quad Q^T A = R$$

Thus we can find R simply by multiplying A on the left by Q^T . For this example, we have

$$R = Q^T A = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & \sqrt{3} \\ 0 & \sqrt{2} \end{bmatrix}$$

Thus, we can write

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

Example 5: For the vectors in Example 3, we have

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

If A is a square nonsingular matrix, then the QR factorization gives us another slick way to solve $A\mathbf{x} = \mathbf{b}$. We have

$$A\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad QR\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad Q^T QR\mathbf{x} = Q^T \mathbf{b} \quad \Rightarrow \quad R\mathbf{x} = Q^T \mathbf{b} \quad \Rightarrow \quad \mathbf{x} = R^{-1} Q^T \mathbf{b}$$

Thus $A\mathbf{x} = \mathbf{b}$ can be solved by multiplying \mathbf{b} by Q^T and then using back substitution to solve $R\mathbf{x} = Q^T\mathbf{b}$.

Example 6: Use the QR factorization of A for the matrix in Example 5 to solve $A\mathbf{x} = \begin{bmatrix} 1 \\ 4 \\ 4 \end{bmatrix}$.

$$Q^T\mathbf{b} = \begin{bmatrix} 3\sqrt{3} \\ 3/\sqrt{2} \\ 3/\sqrt{6} \end{bmatrix}$$

$$\left[\begin{array}{ccc|c} \sqrt{3} & \sqrt{3} & \sqrt{3} & 3\sqrt{3} \\ 0 & \sqrt{2} & 1/\sqrt{2} & 3/\sqrt{2} \\ 0 & 0 & 3/\sqrt{6} & 3/\sqrt{6} \end{array} \right] \Rightarrow \begin{array}{l} x_1 = (3\sqrt{3} - \sqrt{3} - \sqrt{3})/\sqrt{3} = 1 \\ x_2 = (3/\sqrt{2} - 1/\sqrt{2}(1))/\sqrt{2} = 1 \\ x_3 = 1 \end{array} \Rightarrow \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

In []: