Chapter 4: Orthogonality

4.2. Projections

Proposition. *Let A be a matrix. Then $\boldsymbol{N}(A^TA) = \boldsymbol{N}(A)$.*

Proof. If $A\boldsymbol{x} = \boldsymbol{0}$, then of course $A^TA\boldsymbol{x} = \boldsymbol{0}$. Conversely, if $A^TA\boldsymbol{x} = \boldsymbol{0}$, then

$$0 = \boldsymbol{x} \cdot (A^TA\boldsymbol{x}) = \boldsymbol{x}^TA^TA\boldsymbol{x} = (A\boldsymbol{x})^TA\boldsymbol{x} = (A\boldsymbol{x}) \cdot (A\boldsymbol{x}) \ ,$$

so $A\boldsymbol{x} = \boldsymbol{0}$ also. □

Corollary. *If the columns of the matrix A are independent, then A^TA is invertible.*

Proof. Suppose A is $m \times n$, so that its rank is $r(\leq m$, of course). Then A^TA is an $n \times n$ matrix, and $\boldsymbol{N}(A^TA) = \boldsymbol{N}(A)$ consists only the zero vector, so it is invertible. □

Take a subspace V of $\mathbb{R}^m$. We want to figure out how to find the orthogonal projection of each vector $\boldsymbol{b}$ in $\mathbb{R}^m$ onto V, i.e., the closest point $\boldsymbol{p}$ in V to $\boldsymbol{b}$. (That the closest $\boldsymbol{p}$ to $\boldsymbol{b}$ really makes $\boldsymbol{b} - \boldsymbol{p}$ perpendicular to V is clear by geometry, but we can also do it as a miminization problem in multivariate calc — see the next section.) We can see from geometry that the function that assigns to each $\boldsymbol{b}$ its projection $\boldsymbol{p}$ is a linear transformation — it respects addition and scalar multiplication — so there is an $m \times m$ matrix P for which $P\boldsymbol{b} = \boldsymbol{p}$ for each $\boldsymbol{b}$. [WARNING: In the display on page 210, Strang has a typo: P is $m \times m$, not $n \times n$.] To find this P, it's enough to find a matrix that does what we want to an invertible matrix B (because if $PB = MB$, then we can multiply on the right by B^{-1} and say $P = M$.

Choose a basis for V and use it as the columns of a matrix A; suppose $\dim(V) = n$, so that A is $m \times n$. We claim that the projection matrix P that we want is

$$P = A(A^TA)^{-1}A^T \ .$$

This product at least does make sense: Because the columns of A are a basis for V, they are independent, so A^TA is an invertible $n \times n$ matrix; and the dimensions of the factors are right for multiplication. Now let's form a basis for $\mathbb{R}^m$ starting with the columns of A and adding in a basis $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{m-n}$ for the orthogonal complement $V^\perp$ of V. The projection should leave the columns of A unchanged and turn the $\boldsymbol{c}$'s to $\boldsymbol{0}$:

$$\begin{bmatrix} A & \boldsymbol{c}_1 & \ldots & \boldsymbol{c}_{m-n} \end{bmatrix} \overset{P}{\to} \begin{bmatrix} A & O_{m\times(m-n)} \end{bmatrix} \ ;$$

so let's check that the P given by the (ugly) formula above does that:

Multiplying the columns of A by A^T gives A^TA (a matrix of dot products of the columns of A with each other, but we don't need that right now), and multiplying the $\boldsymbol{c}$'s by A^T gives columns of 0's because the $\boldsymbol{c}$'s were orthogonal to the columns of A. So multiplying by the factors in the product gives the sequence of block matrices

$$\begin{aligned} \begin{bmatrix} A & \boldsymbol{c}_1 & \ldots & \boldsymbol{c}_{m-n} \end{bmatrix} &\overset{A^T}{\to} \begin{bmatrix} A^TA & \boldsymbol{0} & \ldots & \boldsymbol{0} \end{bmatrix} \overset{(A^TA)^{-1}}{\to} \begin{bmatrix} I_n & O_{n\times(m-n)} \end{bmatrix} \\ &\overset{A}{\to} \begin{bmatrix} A & O_{m\times(m-n)} \end{bmatrix} \ . \end{aligned}$$

So $A(A^TA)^{-1}A^T$ is the P we want.

Notation: Strang consistently uses $\boldsymbol{b}$ for a general vector $\mathbb{R}^m$, A for the matrix whose columns are a basis for the subspace onto which we are projecting, $\boldsymbol{p}$ for the projection of $\boldsymbol{b}$ onto that subspace and $\boldsymbol{e}$ for the "error vector" $\boldsymbol{b}-\boldsymbol{p}$ (orthogonal to V). He also uses $\widehat{\boldsymbol{x}} = (x_1, \ldots, x_n)$ for the column of coefficients that gives $\boldsymbol{p}$ in terms of the columns of A, i.e., $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$. So, because we know $A\widehat{\boldsymbol{x}} = P\boldsymbol{b} = A(A^TA)^{-1}A^T\boldsymbol{b}$, and that A can be cancelled from the left (its nullspace is just the 0 vector), we see that $\widehat{\boldsymbol{x}} = (A^TA)^{-1}A^T\boldsymbol{b}$, i.e., $\widehat{\boldsymbol{x}}$ is the solution to $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$.

Example. We want to find the orthogonal projection $\boldsymbol{p}$ of $\boldsymbol{b} = (2,0,3)$ on the subspace of $\mathbb{R}^3$ with basis $(1,1,1)$ and $(1,0,-1)$, and to find the projection matrix, and to verify that $\boldsymbol{e} = \boldsymbol{b}-\boldsymbol{p}$ is orthogonal to the subspace:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix}, \qquad (A^TA)^{-1} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix}$$

$$P = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 5/6 & 1/3 & -1/6 \\ 1/3 & 1/3 & 1/3 \\ -1/6 & 1/3 & 5/6 \end{bmatrix},$$

$$\boldsymbol{p} = P\begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 7/6 \\ 5/3 \\ 13/6 \end{bmatrix}, \qquad A^T(\boldsymbol{b}-\boldsymbol{p}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 5/6 \\ -5/3 \\ 5/6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Special Case (silly): A is invertible, i.e., $V = \mathbb{R}^m$: The closest point in V to any $\boldsymbol{b}$ in $\mathbb{R}^m$ is $\boldsymbol{b}$ itself, so $\boldsymbol{p} = \boldsymbol{b}$, $P = I$ and $\widehat{\boldsymbol{x}} = A^{-1}\boldsymbol{b}$.

Special Case (not silly): A is a single nonzero vector $\boldsymbol{a}$, i.e., V is a line: Then $A^TA = \boldsymbol{a}\cdot\boldsymbol{a}$ is a nonzero number, so its inverse is just its reciprocal, and

$$P = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}\cdot\boldsymbol{a}}.$$

Moreover,

$$\boldsymbol{p} = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}\cdot\boldsymbol{a}}\boldsymbol{b} = \frac{\boldsymbol{a}(\boldsymbol{a}^T\boldsymbol{b})}{\boldsymbol{a}\cdot\boldsymbol{a}} = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\boldsymbol{a}\cdot\boldsymbol{a}}\boldsymbol{a}, \quad \text{so} \quad \widehat{\boldsymbol{x}} = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\boldsymbol{a}\cdot\boldsymbol{a}},$$

because in this case $\widehat{\boldsymbol{x}}$ is a single number, the coefficient of $\boldsymbol{a}$ when writing $\boldsymbol{p}$.

Example. We want to find the projection $\boldsymbol{p}$ of $\boldsymbol{b} = (1,2,3)$ onto (the span of) $\boldsymbol{a} = (1,0,-1)$ and the corresponding $\widehat{\boldsymbol{x}}$, and to check that $\boldsymbol{e} = \boldsymbol{b}-\boldsymbol{p}$ is orthogonal to $\boldsymbol{a}$:

$$P = \frac{1}{1+0+1}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\begin{bmatrix} 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & -1/2 \\ 0 & 0 & 0 \\ -1/2 & 0 & 1/2 \end{bmatrix}, \quad \boldsymbol{p} = P\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix},$$

$$\widehat{\boldsymbol{x}} = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\boldsymbol{a}\cdot\boldsymbol{a}} = \frac{1+0+(-3)}{1+0+1} = -1, \qquad \boldsymbol{a}\cdot\boldsymbol{e} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\cdot\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} = 0$$

Note that any projection matrix P, onto the subspace V, say, multiplies all the vectors in $\mathbb{R}^m$ into V, and if we multiply again by P, they don't change. Thus,

$$P^2\boldsymbol{b} = P\boldsymbol{b} \text{ for every } \boldsymbol{b} \text{ in } \mathbb{R}^m \;, \quad \text{i.e.,} \quad P^2 = P \;.$$

The term for a matrix that equals its own square is *idempotent.* There is more about idempotent matrices in the exercises.

4.3. Least Squares Approximation

In this section the objective is to find the n-dimensional subspace of $\mathbb{R}^{n+1}$ that best approximates a set of m points in $\mathbb{R}^m$, with "best" being measured as "minimizing the sum of the squares of the vertical distances from the points to the subspace", where vertical means in the direction of the axis for the last variable. This is the *least squares approximation* to the points. It seemed to me that this question, with vertical distances, was basically different from the last section, which measures orthogonal distances; but by writing both problems in neutral variables (so that I can keep straight which are the constants and which are the variables), we can see they are the same: Sticking to 3-dimensional space, so that we can visualize things, we can write

The projection problem: Given vectors $(p_1, p_2, p_3), (q_1, q_2, q_3), (r_1, r_2, r_3)$ in $\mathbb{R}^3$, find the point $(\hat{r}_1, \hat{r}_2, \hat{r}_3) = g(p_1, p_2, p_3) + h(q_1, q_2, q_3)$ closest to (r_1, r_2, r_3), i.e., minimizing

$$\begin{aligned} E &= (r_1 - \hat{r}_1)^2 + (r_2 - \hat{r}_2)^2 + (r_3 - \hat{r}_3)^2 \\ &= (r_1 - (gp_1 + hq_1))^2 + (r_2 - (gp_2 + hq_2))^2 + (r_3 - (gp_3 + hq_3))^2 \;. \end{aligned}$$

In our earlier notation $\boldsymbol{r} = \boldsymbol{b}$, $\boldsymbol{p}$ and $\boldsymbol{q}$ are the columns of A, and we seek $\widehat{\boldsymbol{r}} = \boldsymbol{p}$ and $\widehat{\boldsymbol{x}} = (g, h)$.

The least squares problem: Given points $(p_1, q_1, r_1), (p_2, q_2, r_2), (p_3, q_3, r_3)$ in $\mathbb{R}^3$, with r on the vertical axis, find the subspace, given by the equation $\hat{r} = gp + hq$, that minimizes the sum of the squares of the vertical distances from the given points to the corresponding points on the subspace, i.e., minimizing

$$\begin{aligned} E &= (r_1 - \hat{r}_1)^2 + (r_2 - \hat{r}_2)^2 + (r_3 - \hat{r}_3)^2 \\ &= (r_1 - (gp_1 + hq_1))^2 + (r_2 - (gp_2 + hq_2))^2 + (r_3 - (gp_3 + hq_3))^2 \;. \end{aligned}$$

The least squares problem, though, arises in a different way: Given data values with n "input" (or "explanatory", or "predictor") variables $x_1, x_2, \ldots, x_n$ and one "output" (or "response") variable y, find the coefficients c_i of the linear equation

$$\hat{y} = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

that best approximates the data. From this point of view, the x-values of the data points become the constants, the entries in the matrix X (which used to be A), the y-values of the data points become $\boldsymbol{b}$, the c's become the coefficients for writing the approximation $\widehat{\boldsymbol{y}}$ for $\boldsymbol{y}$ in terms of x-values of the data points — in other words, what $\widehat{\boldsymbol{x}}$ was in the other version of the problem (ARGH!).

So this time we aren't interested as much in what corresponds to $\boldsymbol{p}$ or P as what corresponds to $\widehat{\boldsymbol{x}}$, i.e., the solution $\boldsymbol{c}$ to $X^TX\boldsymbol{c} = \boldsymbol{y}$, where

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix}$$

At least, m, the number of data points, is usually much larger than n, the number of variables, and we can often control those values in an experiment, so the columns of X are usually independent. Thus, X^TX is usually invertible, and we can solve for $\boldsymbol{c} = (X^TX)^{-1}X^T\boldsymbol{y}$.

Example. Find the plane $\hat{y} = cx_1 + dx_2$ that is the least squares approximation to the data points (0,1,4), (1,0,5), (1,1,8), (-1,2,7):

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \\ -1 & 2 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} 4 \\ 5 \\ 8 \\ 7 \end{bmatrix}, \qquad (X^TX)^{-1} = \begin{bmatrix} 3 & -1 \\ -1 & 6 \end{bmatrix}^{-1} = (1/17)\begin{bmatrix} 6 & 1 \\ 1 & 3 \end{bmatrix}$$

$$\boldsymbol{c} = (X^TX)^{-1}X^T\boldsymbol{y} = (1/17)\begin{bmatrix} 6 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 6 \\ 26 \end{bmatrix} = \begin{bmatrix} 62/17 \\ 84/17 \end{bmatrix};$$

so the best approximation to these points is $\hat{y} = (62/17)x_1 + (84/17)x_2$.

Example. Find the line $\hat{y} = mx + b$ in $\mathbb{R}^2$ that is the least squares approximation to (i.e., is the *regression line* for) the data points (0,1), (1,2), (2,5), (3,6):

There seems to be only one input variable here, but we can create a second having the value 1 at each data point, to take care of the constant term b that means the line is not a subspace:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 5 \\ 6 \end{bmatrix}, \qquad (X^TX)^{-1} = \begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} 0.2 & -0.3 \\ -0.3 & 0.7 \end{bmatrix}$$

$$\boldsymbol{c} = (X^TX)^{-1}X^T\boldsymbol{y} = \begin{bmatrix} 0.2 & -0.3 \\ -0.3 & 0.7 \end{bmatrix}\begin{bmatrix} 42 \\ 24 \end{bmatrix} = \begin{bmatrix} 1.8 \\ 0.8 \end{bmatrix};$$

so the best approximation to these points is $\hat{y} = 1.8x + 0.8$.

Example. The heights h in feet of a ball t seconds after it is thrown upward are (0,0.2), (0.2,3.0), (0.4,5.3), (0.6,7.0), (0.8,7.5), (1.0,5.5). Of course the height should be related to the time by an equation of the form $h = a + bt + ct^2$ where c is negative. What are best approximations for the values of a, b, c? These are the initial height, initial velocity, and half the gravitational constant respectively — we don't know that the ball was thrown on Earth.

Besides the column of 1's, we also add a column of t^2's. Then we use `R` again:

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0.2 & 0.04 \\ 1 & 0.4 & 0.16 \\ 1 & 0.6 & 0.36 \\ 1 & 0.8 & 0.64 \\ 1 & 1 & 1 \end{bmatrix}, \qquad \boldsymbol{h} = \begin{bmatrix} 0.2 \\ 3.0 \\ 5.3 \\ 7.0 \\ 7.5 \\ 5.5 \end{bmatrix}, \qquad (T^TT)^{-1} = \begin{bmatrix} 0.8214286 & -2.946429 & 2.232143 \\ -2.9464286 & 18.169643 & -16.741071 \\ 2.2321429 & -16.741071 & 16.741071 \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix} = (T^TT)^{-1}T^T\boldsymbol{h} = \begin{bmatrix} -0.08571429 \\ 19.88571429 \\ -13.92857143 \end{bmatrix}$$

Apparently the ball was not thrown on Earth; it was thrown out of a shallow hole on a smaller planet.

4.4. Orthogonal Bases and Gram-Schmidt

Vectors are *orthonormal* iff they are pairwise orthogonal and they all have length 1. If the columns of the matrix Q (traditionally that letter) and Q is $m \times n$ where $m \geq n$, then $Q^TQ = I_n$. In particular, if Q is $n \times n$ — and still has orthonormal columns — then Q is invertible with inverse its transpose; it is then called *orthogonal.* (Why not "orthonormal", I don't know, and apparently neither does Strang.)

Suppose we have a basis $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ for $\mathbb{R}^n$ consisting of orthonormal vectors. What does that do for us? Well, for any vector $\boldsymbol{v}$ in $\mathbb{R}^n$, suppose we have written $\boldsymbol{v}$ in terms of the $\boldsymbol{q}$'s:

$$\boldsymbol{v} = c_1 \boldsymbol{q}_1 + \cdots + c_n \boldsymbol{q}_n \ .$$

Then for each k from 1 to n:

$$\boldsymbol{q}_k^T \boldsymbol{v} = c_1 \boldsymbol{q}_k^T \boldsymbol{q}_1 + \cdots + c_n \boldsymbol{q}_k^T \boldsymbol{q}_n \qquad = c_1(0) + \cdots + c_k(1) + \cdots + c_n(0) = c_k \ .$$

So we can find the coefficients to write $\boldsymbol{v}$ in terms of $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ just by taking dot products with the $\boldsymbol{q}$'s

We want to start with a basis $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ for $\mathbb{R}^n$ and build a new basis $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ that consists of orthonormal vectors.

How do we get from the $\boldsymbol{a}$'s to the $\boldsymbol{q}$'s? Well, the first step is easy: We just "normalize" $\boldsymbol{a}_1$, i.e., give it length 1, by dividing by its length:

$$\boldsymbol{q}_1 = \boldsymbol{a}_1 / ||\boldsymbol{a}_1|| \ .$$

Then we have to replace $\boldsymbol{a}_2$ with one that is orthogonal to $\boldsymbol{q}_1$. By Section 4.2, we have a way to get one: project $\boldsymbol{a}_2$ onto $\boldsymbol{q}_1$, and subtract the result from $\boldsymbol{a}_2$. Then we normalize that result:

$$\boldsymbol{e}_2 = \boldsymbol{a}_2 - \frac{\boldsymbol{q}_1^T \boldsymbol{a}_2}{\boldsymbol{q}_1^T \boldsymbol{q}_1} \boldsymbol{q}_1 = \boldsymbol{a}_2 - (\boldsymbol{q}_1^T \boldsymbol{a}_2) \boldsymbol{q}_1 \ , \qquad \boldsymbol{q}_2 = \boldsymbol{e}_2 / ||\boldsymbol{e}_2|| \ ;$$

the projection formula is simpler because $\boldsymbol{q}_1$ have length 1. Then the next step: subtract from $\boldsymbol{a}_3$ its projections on $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ (which gives something orthogonal to the first two $\boldsymbol{q}$'s), and normalize the result.

$$\boldsymbol{e}_3 = \boldsymbol{a}_3 - (\boldsymbol{q}_1^T \boldsymbol{a}_3) \boldsymbol{q}_1 - (\boldsymbol{q}_2^T \boldsymbol{a}_3) \boldsymbol{q}_2 \ , \qquad \boldsymbol{q}_3 = \boldsymbol{e}_3 / ||\boldsymbol{e}_3|| \ .$$

And so on.

And what is the process of going back from the $\boldsymbol{q}$'s to the $\boldsymbol{a}$'s? Using the dot product fact that we noted earlier, we get

$$\boldsymbol{a}_k = \sum_{i=1}^{n} (\boldsymbol{q}_i^T \boldsymbol{a}_k) \boldsymbol{q}_i \ .$$

If we write this in terms of matrices, we get

$$\begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^T \boldsymbol{a}_1 & \boldsymbol{q}_1^T \boldsymbol{a}_1 & \dots & \boldsymbol{q}_1^T \boldsymbol{a}_1 \\ \boldsymbol{q}_2^T \boldsymbol{a}_1 & \boldsymbol{q}_2^T \boldsymbol{a}_2 & \dots & \boldsymbol{q}_2^T \boldsymbol{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{q}_n^T \boldsymbol{a}_1 & \boldsymbol{q}_n^T \boldsymbol{a}_2 & \dots & \boldsymbol{q}_n^T \boldsymbol{a}_n \end{bmatrix} .$$

If we call A the matrix with columns the $\boldsymbol{a}$'s, Q the matrix with columns the $\boldsymbol{q}$'s, and R the matrix of dot products, then $A = QR$. Moreover, look at the entries $\boldsymbol{q}_i^T \boldsymbol{a}_j$ in the lower triangle of R, i.e., $i > j$. Then $\boldsymbol{a}_j$ can be written as a combination of the $\boldsymbol{q}$'s up to j; and $\boldsymbol{q}_i$ is orthogonal to the earlier $\boldsymbol{q}$'s, so it is orthogonal to $\boldsymbol{a}_j$. So the lower triangle is all 0's, i.e., R is upper triangular.

Example. The vector $(1, -2, 2)/3$ has length 1. We want to find an basis for $\mathbb{R}^3$ that includes it, and then write the corresponding A as QR where Q is orthogonal and R is upper triangular:

The given vector together with (0,1,0) and (0,0,1) is a basis for $\mathbb{R}^3$, so we can use this basis

$$\boldsymbol{a}_1 = \boldsymbol{q}_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1/3 & 0 & 0 \\ -2/3 & 1 & 0 \\ 2/3 & 0 & 1 \end{bmatrix}.$$

Then

$$\boldsymbol{e}_2 = \boldsymbol{a}_2 - (\boldsymbol{q}_1^T \boldsymbol{a}_2)\boldsymbol{q}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - (-2/3) \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 2/9 \\ 5/9 \\ 4/9 \end{bmatrix}, \quad \boldsymbol{q}_2 = \begin{bmatrix} 2/\sqrt{45} \\ 5/\sqrt{45} \\ 4/\sqrt{45} \end{bmatrix}$$

and

$$\boldsymbol{e}_3 = \boldsymbol{a}_3 - (\boldsymbol{q}_1^T \boldsymbol{a}_3)\boldsymbol{q}_1 - (\boldsymbol{q}_2^T \boldsymbol{a}_3)\boldsymbol{q}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - (2/3) \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix} - (4/\sqrt{45}) \begin{bmatrix} 2/\sqrt{45} \\ 5/\sqrt{45} \\ 4/\sqrt{45} \end{bmatrix}$$
$$= \begin{bmatrix} -2/5 \\ 0 \\ 1/5 \end{bmatrix}, \quad \boldsymbol{q}_3 = \begin{bmatrix} -2/\sqrt{5} \\ 0 \\ 1/\sqrt{5} \end{bmatrix}$$

We could check that the $\boldsymbol{q}$'s are pairwise orthogonal and each have length 1. We get

$$Q = \begin{bmatrix} 1/3 & 2/\sqrt{45} & -2/\sqrt{5} \\ -2/3 & 5/\sqrt{45} & 0 \\ 2/3 & 4/\sqrt{45} & 1/\sqrt{5} \end{bmatrix}, \quad R = \begin{bmatrix} 1 & -2/3 & 2/3 \\ 0 & 5/\sqrt{45} & 4/\sqrt{45} \\ 0 & 0 & 1/\sqrt{5} \end{bmatrix}.$$