Lectures - Week 6
Linear Least Squares, Orthonormal vectors and the QR Decomposition

Linear Least Squares Method

As an example of an application where these interconnections between the spaces is useful, we consider the linear least squares problem. Suppose we are given a set of data points $\{(x_i, f_i)\}$, $i = 1, \ldots, n$. These could be measurements from an experiment or obtained simply by evaluating a function at some points. Suppose further that we want to fit a polynomial to these data points in some way. As a concrete example suppose we have the three points specifically given by

$$(1, 2.2) \qquad (.8, 2.4) \qquad (0, 4.25)$$

and we expect them to lie in a line but due to experimental error, they don't. We would like to draw a line and have the line be the best representation of the points. If we only have two points then the line will pass through both points and so the error is zero at each point. However, if we have more than two data points, then we can't find a line that passes through the three points (unless they happen to be collinear) so we have to find a line which is a good approximation in some sense. Of course we need to define what we mean by a good representation. We could create an error vector of length n and each component measures the difference $|f_i - y(x_i)|$ where $y = a_1 x + a_0$ is the line we fit the data with. Then we can take a norm of this error vector and our goal is to find the line which minimizes this error vector. Of course we have to specify what norm we want to use. The linear least squares problem finds the line which minimizes this difference in the ℓ_2 (Euclidean) norm.

Example Consider the data points given above. Our ultimate goal is to find a line which fits the data in a linear least squares sense. For now, write the overdetermined system assuming our polynomial is $p_1(x) = a_0 + a_1 x$ and state when it has a solution.

Our equations are

$$\begin{aligned} a_0 + a_1(1) &= 2.1 \\ a_0 + a_1(.8) &= 2.5 \\ a_0 + a_1(0) &= 4.1 \end{aligned}$$

Writing this as a matrix problem $A\vec{x} = \vec{b}$ we have

$$\begin{pmatrix} 1 & 1 \\ 1 & 0.8 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 2.2 \\ 2.4 \\ 4.25 \end{pmatrix}$$

Now we know that this over-determined problem has a solution if the right hand side is in $\mathcal{R}(A)$ (i.e., is a linear combination of the columns of the coefficient matrix A). Here the rank of A is clearly 2 and thus not all of R^3. Moreover, $(2.2, 2.4, 4.25)^T$ is not in the $\mathcal{R}(A)$, i.e., not in the span$\{(1,1,1)^T,(1,0.8,0)^T\}$ and so the system doesn't have a solution. This just means that we can't find a line that passes through all three points.

Because we can't solve this over-determined problem $A\vec{x}=\vec{b}$ in general we find a $\vec{z}$ which is a good approximation. To this end, we look at the residual $\vec{b}-A\vec{z}$ and make this as small as possible; of course we must choose a norm. The linear least squares problem uses the Euclidean length, i.e., ℓ_2 norm. Looking at the ith component of the residual $\vec{b}-A\vec{z}$ we see that b_i is just the y coordinate of the data we are given and $(A\vec{z})_i$ is our equation evaluated at x_i so each component of the residual is the vertical difference at x_i, $f_i - y(x_i)$.

Example If our data had been

$$(1,2.1) \qquad (0.8,2.5) \qquad (0,4.1)$$

then would we have had a solution to the over-determined system?
Our matrix problem $A\vec{x}=\vec{b}$ is

$$\begin{pmatrix} 1 & 1 \\ 1 & 0.8 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 2.1 \\ 2.5 \\ 4.1 \end{pmatrix}$$

and we notice that in this case, the right hand side is in $\mathcal{R}(A)$ because

$$\begin{pmatrix} 2.1 \\ 2.5 \\ 4.1 \end{pmatrix} = 4.1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 2 \begin{pmatrix} 1 \\ 0.8 \\ 0 \end{pmatrix}$$

and thus the system is solvable and we have the line $4.1-2x$ which passes through all three points.

Consider the over-determined system $A\vec{x}=\vec{b}$ where A is $m\times n$ with $m>n$. The linear least squares problem is to

$$\text{find a vector } \vec{x} \text{ which minimizes the } \ell_2 \text{ norm of the residual} \quad \|\vec{b}-A\vec{z}\|_2 \quad \text{for all } \vec{z}\in\mathbf{R}^n$$

We note that minimizing the ℓ_2 norm of the residual is equivalent to minimizing its square. This often easier to work with because we avoid dealing with square roots. So we rewrite the problem as

$$\text{find a vector } \vec{x} \text{ which minimizes the square of the } \ell_2 \text{ norm } \|\vec{b}-A\vec{z}\|_2^2 \text{ for all } \vec{z}\in\mathbf{R}^n$$

Theorem The linear least squares problem always has a solution. It is unique if A has linearly independent columns. The solution of the problem can be found by solving the *normal equations*

$$A^T A\vec{y} = \vec{b}\,.$$

Before we prove this, lets look at the matrix $A^T A$. Clearly it is symmetric because

$$(A^T A)^T = A^T (A^T)^T = A^T A$$

but is it positive definite? Recall that a matrix B is positive definite if $\vec{x}^T B\vec{x} > 0$ for all $\vec{x} \neq 0$. We have

$$\vec{x}^T (A^T A)\vec{x} = (\vec{x}^T A^T)(A\vec{x}) = (A\vec{x})^T (A\vec{x}) = \vec{y}^T \vec{y} \qquad \text{where } \vec{y} = A\vec{x}$$

Now $y^T y$ is just the scalar (or inner or dot) product of a vector with itself and also it is the square of the Euclidean length of $\vec{y}$ which is always non-negative. It is only zero if $\vec{y} = \vec{0}$. Can $\vec{y}$ ever be zero? Remember that $y = A\vec{x}$ so if $\vec{x} \in \mathcal{N}(A)$ then $\vec{y} = \vec{0}$. When can the rectangular matrix A have something in the null space other than the zero vector? If we can take a linear combination of the columns of A (with coefficients nonzero) and get zero, i.e., if the columns of A are linearly dependent. Another way to say this is that if the columns of A are linearly independent, then $A^T A$ is positive definite; otherwise it is positive semi-definite (meaning that $x^T A^T Ax \geq 0$). Notice in our theorem we have that the solution is unique if A has linearly independent columns.

Proof First we show that the problem always has a solution. Recall that $\mathcal{R}(A)$ and $\mathcal{N}(A^T)$ are orthogonal complements in R^m. This tells us that we can write any vector in R^m as the sum of a vector in $\mathcal{R}(A)$ and one in $\mathcal{N}(A^T)$. To this end we write $\vec{b} = \vec{b}_1 + \vec{b}_2$ where $\vec{b}_1 \in \mathcal{R}(A)$ and $\vec{b}_2 \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$. Now we have

$$\vec{b} - A\vec{x} = \vec{b}_1 + \vec{b}_2 - A\vec{x} = \vec{b}_2$$

because $\vec{b}_1 \in \mathcal{R}(A)$ and so the equation $A\vec{x} = \vec{b}_1$ is always solvable. When we take the ℓ_2 norm we see that the residual is $\|\vec{b}_2\|_2$; we can never get rid of this unless $\vec{b} \in \mathcal{R}(A)$ entirely. So the problem is always solvable. When does $A\vec{x} = \vec{b}_1$ have a unique solution? It is unique when the columns of A are linearly independent. Lastly we must show that the way to find the solution $\vec{x}$ is by solving the normal equations. If we knew what $\vec{b}_1$ was, then we could simply solve $A\vec{x} = \vec{b}f_1$ but we don't know what the decomposition of $\vec{b} = \vec{b}_1 + \vec{b}_2$ is, simply that it is guaranteed to exist. To demonstrate that the $\vec{x}$ which minimizes $\|\vec{b} - A\vec{x}\|_2$ is found by solving $A^T A\vec{x} = A^T\vec{b}$ we first note that these normal equations can be written as $A^T(\vec{b} - A\vec{x}) = \vec{0}$ which is just A^T times the residual vector. From what we have already done we know that

$$A^T(\vec{b} - A\vec{x}) = A^T(\vec{b}_1 + \vec{b}_2 - A\vec{x}) = A^T(\vec{b}_2)$$

Recall that $\vec{b}_2 \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$ which means that $A^T\vec{b}_2 = \vec{0}$ and we have that

$$A^T(\vec{b} - A\vec{x}) = \vec{0} \Rightarrow A^T A\vec{x} = A^T\vec{b}$$

So the proof relied upon the fact that $\mathcal{R}(A)$ and $\mathcal{N}(A^T)$ are orthogonal complements and that this implies we can write any vector as the sum of a vector in $\mathcal{R}(A)$ and its orthogonal complement.

Example We return to our previous example and now determine the line which fits the data in the linear least squares sense; after we obtain the line we will compute the ℓ_2 norm of the residual.

We now know that to solve the linear least squares problem has a solution and in our case it is unique because A has linearly independent columns. All we have to do is form the normal equations and solve as usual. The normal equations

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0.8 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0.8 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0.8 & 0 \end{pmatrix} \begin{pmatrix} 2.2 \\ 2.4 \\ 4.25 \end{pmatrix}$$

are simplified as

$$\begin{pmatrix} 3.0 & 1.8 \\ 1.8 & 1.64 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 8.85 \\ 4.12 \end{pmatrix}$$

which has the solution () giving the line $y(x) = 4.225 - 2.125x$. If we calculate the residual vector we have

$$\begin{pmatrix} 2.2 - y(1) \\ 2.4 - y(0.8) \\ 4.25 - y(0) \end{pmatrix} = \begin{pmatrix} 0.1 \\ -0.125 \\ 0.025 \end{pmatrix}$$

which has an ℓ_2 norm of 0.162019.

We said that we only talked about the inverse of a square matrix. However, one can define a *pseudo-inverse* of a rectangular matrix. If A is an $m \times n$ matrix with linearly independent columns then a pseudo-inverse (or sometimes called left inverse of A) is $(A^T A)^{-1} A^T$. Note that

$$\left[(A^T A)^{-1} A^T \right] A = (A^T A)^{-1} (A^T A) = I$$

Example Find the quadratic polynomial which fits the following data in a linear least squares sense and calculate the ℓ_2 norm of the residual vector.

$$(0,0) \qquad (1,1) \qquad (3,2) \qquad (4,5)$$

In this case we seek a polynomial of the form $p(x) = a_0 + a_1 x + a_2 x^2$. Our over determined system is

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 5 \end{pmatrix}$$

so that the normal equations are

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 1 & 9 & 16 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 1 & 9 & 16 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \\ 5 \end{pmatrix}$$

leading to the square system

$$\begin{pmatrix} 4 & 8 & 26 \\ 8 & 26 & 92 \\ 26 & 92 & 338 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 8 \\ 27 \\ 99 \end{pmatrix}$$

Solving this we get $a_0 = 3/10$, $a_1 = 7/30$, $a_2 = 1/3$. Our residual vector is

$$\vec{r} = \begin{pmatrix} 0 - p(0) \\ 1 - p(1) \\ 2 - p(3) \\ 5 - p(4) \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.6 \\ 0.6 \\ 0.3 \end{pmatrix}$$

Orthonormal vectors

We know that two vectors, $\vec{x}$, $\vec{y}$, are *orthogonal* provided $\vec{x}^T\vec{y} = 0$. We give a special name to vectors which are orthogonal and have Euclidean length one.

Definition A set of vectors $\{\vec{v}_j\}_{j=1}^n$ are called *orthonormal* provided

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

Example Is there a connection between linear independence and orthogonal or orthonormal vectors?

We claim that a set of nonzero orthogonal vectors (and thus orthonormal vectors) are always linearly independent. To see this we will use proof by contradiction; that is, we will assume that the orthogonal vectors are linearly dependent and get a contradiction. Let $\{\vec{v}_j\}_{j=1}^n$ be a set of *orthogonal* vectors and suppose they are linearly dependent. Then there are constants c_i where not all $c_i = 0$ such that

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0}$$

The contradiction we will obtain is that $c_i = 0$ for all i. We now take $\vec{v}_j$ for arbitrary j and dot it into each side of this equation. We obtain

$$\vec{v}_j^T(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n) = \vec{v}_j^T\vec{0} \Rightarrow c_1\vec{v}_j^T\vec{v}_1 + c_2\vec{v}_j^T\vec{v}_2 + \cdots + c_n\vec{v}_j^T\vec{v}_n) = 0$$

But because the vectors are orthogonal all the terms but one disappear to give

$$c_j\vec{v}_j^T\vec{v}_j = 0 \Rightarrow c_j = 0$$

Because j was arbitrary it must hold for all j and thus all constants $c_j = 0$. Thus a set of orthogonal or orthonormal vectors are linearly independent.

Example Show that the following set of vectors form an orthonormal basis for R^3.

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{v}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$

Clearly each vector has Euclidean length one and if we take the dot product of any vector with another it is zero. We have

$$\vec{v}_1^T \vec{v}_2 = 1 \cdot 0 + 0(\frac{1}{\sqrt{2}} + 0(\frac{1}{\sqrt{2}} = 0 \qquad \vec{v}_1^T \vec{v}_3 = 0$$

and

$$\vec{v}_2^T \vec{v}_3 = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0$$

These vectors form a basis because there are three of them and they are linearly independent because they are orthonormal.

What characteristics does a square matrix have if its columns are orthonormal? Let U be an $m \times n$ matrix whose columns are orthonormal vectors $\vec{u}_j$ in R^m. Then $U^T U = I$ so that $U^{-1} = U^T$ and U is an orthogonal matrix. Consequently the columns of an orthogonal matrix are orthonormal.

An orthogonal matrix is very special. It preserves the Euclidean length of a vector, i.e., if Q is orthogonal

$$\|Q\vec{x}\|_2 = \|\vec{x}\|_2$$

To see this we write the ℓ_2 norm as an inner product and use the definition of an orthogonal matrix

$$\|Q\vec{x}\|_2^2 = (Q\vec{x})^T (Q\vec{x}) = \vec{x}^T Q^T Q \vec{x} = \vec{x}^T \vec{x} = \|\vec{x}\|_2^2$$

We will also see that $\mathcal{K}_2(Q) = 1$ which means it is the best conditioned matrix available.

Suppose we have a set of linearly independent vectors in R^n and we want to use them to generate a set of orthonormal vectors. There is a systematic way to do this called the Gram Schmidt Orthogonalization Method. Note that if we make the set of vectors orthogonal, then it is easy to make them orthonormal – we just divide each vector by its Euclidean length. For example we can make $\vec{v}^1 = (1,2)^T$ have length one by dividing each term by $\sqrt{5}$. To see how this method works we will first take a basis in R^2 and convert it to an orthonormal basis. Then we can see how this can be generalized to R^n.

Example The two vectors $\vec{v}_1 = (1,2)^T$ and $\vec{v}_2 = (2,3)^T$ in R^2 form a basis; use them to generate an orthonormal basis.

Our strategy will be to construct two orthogonal vectors, say $\vec{u}_1$, $\vec{u}_2$ (i.e., $\vec{u}_1^T \vec{u}_2 = 0$) from $\vec{v}_1, \vec{v}_2$ and then normalize them to get $\vec{e}_1, \vec{e}_2$. We will take $\vec{u}_1 = \vec{v}_1$ to start and then construct $\vec{u}_2$ using $\vec{u}_1$ and $\vec{v}_2$. Recall that in R^2 the projection of a vector $\vec{v}$ onto $\vec{u}$ denoted $\mathrm{proj}_{\vec{u}}\vec{v}$ is a vector given by dropping a perpendicular from $\vec{v}$ to $\vec{u}$ and computing the length of this line segment and then multiplying it by a unit vector in the direction of $\vec{u}$. (A unit vector in the direction of $\vec{u}$ is just $\vec{u}/\|\vec{u}\|_2$.) If θ is the angle between the two vectors then the length of this projection, denoted $\text{proj}_{\vec{u}}\vec{v}$ is found from

$$\cos\theta = \frac{\text{proj}_{\vec{u}}\vec{v}}{\|\vec{v}\|_2} \Rightarrow \text{proj}_{\vec{u}}\vec{v} = \|\vec{v}\|_2 \cos\theta$$

Note that $\vec{v}$ can be written as the sum of the projection vector plus a vector that is orthogonal to $\vec{u}$. Using the fact that $\vec{u}\cdot\vec{v} = \|\vec{u}\|_2\|\vec{v}\|_2\cos\theta$ we get that the length of the projection is given by

$$\text{proj}_{\vec{u}}\vec{v} = \|\vec{v}\|_2 \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|_2\|\vec{v}\|_2} = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|_2}\,.$$

Then a vector with this magnitude is found by multiplying it by the unit vector $\vec{u}/\|\vec{u}\|_2$ so we have

$$\vec{\text{proj}}_{\vec{u}}\vec{v} = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|_2}\frac{\vec{u}}{\|\vec{u}\|_2} = \frac{\vec{u}\cdot\vec{v}}{\vec{u}\cdot\vec{u}}\,\vec{u}$$

To create our orthonormal basis $\vec{e}_1, \vec{e}_2$ we set

$$\vec{u}_1 = \vec{v}_1 \qquad \text{and} \qquad \vec{e}_1 = \frac{\vec{u}_1}{\|\vec{u}_1\|_2}$$

$$\vec{u}_2 = \vec{v}_2 - \vec{\text{proj}}_{\vec{u}_1}\vec{v}_2 \qquad \text{and} \qquad \vec{e}_2 = \frac{\vec{u}_2}{\|\vec{u}_2\|_2}$$

Note that $\vec{u}_1, \vec{u}_2$ are orthogonal because we have subtracted off the part that is not orthogonal; this can be formally shown by

$$\vec{u}_1\cdot\vec{u}_2 = \vec{u}_1\cdot\vec{v}_2 - \vec{u}_1\cdot\Big(\frac{\vec{u}_1\cdot\vec{v}_2}{\vec{u}_1\cdot\vec{u}_1}\,\vec{u}_1\Big) = \vec{u}_1\cdot\vec{v}_2 - (\vec{u}_1\cdot\vec{v}_2)\frac{\vec{u}_1\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1} = 0$$

For our specific vectors we have

$$\vec{e}_1 = \frac{1}{\sqrt{5}}(1,2)^T = (\frac{1}{\sqrt{5}}, \frac{2}{5})^T$$

and

$$\vec{u}_2 = (2,3)^T - \frac{(1,2)^T(2,3)}{5}(1,2)^T = (2,3)^T - \frac{8}{5}(1,2)^T = (\frac{2}{5}, -\frac{1}{5})^T$$

and thus

$$\|(\frac{2}{5}, -\frac{1}{5})\| = \frac{1}{\sqrt{5}} \Rightarrow \vec{e}_2 = \frac{1}{\sqrt{5}}(\frac{2}{5}, -\frac{1}{5})^T$$

Clearly our vectors are orthogonal since

$$\frac{1}{\sqrt{5}}(1,2)^T\frac{1}{\sqrt{5}}(\frac{2}{5}, -\frac{1}{5})^T = 0$$

This algorithm can be extended to vectors in $\mathbf{R}^3$ easily by calculating $\vec{u}_1$, $\vec{u}_2$ as above and when calculating $\vec{u}_3$ we just subtract off the component of both $\vec{u}_1$ and $\vec{u}_2$ in the direction of $\vec{v}_3$, i.e.,

$$\vec{u}_3 = \vec{v}_3 - \vec{\text{proj}}_{\vec{u}_1}\vec{v}_3 - \vec{\text{proj}}_{\vec{u}_2}\vec{v}_3$$

Clearly this can be extended to $\mathbf{R}^n$ in an analogous manner.

QR Decompositions

We have already looked at the decomposition of a square matrix A as $A = LU$ where L was unit lower triangular and U was upper triangular. We now want to consider another decomposition where an orthogonal matrix is used and A can be rectangular. We first look at the case where A is square.

Theorem Let A be an $n \times n$ matrix with linearly independent columns. Then A can be uniquely written as

$$A = QR$$

where Q is an $n \times n$ orthogonal matrix and R is upper triangular $n \times n$ matrix with positive diagonal entries.

To prove this result we consider the $n \times n$ matrix $A^T A$ which we know is symmetric and because A has linearly independent columns it is also positive definite. From previous work we know that a symmetric positive definite matrix has a unique LL^T decomposition where L is lower triangular with positive diagonal entries. We first show that $Q = A(L^T)^{-1}$ is an orthogonal matrix. To this end, we must show that $Q^T Q = I$. Using the fact that $A^T A = LL^T$ we have

$$Q^T Q = (A(L^T)^{-1})^T (A(L^T)^{-1}) = L^{-1} A^T A (L^T)^{-1} = L^{-1} L L^T (L^T)^{-1} = I$$

So if A is a square matrix, then we take $R = L^T$ and because we have shown that $Q = A(L^T)^{-1} = AR^{-1}$ is an orthogonal matrix then we have $A = QR$. R has positive diagonal entries because L did.

To show that the decomposition is unique, we assume there are two such decompositions and get a contradiction. Let

$$A = Q_1 R_1 \qquad \text{and} \qquad A = Q_2 R_2$$

where $Q_1^T Q_1 = I$ and $Q_2^T Q_2 = I$ and $R_1 \neq R_2$. Now writing $A^T A$ with each of these two decompositions gives

$$A^T A = (Q_1 R_1)^T (Q_1 R_1) = R_1^T Q_1^T Q_1 R_1 = R_1^T R_1$$

and

$$A^T A = (Q_2 R_2)^T (Q_2 R_2) = R_2^T Q_2^T Q_2 R_2 = R_2^T R_2$$

Thus

$$A^T A = R_1^T R_1 = R_2^T R_2$$

But this says that there are two different LL^T decompositions of A^TA where each L has positive diagonal entries. However, we have proved that the LL^T decomposition is unique if we choose the signs of the diagonal entries. Consequently we have a contradiction and thus the decomposition is unique.

The proof of this theorem actually gives us a way to construct a QR decomposition of a matrix. We first form A^TA, do a Cholesky decomposition and thus have R and form $Q = AR^{-1}$. This can be done by hand, but is NOT a good approach computationally.

It turns out that an analogous result holds if A is an $m \times n$ matrix.

Theorem Let A be an $m \times n$ matrix with linearly independent columns. Then A can be uniquely written as

$$A = QR$$

where Q is an $m \times m$ orthogonal matrix and R is upper triangular $m \times n$ matrix with positive diagonal entries.

We note that because A is $m \times n$ and has linearly independent columns this says that $n \leq m$.

The QR decomposition can be used to solve a square linear system $A\vec{x} = \vec{b}$. We write

$$A\vec{x} = \vec{b}, A = QR \Rightarrow QR\vec{x} = \vec{b} \Rightarrow R\vec{x} = Q^T\vec{b}$$

Once we form $Q^T\vec{b}$ we are left with an upper triangular system to solve by a back solve. Computationally the QR decomposition is still $\mathcal{O}(n^3)$ for an $n \times n$ matrix but the coefficient of n^3 is larger than for LU decomposition so it is more expensive. However, because an orthogonal matrix is so well-conditioned, it is useful if you know that A is ill-conditioned.

The QR decomposition can be used to easily solve the linear least squares problem. Recall that the unique solution to the linear least squares problem when A has linearly independent columns is found from the normal equations

$$A^TA\vec{x} = A^T\vec{b} \Rightarrow \vec{x} = (A^TA)^{-1}A^T\vec{b}$$

Thus if we have that $A = QR$ we have

$$\vec{x} = (A^TA)^{-1}A^T\vec{b} = \left[(QR)^T(QR)\right]^{-1}(QR)^T\vec{b} = \left[R^TQ^TQR\right]^{-1}R^TQ^T\vec{b}$$

$$\Rightarrow \vec{x} = \left[R^TR\right]^{-1}R^TQ^T\vec{b} = R^{-1}R^{-T}R^TQ^T\vec{b} = R^{-1}Q^T\vec{b}$$

Thus once we have the decomposition $A = QR$ then we perform the matrix times vector multiplication $Q^T\vec{b}$ and then solve the upper triangular system

$$R\vec{x} = Q^T\vec{b}$$