QR DECOMPOSITION

MARIA CHRYSAFIS

Any real square matrix A can be decomposed into a product of an orthogonal matrix Q and an upper triangular matrix R; such a decomposition is called QR-decomposition. There are various ways to find Q and R, and in particular, the one I will talk about in this paper is called the Gram Schmidt algorithm. It is worth noting that the Gram Schmidt algorithm is not the only way to find suitable Q and R: we could've also used Householder transformations and Givens rotations.

1. Definitions

Before we begin discussing the Gram Schmidt decomposition, it is worth establishing some standard definitions that we will use later in the paper. The first is that of an *inner product*, for which we say that:

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^T \boldsymbol{B}.$$

The second definition is that of a *projection*, for which we denote:

$$\operatorname{proj}_{\boldsymbol{b}} \boldsymbol{a} = \frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} \mathbf{b}.$$

2. Gram Schmidt Decomposition

To perform the Gram Schmidt decomposition on an arbitrary $n \times n$ real square matrix $\boldsymbol{A}$, let the column vectors be $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_n$. Denote for $1 \leq k \leq n$:

$$\boldsymbol{u}_k = \boldsymbol{v}_k - \sum_{j=1}^{k-1} \operatorname{proj}_{\boldsymbol{u}_j} \boldsymbol{v}_k.$$

$$\boldsymbol{e}_k = \boldsymbol{u}_k \cdot \frac{1}{||\boldsymbol{u}_k||}.$$

We can let our matrix $\boldsymbol{Q}$ to be

$$\boldsymbol{Q} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}$$

Date: March 21, 2023.

and $\boldsymbol{R}$ to be

$$R = \begin{bmatrix} \langle \mathbf{e}_1, \mathbf{a}_1 \rangle & \langle \mathbf{e}_1, \mathbf{a}_2 \rangle & \langle \mathbf{e}_1, \mathbf{a}_3 \rangle & \cdots & \langle \mathbf{e}_1, \mathbf{a}_n \rangle \\ 0 & \langle \mathbf{e}_2, \mathbf{a}_2 \rangle & \langle \mathbf{e}_2, \mathbf{a}_3 \rangle & \cdots & \langle \mathbf{e}_2, \mathbf{a}_n \rangle \\ 0 & 0 & \langle \mathbf{e}_3, \mathbf{a}_3 \rangle & \cdots & \langle \mathbf{e}_3, \mathbf{a}_n \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \langle \mathbf{e}_n, \mathbf{a}_n \rangle \end{bmatrix}.$$

The central claim is that not only is $\boldsymbol{Q}$ an orthogonal matrix, but that moreover, $\boldsymbol{Q} \cdot \boldsymbol{R} = \boldsymbol{A}$. This is the crux of the Gram Schmidt decomposition algorithm. We will prove that it actually works later on in this paper.

Note. The Gram Schmidt decomposition algorithm also provides a way to numerically compute the determinant of a matrix. To see this, notice that if we make $\boldsymbol{Q}$ have strictly positive diagonal entries, then $\det(\boldsymbol{QR}) = \det(\boldsymbol{Q}) \det(\boldsymbol{R}) = \det(\boldsymbol{R})$ since $\boldsymbol{Q}$ has determinant 1 by virtue of being orthogonal. Since $\boldsymbol{R}$ is upper diagonal, its determinant is simply the product of its diagonal elements, which can easily be computed. The complexity of using this method to compute the determinant is $O(N^3)$, which is no better than the naive method.

For now, let's start with an example.

2.1. **Example.** Suppose we want to apply QR decomposition to the following matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Let us denote the column vectors

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Performing Gram–Schmidt, we see that:

$$\boldsymbol{u}_1 = \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

$$\boldsymbol{u}_2 = \boldsymbol{v}_2 - \operatorname{proj}_{\boldsymbol{u}_1} \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

$$\boldsymbol{u}_3 = \boldsymbol{v}_3 - \operatorname{proj}_{\boldsymbol{u}_1} \boldsymbol{v}_3 - \operatorname{proj}_{\boldsymbol{u}_2} \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{bmatrix}.$$

We can normalize $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$ into $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ after the fact:

$$\boldsymbol{e}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{e}_3 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ -\frac{\sqrt{2}}{2} \end{bmatrix}.$$

As such,

$$\boldsymbol{R} = \begin{bmatrix} \langle \boldsymbol{e_1}\ ,\ \boldsymbol{a_1} \rangle & \langle \boldsymbol{e_2}\ ,\ \boldsymbol{a_2} \rangle & \langle \boldsymbol{e_1}\ ,\ \boldsymbol{a_3} \rangle \\ 0 & \langle \boldsymbol{e_1}\ ,\ \boldsymbol{a_2} \rangle & \langle \boldsymbol{e_2}\ ,\ \boldsymbol{a_3} \rangle \\ 0 & 0 & \langle \boldsymbol{e_3}\ ,\ \boldsymbol{a_3} \rangle \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & 1 & 1 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

and

$$\boldsymbol{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}}. \end{bmatrix}$$

We can verify that $\boldsymbol{Q}$ is indeed orthogonal and that $\boldsymbol{R}$ is indeed upper triangular.

Thus, it does indeed work for the example before, but what about more generally?

Claim 2.1. *When $1 \leq i < j \leq n$, $\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = 0$. That is, $\boldsymbol{Q}$ is an orthogonal matrix.*

Proof. To show this, we can use induction on n, the size of our source matrix $\boldsymbol{A}$. When $n = 1$, it is trivially clear that the inductive hypothesis holds, so we can proceed to the case when $n > 1$:

$$\begin{aligned} \langle \boldsymbol{e}_i, \boldsymbol{e}_n \rangle &= \langle \boldsymbol{e}_i, \boldsymbol{v}_n - \sum_{k=1}^{n-1} \operatorname{proj}_{\boldsymbol{e}_k} \boldsymbol{v}_n \rangle \\ &= \langle \boldsymbol{e}_i, \boldsymbol{v}_n - \sum_{k=1}^{n-1} \mathbf{e}_k \cdot \frac{\langle \boldsymbol{e}_k\ ,\ \boldsymbol{v_n} \rangle}{||\boldsymbol{e}_k||} \rangle \\ &= \langle \boldsymbol{e}_i, \boldsymbol{v}_n \rangle - \langle \boldsymbol{e}_i, \sum_{k=1}^{n-1} \boldsymbol{e}_k \frac{\langle \boldsymbol{e}_k\ ,\ \boldsymbol{v_n} \rangle}{||\boldsymbol{e}_k||} \rangle \end{aligned}$$

We know by our inductive hypothesis that $\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = 0$ when $i \neq j$ and that $\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = 1$ when $i = j$. As such, we see that:

$$\begin{aligned} \langle \boldsymbol{e}_i, \boldsymbol{e}_n \rangle &= \langle \boldsymbol{e}_i, \boldsymbol{v}_n \rangle - \langle \boldsymbol{e}_i, \sum_{k=1}^{n-1} \boldsymbol{e}_k \frac{\langle \boldsymbol{e}_k\ ,\ \boldsymbol{v}_n \rangle}{||\boldsymbol{e}_k||} \rangle \\ &= \langle \boldsymbol{e}_i, \boldsymbol{v}_n \rangle - \langle \boldsymbol{e}_i, \boldsymbol{v}_n \rangle \\ &= 0, \end{aligned}$$

from which it follows that $\boldsymbol{Q}$ is indeed orthogonal. ∎

Theorem 2.2. *If $\boldsymbol{A}$ has nonzero determinant, then $\boldsymbol{A}$ has a unique QR decomposition with positive diagonal elements in $\boldsymbol{R}$.*

Proof. To see why this is the case, assume by sake of contradiction that $\boldsymbol{A} = \boldsymbol{Q_1 R_1} = \boldsymbol{Q_2 R_2}$ for distinct matrices $\boldsymbol{Q_1}$ and $\boldsymbol{Q_2}$. Then, suppose we define a matrix $\boldsymbol{M}$ such that:

$$\boldsymbol{M} = \boldsymbol{R_1}\boldsymbol{R_2}^{-1} = \boldsymbol{Q_1}^{-1}\boldsymbol{Q_2}.$$

Then, we can write:

$$\boldsymbol{A}\boldsymbol{A}^T = (\boldsymbol{Q_1 R_1})^T \boldsymbol{Q_1 R_1} = \boldsymbol{R_1}^T \boldsymbol{Q_1}^T \boldsymbol{Q_1 R_1} = \boldsymbol{R_1}^T \boldsymbol{R_1},$$

where we utilize the fact that $\boldsymbol{Q_1}^T\boldsymbol{Q_1} = \boldsymbol{I}$, since $\boldsymbol{Q_1}$ is orthogonal. However, we can also write

$$\boldsymbol{A}\boldsymbol{A}^T = (\boldsymbol{Q_2 R_2})^T \boldsymbol{Q_2 R_2} = \boldsymbol{R_2}^T \boldsymbol{Q_2}^T \boldsymbol{Q_2 R_2} = \boldsymbol{R_2}^T \boldsymbol{R_2},$$

where we again utilize the fact that $\boldsymbol{Q_2}$ is also orthogonal.

This would imply that the Cholesky decomposition isn't unique, since we can write $\boldsymbol{A}\boldsymbol{A}^T$ in two ways as a product of upper triangular matrices. Since this cannot be the case, our original assumption that $\boldsymbol{Q_1}$ and $\boldsymbol{Q_2}$ are unique must be mistaken. Thus, we must have a unique $\boldsymbol{QR}$ decomposition. ∎

3. Applications to Least Squares

The least squares problem asks us to find the $\boldsymbol{x}$ for which $||\boldsymbol{Ax} - \boldsymbol{b}||$ is minimized. In other words, we'd like to minimize

$$||\boldsymbol{Ax} - \boldsymbol{b}|| = (\boldsymbol{Ax} - \boldsymbol{b})^T(\boldsymbol{Ax} - \boldsymbol{b}) = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} - \boldsymbol{b}^T\boldsymbol{Ax} + \boldsymbol{b}^T\mathrm{b}.$$

To minimize this, we need the derivative with respect to $\boldsymbol{x}$ to be 0, so we need

$$2\boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{A}^T\boldsymbol{b} - \mathrm{b}^T\mathrm{A} = 0 \iff \boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{A}^T\boldsymbol{b} \iff \boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}.$$

Now, of course, we can always find $\boldsymbol{x}$ by finding the inverse of $\boldsymbol{A^T} \cdot \boldsymbol{A}$. However, we could also do it differently, using QR decomposition. If we let $\boldsymbol{A} = \boldsymbol{QR}$, then it becomes

$$(\boldsymbol{QR})^T\boldsymbol{QRx} = (\boldsymbol{QR})^T b \iff \boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{x} = (\boldsymbol{QR})^T\boldsymbol{b} \iff \boldsymbol{R}^T\boldsymbol{Rx} = \boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{b}.$$

Thus, we can solve

$$\boldsymbol{Rx} = \boldsymbol{Q}^T\boldsymbol{b}$$

instead, which can be solved with back substitution, by virtue of $\boldsymbol{R}$ being a square matrix. This is guaranteed to yield the $\boldsymbol{x}$ for which $||\boldsymbol{Ax} - \boldsymbol{b}||$ is minimized without using matrix inverse.