QR and least square

Zecheng Zhang

March 3, 2023

1 QR factorization

We study $A \in \mathbb{R}^{m\times n}$ matrix with linearly independent columns. QR algorithm is a key algorithm in numerical linear algebra. We want to study the column space of A.

Recall the Gram–Schmidt process for producing an orthogonal or an orthonormal basis for any nonzero subspace of $\mathbb{R}^n$. Given a basis $\{x_1, ..., x_p\}$ for a nonzero subspace W, define

$$\begin{aligned} q_1 &= a_1/r_{11} \\ q_2 &= a_2/r_{22} - \frac{r_{12}}{r_{22}}q_1 \\ q_3 &= a_3/r_{33} - \frac{r_{13}}{r_{33}}q_1 - \frac{r_{23}}{r_{33}}q_2 \\ &\cdots \\ q_p &= a_p/r_{pp} - \frac{r_{1p}}{r_{pp}}a_1 - \frac{r_{2p}}{r_{pp}}q_2 - \frac{r_{(p-1)p}}{r_{pp}}q_{p-1}, \end{aligned}$$

where $r_{ij} = q_i^T a_j$ and $r_{jj} = \|a_j - \sum_{i=1}^{j} r_{ij}q_i\|$. Then $\{q_1, ..., q_p\}$ is an orthonormal basis for W, i.e., $span\{a_1, a_2, ..., a_p\} = span\{q_1, q_2, ..., q_p\}$.

Theorem 1.1. If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

Proof. Let $a_1, ..., a_n$ be columns of A. Perform Gram-Schmidt, we obtain $Q = [q_1, ..., q_n]$, which is an orthonormal set of vectors whose span is $col(A)$. For a_k, a_k is in $span\{a_1, ..., a_k\} = span\{q_1, ..., q_k\}$. That is there exists $r_{1k}, ..., r_{kk}$ such that $a_k = r_{1k}q_1 + ... + r_{kk}q_k + 0q_{k+1}...0q_n$. Without loss of generality, we assume $r_{kk} > 0$, otherwise multiply r_{kk} and q_k by -1 simultaneously. Denote $Q = [q_1, q_2, ..., q_n]$, $R = [r_1, ..., r_n]$ where $r_k = [r_{1k}, ..., r_{kk}, 0, ..., 0]^t \in \mathbb{R}^n$, recall the matrix multiplication we have $A = QR$. We now claim that R is upper triangular with a positive diagonal (easy to verify) and invertible. Recall $rank(QR) \leq min(rank(Q), rank(R))$. Since $rank(A) = n = rank(Q)$, this implies that $rank(R) = n$. □

When $m > n$, we can append $m - n$ columns to Q to make it a $m \times m$ unitary matrix $\tilde{Q}$. In this process, we will append $m - n$ 0 rows to R to obtain $\tilde{R}$. We call $A = \tilde{Q}\tilde{R}$ full QR of A.

2 Modified QR

The GS-QR algorithm is not numerically stable. For the moment, a stable algorithm is one that is not too sensitive to the effects of rounding off errors. The modified GS is the way to improve

Algorithm 1: Gram Schmidt

Data: $n \geq 0$

for $j = 1$ *to* n **do**
    $v_j = a_j$
    for $i = 1$ *to* $j - 1$ **do**
        $r_{ij} = q_i^t a_j$
        $v_j = v_j - r_{ij} q_i$
    $r_{jj} = \|v_j\|_2$
    $q_j = v_j / r_{jj}$

the stability of the QR algorithm. GS can be expressed as an orthogonal projection:

$$q_1 = \frac{P_1 a_1}{\|P_1 a_1\|}, q_2 = \frac{P_2 a_2}{\|P_2 a_2\|}, ..., q_n = \frac{P_n a_n}{\|P_n a_n\|}, \tag{1}$$

where $P_j \in \mathbb{R}^{m \times m}$ denotes the orthogonal projector onto space spanned by $\{q_1, ...q_{j-1}\}$.

For each j, the GS algorithm computes a single orthogonal projection of rank $m - (j - 1)$, $v_j = P_j a_j$. Recall that: $P_{\perp q}$ denotes the rank $m - 1$ orthogonal projection onto the space orthogonal to q. By the definition of P_j, we can verify (without proof here):

$$P_j = P_{\perp q_{j-1}} ... P_{\perp q_2} P_{\perp q_1}, \tag{2}$$

and $P_1 = I$. As a result,

$$v_j = P_j a_j = P_{\perp q_{j-1}} ... P_{\perp q_2} P_{\perp q_1} a_j. \tag{3}$$

Specifically,

$$\begin{aligned}
v_j^1 &= a_j, \\
v_j^2 &= P_{\perp q_1} v_j^1 = v_j^1 - q_1 q_1^t v_j^1, \\
v_j^3 &= P_{\perp q_2} v_j^2 = v_j^2 - q_2 q_2^t v_j^2, \\
&... \quad ... \\
v_j &= P_{\perp q_{j-1}} v_j^{j-1} = v_j^{j-1} - q_{j-1} q_{j-1}^t v_j^{j-1}.
\end{aligned}$$

We summarize the algorithm in 2.

Algorithm 2: Modified Gram Schmidt

for $i = 1$ *to* n **do**
    $v_j = a_j$
for $i = 1$ *to* n **do**
    $r_{ii} = \|v_i\|$
    $q_i = v_i / r_{ii}$
    for $j = i + 1$ *to* n **do**
        $r_{ij} = q_i^t v_j$
        $v_j = v_j - r_{ij} q_i$