Matrix Computation 2011.03.24

Lecture Notes 4: QR Factorization

Professor: Zhihua Zhang *Scribe:Yizhi Liu*

1 QR Factorization

1.1 Orthonormal Basis and QR Factorization

Consider $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, where $m \geq n$ and $\text{rank}(\boldsymbol{A}) = n$. We write $\boldsymbol{A}$ as $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n)$. Then

$$\text{span}(\boldsymbol{a}_1) \subseteq \text{span}(\boldsymbol{a}_1, \boldsymbol{a}_2) \subseteq \cdots \subseteq \text{span}(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$$

And related *orthonormal* basis such

$$\text{span}(\boldsymbol{q}_1) \subseteq \text{span}(\boldsymbol{q}_1, \boldsymbol{q}_2) \subseteq \cdots \subseteq \text{span}(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n)$$

With

$$\begin{aligned}
\boldsymbol{a}_1 &= r_{11}\boldsymbol{q}_1 \\
\boldsymbol{a}_2 &= r_{12}\boldsymbol{q}_1 + r_{22}\boldsymbol{q}_2 \\
&\vdots \\
\boldsymbol{a}_n &= r_{1n}\boldsymbol{q}_1 + \cdots + r_{nn}\boldsymbol{q}_n
\end{aligned} \tag{1}$$

we can define $\underset{m\times n}{\boldsymbol{Q}} = (\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n)$, $\underset{n\times n}{\boldsymbol{R}} = \begin{pmatrix} r_{11} & r_{12} & & \cdots & r_{1n} \\ 0 & r_{22} & \ddots & & \vdots \\ & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & \ddots & r_{(n-1)n} \\ 0 & \cdots & & 0 & r_{nn} \end{pmatrix}$, such that

$$\underset{m\times n}{\boldsymbol{A}} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n) = \underset{m\times n}{\boldsymbol{Q}}\,\underset{n\times n}{\boldsymbol{R}}$$

where $\boldsymbol{Q}$ is *column orthogonal, i.e.,* $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_n$.

1.2 Full QR Factorization

With standard result from introductory linear algebra, there exists $\tilde{\boldsymbol{Q}} \in \mathbb{R}^{m\times(m-n)}$ such that $\underset{m\times m}{\boldsymbol{Q}_{\text{full}}} = \begin{pmatrix} \boldsymbol{Q}, & \tilde{\boldsymbol{Q}} \end{pmatrix}$ is orthogonal. Write $\underset{m\times n}{\boldsymbol{R}_{\text{full}}} = \begin{pmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{pmatrix}$, we have *full* QR factorization

$$\boldsymbol{A} = \boldsymbol{Q}_{\text{full}}\boldsymbol{R}_{\text{full}}$$

1.3 Classical Gram-Schmidt (CGS)

With (1), we have $\boldsymbol{a}_1 = r_{11}\boldsymbol{q}_1 \;\Rightarrow\; \|\boldsymbol{a}_1\|_2 = |r_{11}|$. Let

$$r_{11} = \|\boldsymbol{a}_1\|_2 \qquad \boldsymbol{q}_1 = \frac{\boldsymbol{a}_1}{r_{11}}$$

Similarly, $\boldsymbol{a}_2 = r_{12}\boldsymbol{q}_1 + r_{22}\boldsymbol{q}_2 \;\Rightarrow\; \boldsymbol{q}_1^T\boldsymbol{a}_2 = |r_{12}|$. Let

$$r_{12} = \boldsymbol{q}_1^T\boldsymbol{a}_2 \qquad r_{22} = \|\boldsymbol{a}_2 - r_{12}\boldsymbol{q}_1\|_2 \qquad \boldsymbol{q}_2 = \frac{\boldsymbol{a}_2 - r_{12}\boldsymbol{q}_1}{r_{22}}$$

Notice that

$$\boldsymbol{a}_j = r_{1j}\boldsymbol{q}_1 + r_{2j}\boldsymbol{q}_2 + \cdots + r_{jj}\boldsymbol{q}_j \qquad \boldsymbol{q}_i^T\boldsymbol{a}_j = r_{ij}\ (i \neq j)$$

This leads to the *classical Gram-Schmidt* (CGS) algorithm for computing QR Factorization.

Input: $\boldsymbol{A}_{(m\times n)}$ with rank$(\boldsymbol{A}) = n$
Output: $\boldsymbol{Q}_{(m\times n)}$, $\boldsymbol{R}_{(n\times n)}$
$\boldsymbol{R}(1,1) = \|\boldsymbol{A}(:,1)\|_2$;
$\boldsymbol{Q}(:,1) = \boldsymbol{A}(:,1)/\boldsymbol{R}(1,1)$;
for $k = 2:n$ **do**
  $\boldsymbol{R}(1:k-1,k) = \boldsymbol{Q}(1:m,1:k-1)^T\boldsymbol{A}(1:m,k)$;
  $\boldsymbol{z} = \boldsymbol{A}(1:m,k) - \boldsymbol{Q}(1:m,1:k-1)\boldsymbol{R}(1:k-1,k)$;
  $\boldsymbol{R}(k,k) = \|\boldsymbol{z}\|_2$;
  $\boldsymbol{Q}(1:m,k) = \boldsymbol{z}/\boldsymbol{R}(k,k)$;
end
return $\boldsymbol{Q}$,$\boldsymbol{R}$;

Algorithm 1: Classical Gram-Schmidt

1.4 Existence of QR Factorization

Theorem 1.1 *Every $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ $(m \geq n)$ has a QR Factorization.*

Proof: If rank$(\boldsymbol{A}) = n$, then use CGS algorithm we get the conclusion directly.
Otherwise, if $q = \text{rank}(\boldsymbol{A}) < n$, we partition $\boldsymbol{A}$ as

$$\underset{m\times n}{\boldsymbol{A}} = \begin{pmatrix} \underset{m\times q}{\boldsymbol{A}_1}, & \underset{m\times(n-q)}{\boldsymbol{A}_2} \end{pmatrix}$$

Since $q = \text{rank}(\boldsymbol{A})$, the columns of $\boldsymbol{A}_1$ are the bases of ran$(\boldsymbol{A})$. Namely, $\boldsymbol{A}_2$ can be represented by $\boldsymbol{A}_1$. Write

$$\boldsymbol{A}_2 = \boldsymbol{A}_1\boldsymbol{X} = \boldsymbol{Q}_1\boldsymbol{R}_1\boldsymbol{X} \overset{\boldsymbol{R}_2=\boldsymbol{R}_1\boldsymbol{X}}{=\!=\!=} \boldsymbol{Q}_1\boldsymbol{R}_2$$

Then,

$$\begin{aligned} \boldsymbol{A} &= \begin{pmatrix} \boldsymbol{Q}_1\boldsymbol{R}_1, & \boldsymbol{Q}_1\boldsymbol{R}_2 \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{Q}_1, & \boldsymbol{Q}_1 \end{pmatrix}\begin{pmatrix} \boldsymbol{R}_1 & \boldsymbol{R}_2 \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\ &= \boldsymbol{Q}\boldsymbol{R} \qquad \square \end{aligned}$$

Theorem 1.2 *Every* $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ $(m \geq n)$ *of* full rank *has a unique* reduced *QR factorization* $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ *with* $r_{jj} > 0$.

1.5 Modified Gram-Schmidt (MGS)

Input: $\boldsymbol{A}_{(m\times n)}$ with rank$(\boldsymbol{A}) = n$
Output: $\boldsymbol{Q}_{(m\times n)}$, $\boldsymbol{R}_{(n\times n)}$
for $k = 1 : n$ **do**
    $\boldsymbol{R}(k,k) = \|\boldsymbol{A}(1:m,k)\|_2$;
    $\boldsymbol{Q}(1:m,k) = \boldsymbol{A}(1:m,k)/\boldsymbol{R}(k,k)$;
    for $j = k+1 : n$ **do**
        $\boldsymbol{R}(k,j) = \boldsymbol{Q}(1:m,k)^T\boldsymbol{A}(1:m,j)$;
        $\boldsymbol{A}(1:m,j) = \boldsymbol{A}(1:m,j) - \boldsymbol{Q}(1:m,k)\boldsymbol{R}(k,j)$;
    end
end
return $\boldsymbol{Q}$,$\boldsymbol{R}$;

Algorithm 2: Modified Gram-Schmidt

Remarks: In the kth step of MGS, the kth column of $\boldsymbol{Q}$ and the kth row of $\boldsymbol{R}$ are determined. Comparing with CGS, MGS is more robust.

Derivation of MGS:

$$\boldsymbol{A} = \sum_{i=1}^{n} \boldsymbol{q}_i \boldsymbol{r}_i^T$$
$$\Rightarrow \boldsymbol{A} - \sum_{i=1}^{k-1} \boldsymbol{q}_i \boldsymbol{r}_i^T = \sum_{i=1}^{n} \boldsymbol{q}_i \boldsymbol{r}_i^T - \sum_{i=1}^{k-1} \boldsymbol{q}_i \boldsymbol{r}_i^T = \begin{pmatrix} \underset{k-1}{\boldsymbol{0}}, & \underset{n-k+1}{\boldsymbol{A}^{(k)}} \end{pmatrix}$$

Write $\boldsymbol{A}^{(k)}$ as $\boldsymbol{A}^{(k)} = \begin{pmatrix} \underset{1}{\boldsymbol{z}}, & \underset{n-k}{\boldsymbol{B}} \end{pmatrix}$, we have $\boldsymbol{q}_k^T\boldsymbol{A}^{(k)} = \begin{pmatrix} \boldsymbol{q}_k^T\boldsymbol{z}, & \boldsymbol{q}_k^T\boldsymbol{B} \end{pmatrix}$. Thus,

$$\boldsymbol{q}_k^T(\boldsymbol{A} - \sum_{i=1}^{k-1} \boldsymbol{q}_i \boldsymbol{r}_i^T) = \begin{pmatrix} \boldsymbol{0}, & \boldsymbol{q}_k^T\boldsymbol{A}^{(k)} \end{pmatrix} = \begin{pmatrix} \underset{k-1}{\boldsymbol{0}}, & \underset{1}{\boldsymbol{q}_k^T\boldsymbol{z}}, & \underset{n-k}{\boldsymbol{q}_k^T\boldsymbol{B}} \end{pmatrix}$$

On the other hand,

$$\boldsymbol{q}_k^T(\boldsymbol{A} - \sum_{i=1}^{k-1} \boldsymbol{q}_i \boldsymbol{r}_i^T) = \boldsymbol{q}_k^T\boldsymbol{A} = \boldsymbol{q}_k^T \sum_{i=1}^{n} \boldsymbol{q}_i \boldsymbol{r}_i^T = \boldsymbol{r}_k^T$$

Thus,

$$\boldsymbol{r}_k^T = \begin{pmatrix} \underset{k-1}{\boldsymbol{0}}, & \underset{1}{\boldsymbol{q}_k^T\boldsymbol{z}}, & \underset{n-k}{\boldsymbol{q}_k^T\boldsymbol{B}} \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{q}_k = \frac{\boldsymbol{z}}{r_{kk}}$$

To derive the inner loop, notice that

$$
\begin{aligned}
\left(\mathbf{0}, \quad \boldsymbol{A}^{(k+1)}\right) &= \boldsymbol{A} - \sum_{i=1}^{k} \boldsymbol{q}_i \boldsymbol{r}_i^T = \boldsymbol{A} - \sum_{i=1}^{k-1} \boldsymbol{q}_i \boldsymbol{r}_i^T - \boldsymbol{q}_k \boldsymbol{r}_k^T \\
&= \left(\underset{k-1}{\mathbf{0}}, \quad \underset{n-k+1}{\boldsymbol{A}^{(k)}}\right) - \boldsymbol{q}_k \boldsymbol{r}_k^T \\
&= (\mathbf{0}, \quad \boldsymbol{z}, \quad \boldsymbol{B}) - \boldsymbol{q}_k \boldsymbol{r}_k^T \\
&= \left(\underset{k-1}{\mathbf{0}}, \quad \boldsymbol{z}, \quad \boldsymbol{B}\right) - \boldsymbol{q}_k \left(\underset{k-1}{\mathbf{0}}, \quad r_{kk}, \, r_{k,k+1}, \, \ldots, \, r_{kn}\right) \\
&= \left(\mathbf{0}, \quad \mathbf{0}, \quad \boldsymbol{B} - \boldsymbol{q}_k (r_{k,k+1}, \, \ldots, \, r_{kn})\right) \qquad \square
\end{aligned}
$$