School of Information Science and Technology
ShanghaiTech University

QR decompositions

Theorems and Algorithms

Sihang Xu
xush@shanghaitech.edu.cn

December 15, 2020

Agenda

Introduction

The QR decomposition of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is an orthonormal matrix Q and an upper triangular matrix $\mathbb{R}$ such that $\mathbf{A} = \mathbf{QR}$:

$$\begin{bmatrix} \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ \\ \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ 0 & & & r_{nn} \end{bmatrix}$$

Where $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_n]$ is orthogonal and $\mathbf{R}$ is upper triangular.

$$\begin{bmatrix} \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ \\ \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ 0 & & & r_{nn} \end{bmatrix}$$

One way of computing QR decomposition is to design a process to produce $\mathbf{q}_1, \ldots, \mathbf{q}_n$, such that

- $\mathbf{a}_k \in \text{span}(\mathbf{q}_1, \ldots, \mathbf{q}_k)$, for $k = 1, \ldots, n$

The other way of computing QR decomposition is to design a series of orthonormal matrix $\mathbf{Q}_1, \ldots, \mathbf{Q}_k$, such that $\mathbf{Q}_k \cdots \mathbf{Q}_1 \mathbf{A}$ is an upper triangular matrix R.

Gram-Schmidt Process

Given linearly independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$, (Classical) Gram-Schmidt process calculate $\mathbf{q}_1, \ldots, \mathbf{q}_n$ with the following steps:

- Let $\mathbf{q}_1 = \mathbf{a}_1 / ||\mathbf{a}_1||$.
- For $j = 2, \ldots, n$, let

$$\bar{\mathbf{q}}_j = \mathbf{a}_j - (\mathbf{a}_j^\top \mathbf{q}_1)\mathbf{q}_1 - \cdots - (\mathbf{a}_j^\top \mathbf{q}_{j-1})\mathbf{q}_{j-1}$$

  and

$$\mathbf{q}_j = \bar{\mathbf{q}}_j / ||\mathbf{q}_j||$$

- Then $\mathbf{q}_1, \ldots, \mathbf{q}_n$ is an orthonormal basis of $\text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$

We can validate $\mathbf{q}_1, \ldots, \mathbf{q}_n$ has property that

$$\text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_j) = \text{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j)$$

for all $j = 1, \ldots, n$.

We can proof the claim by induction.

- For $j = 1$, it's obvious that $\text{span}(\mathbf{a}_1) = \text{span}(\mathbf{q}_1)$.
- If for $j = k$, $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_k) = \text{span}(\mathbf{q}_1, \dots, \mathbf{q}_k)$, then for $j = k + 1$, recall the step of in the process,

$$\bar{\mathbf{q}}_{k+1} = \mathbf{a}_{k+1} - (\mathbf{a}_{k+1}^\top \mathbf{q}_1)\mathbf{q}_1 - \cdots - (\mathbf{a}_{k+1}^\top \mathbf{q}_k)\mathbf{q}_k$$

We have $\bar{\mathbf{q}}_{k+1} \in \text{span}(\mathbf{q}_1, \dots, \mathbf{q}_k, \mathbf{a}_{k+1}) = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_k, \mathbf{a}_{k+1})$, and $\mathbf{a}_{k+1} \in \text{span}(\mathbf{q}_1, \dots, \mathbf{q}_k, \bar{\mathbf{q}}_{k+1}) = \text{span}(\mathbf{q}_1, \dots, \mathbf{q}_k, \mathbf{q}_{k+1})$. Hence

$$\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_{k+1}) \subset \text{span}(\mathbf{q}_1, \dots, \mathbf{q}_{k+1})$$

$$\text{span}(\mathbf{q}_1, \dots, \mathbf{q}_{k+1}) \subset \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_{k+1})$$

Which means $\text{span}(\mathbf{q}_1, \dots, \mathbf{q}_{k+1}) = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_{k+1})$ and the proof is done.

For the property, we know if $\mathbf{q}_1, \ldots, \mathbf{q}_k$ is the output of Gram-Schmidt process with input $\mathbf{a}_1, \ldots, \mathbf{a}_n$, then we have

$$\mathbf{a}_k \in \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_k) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_k)$$

Let $r_{ij} = \mathbf{a}_i^T \mathbf{q}_j$ for $i \geq j$, and we can gain the QR decomposition:

$$\begin{bmatrix} \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ \\ \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ 0 & & & r_{nn} \end{bmatrix}$$

Modified Gram-Schmidt Process for QR

In the classical Gram-Schmidt process, we compute $\bar{\mathbf{q}}_k$ by

$$\bar{\mathbf{q}}_k = \mathbf{a}_k - (\mathbf{a}_k^\top \mathbf{q}_1)\mathbf{q}_1 - \cdots - (\mathbf{a}_k^\top \mathbf{q}_{k-1})\mathbf{q}_{k-1}$$

In Modified Gram-Schmidt process, we use some smaller steps instead:

$$\begin{aligned}
\tilde{q}_k^{(1)} &= \mathbf{a}_k - (\mathbf{q}_1^T \mathbf{a}_k)\mathbf{q}_1 \\
\tilde{q}_k^{(2)} &= \mathbf{a}_k - (\mathbf{q}_2^T \tilde{q}_k^{(1)})\mathbf{q}_2 \\
&\vdots \\
\tilde{q}_k^{(k-1)} &= \mathbf{a}_k - (\mathbf{q}_{k-1}^T \tilde{q}_k^{(1)})\mathbf{q}_{k-1}
\end{aligned}$$

Modified Gram-Schmidt process has greater numerical stability

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, the idea of Householder QR can be described as following: Design a series orthonormal matrix $\mathbf{Q}_1, \ldots, \mathbf{Q}_k$, such that $\mathbf{Q}_1\mathbf{A}, \mathbf{Q}_2\mathbf{Q}_1\mathbf{A}, \ldots$ has following forms:

$$\mathbf{Q}_1\mathbf{A} = \begin{bmatrix} \times & \times & \cdots & \times \\ 0 & \times & \cdots & \times \\ \vdots & \vdots & & \vdots \\ 0 & \times & \cdots & \times \end{bmatrix}, \mathbf{Q}_2\mathbf{Q}_1\mathbf{A} = \begin{bmatrix} \times & \times & \cdots & \times \\ 0 & \times & \cdots & \times \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \times \end{bmatrix}, \ldots$$

And finally $\mathbf{Q}_k \cdots \mathbf{Q}_1\mathbf{A}$ is a upper triangular matrix $\mathbf{R}$, and then let $\mathbf{Q} = (\mathbf{Q}_k \cdots \mathbf{Q}_1)^T = \mathbf{Q}_1^T \cdots \mathbf{Q}_k^T$, we obtain $\mathbf{A} = \mathbf{QR}$.

Householder Reflection

Let $\mathbf{v} \in \mathbb{R}^m, \mathbf{v} \neq 0$. Let

$$\mathbf{H} = \mathbf{I} - \frac{2}{||\mathbf{v}||}\mathbf{v}\mathbf{v}^T$$

Then H is a reflection matrix, which is orthogonal.

For a vector $\mathbf{x}$, it can be verified

$$\mathbf{v} = \mathbf{x} \mp ||\mathbf{x}||\mathbf{e}_1 \Rightarrow \mathbf{H}\mathbf{x} = \pm||x||_2\mathbf{e}_1$$

Let $\mathbf{v} = \mathbf{a}_1 \mp ||\mathbf{a}_1||\mathbf{e}_1 \Rightarrow$ and $\mathbf{H}_1$ be the corresponding Householder reflection matrix. Apply $\mathbf{H}_1$ on $\mathbf{A}$, we can gain $\mathbf{A}^{(1)}$ with desired form:

$$\mathbf{A}^{(1)} = \mathbf{H}_1\mathbf{A} = \begin{bmatrix} \times & \times & \cdots & \times \\ 0 & \times & \cdots & \times \\ \vdots & \vdots & & \vdots \\ 0 & \times & \cdots & \times \end{bmatrix},$$

Householder QR

The same things can be done on $\mathbf{a}_2, \dots, \mathbf{a}_n$, with slightly different step. Assume at step $k+1$, we have $\mathbf{a}_1^{(k)}, \mathbf{a}_2^{(k)}, \dots, \mathbf{a}_k^{(k)}$ are already become "upper-triangular".

We do the Householder reflection on $\mathbf{a}_{k+1,k+1:m}^{(k)}$, which is $\mathbf{a}_{k+1}^{(k)}$ with the first k row are truncated, and obtain a smaller matrix $\bar{\mathbf{H}}_{k+1} \in \mathbb{R}^{m-k\times m-k}$. Then we can calculate $\mathbf{H}_{k+1}$ by

$$\mathbf{H}_{k+1} = \begin{bmatrix} \mathbf{I}_{k\times k} & O_{k\times m-k} \\ O_{m-k\times k} & \bar{\mathbf{H}}_{k+1} \end{bmatrix}$$

And $\mathbf{H}_{k+1}$ will eliminate the lower part of $\mathbf{a}_{k+1}^{(k)}$:

$$\mathbf{H}_{k+1} \begin{bmatrix} \mathbf{R}_k & \mathbf{a}_{k+1,1:k}^{(k)} & \cdots \\ 0 & \mathbf{a}_{k+1,k+1}^{(k)} & \cdots \\ O & \mathbf{a}_{k+1,k+2:m}^{(k)} & \cdots \end{bmatrix} = \begin{bmatrix} \mathbf{R}_k & \mathbf{a}_{k+1,1:k}^{(k)} & \cdots \\ 0 & \mathbf{a}_{k+1,k+1}^{(k)} & \cdots \\ O & 0 & \cdots \end{bmatrix}$$

After applying $\mathbf{H}_1, \ldots, \mathbf{H}_n$, $\mathbf{A}^{(n)}$ is an upper triangular matrix. Then we have

$$(\mathbf{H}_n \ldots \mathbf{H}_1)\mathbf{A} = \mathbf{A}^{(n)} \Rightarrow \mathbf{A} = \mathbf{H}_1^T \ldots \mathbf{H}_n^T \mathbf{A}^{(n)}$$

Then $\mathbf{Q} = \mathbf{H}_1^T \ldots \mathbf{H}_n^T, \mathbf{R} = \mathbf{A}^{(n)}$ is exactly the QR decomposition of A.

Formally, Householder QR for $\mathbf{A} \in \mathbb{R}^{m \times n}$ has the following steps:

1. Let $\mathbf{A}^{(0)} = \mathbf{A}$.
2. For $k = 1, \ldots, n$
   - Calculate $\mathbf{H}_{k+1}$ from $\mathbf{A}^{(k)}$ by Householder reflection
   - Let $\mathbf{A}^{(k+1)} = \mathbf{H}_{k+1}\mathbf{A}^{(k)}$
3. Output $\mathbf{Q} = \mathbf{H}_1^T \ldots \mathbf{H}_n^T, \mathbf{R} = \mathbf{A}^{(n)}$

Introduction

The core idea of Givens QR is the same as Householder QR: using orthogonal transformations to eliminate the lower part entries to obtain an upper triangular **R**, and the inverse of the orthogonal transformations multiple up to **Q**.

In the Givens QR, rotations is used to eliminate entries, and its step is very similar to Gauss Elimination.

$$\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow{\mathbf{J}_{1,2}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow{\mathbf{J}_{1,3}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow{\mathbf{J}_{1,4}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$$

Definition

Define the Givens Rotation as:

$$
\mathbf{J}(i,k,\theta) = \begin{array}{c c} & \begin{array}{ccccc} & i & & k & \end{array} \\ \begin{array}{c} \\ i \\ \\ k \\ \\ \end{array} & \begin{bmatrix} 1 & & & & \\ & c & & s & \\ & & 1 & & \\ & -s & & c & \\ & & & & 1 \end{bmatrix} \end{array}
$$

where $c = \cos\theta$ and $s = \sin\theta$.

It's easy to validate that $\mathbf{J}(i,k,\theta)$ is orthogonal.

Givens Rotation on vectors

Consider $\mathbf{x} = [x_1, \ldots, x_n]^T$. Let $\mathbf{y} = \mathbf{J}(i, k, \theta)$, we can conclude that

- $y_j = x_j$ for $j \neq i, k$, the givens rotation leaves the irrelevant rows unchanged.
- $y_i = cx_i + sx_k$
- $y_k = -sx_i + cx_k$

Let $y_k = 0$, we have $\sin\theta x_i = \cos\theta x_k$, and solve the equation gives

$$\theta = \begin{cases} \arctan(x_k/x_i) & , x_i \neq 0 \\ \pi/2 \text{ or } 3\pi/2 & , x_i = 0 \end{cases}$$

Givens QR

And the process to compute Givens QR for a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ can be described as follows:

1. Let $\mathbf{Q} = \mathbf{I}, \mathbf{R} = \mathbf{A}$
2. For $i = 1, \ldots, n$
   For $k = i+1, \ldots, m$
   - Compute θ_{ik} from x_{ii} and x_{ik}
   - $\mathbf{Q} = \mathbf{Q} \times \mathbf{J}(i, k, \theta_{ik}), \mathbf{R} = \mathbf{J}(i, k, \theta_{ik})$
3. Output $\mathbf{Q}$ and $\mathbf{R}$

Thank you!