Givens Rotations

- Example: Let

$$\mathbf{G}(\theta) = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}$$

where $c = \cos(\theta), s = \sin(\theta)$ for some θ. Consider $\mathbf{y} = \mathbf{G}(\theta)\mathbf{x}$:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} cx_1 + sx_2 \\ -sx_1 + cx_2 \end{bmatrix}.$$

It can be verified that

- $\mathbf{G}(\theta)$ is orthogonal;
- $y_2 = 0$ if $\theta = \tan^{-1}(x_2/x_1)$, or equivalently if

$$c = \frac{x_1}{\sqrt{x_1^2 + x_2^2}}, \quad s = \frac{x_2}{\sqrt{x_1^2 + x_2^2}}.$$

Givens Rotations

- Givens rotation/transformation:

$$
\begin{array}{cc}
 & \begin{array}{ccccc} & i & & k & \\ & \downarrow & & \downarrow & \end{array} \\
\mathbf{G}(i,k,\theta) = & \begin{bmatrix} \mathbf{I} & & & & \\ & c & & s & \\ & & \mathbf{I} & & \\ & -s & & c & \\ & & & & \mathbf{I} \end{bmatrix} \begin{array}{c} \\ \leftarrow i \\ \\ \leftarrow k \\ \\ \end{array}
\end{array}
$$

where $c = \cos(\theta)$, $s = \sin(\theta)$.

- $\mathbf{G}(i,k,\theta)$ is orthogonal
- let $\mathbf{y} = \mathbf{G}(i,k,\theta)\mathbf{x}$. It holds that

$$
y_j = \begin{cases} cx_i + sx_k, & j = i \\ -sx_i + cx_k, & j = k \\ x_j, & j \neq i,k \end{cases}
$$

- y_k is forced to zero if we choose $\theta = \tan^{-1}(x_k/x_i)$.

Givens QR

- Example: consider a 4×3 matrix. Givens QR (from top to bottom) can be

$$\mathbf{A} = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{1,2}^{(1)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{1,3}^{(1)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{1,4}^{(1)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$$

$$\xrightarrow{\mathbf{G}_{2,3}^{(2)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{2,4}^{(2)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{3,4}^{(3)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{R}$$

Givens QR

or (from bottom to top)

$$\mathbf{A} = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{3,4}^{(1)}} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{2,3}^{(1)}} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{1,2}^{(1)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$$

$$\xrightarrow{\mathbf{G}_{3,4}^{(2)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{2,3}^{(2)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{\mathbf{G}_{3,4}^{(3)}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{R}$$

where $\mathbf{A} \xrightarrow{\mathbf{G}} \mathbf{B}$ means $\mathbf{B} = \mathbf{GA}$; $\mathbf{G}_{i,k}^{(j)} = \mathbf{G}^{(j)}(i,k,\theta)$, with θ chosen to zero out the (k,j)th entry of the matrix transformed by $\mathbf{G}_{i,k}^{(j)}$.

Givens QR

- **Givens QR:** assume $m \geq n$. Perform a sequence of Givens rotations to annihilate the lower triangular parts of $\mathbf{A}$ to obtain $\mathbf{R}$, say

$$\underbrace{(\mathbf{G}_{n,m}^{(n)} \cdots \mathbf{G}_{n,n+2}^{(n)} \mathbf{G}_{n,n+1}^{(n)}) \cdots (\mathbf{G}_{2m}^{(2)} \cdots \mathbf{G}_{24}^{(2)} \mathbf{G}_{23}^{(2)})(\mathbf{G}_{1m}^{(1)} \cdots \mathbf{G}_{13}^{(1)} \mathbf{G}_{12}^{(1)})}_{=\mathbf{Q}^T} \mathbf{A} = \mathbf{R}$$

  where $\mathbf{R}$ takes the upper triangular form, and $\mathbf{Q}$ is orthogonal.

- the Givens QR procedure is a process of "orthogonal triangularization"
- complexity (for $m \geq n$): $\mathcal{O}(3n^2(m - n/3))$ for $\mathbf{R}$ only
- not as efficient as Householder QR for general (and dense) $\mathbf{A}$'s
  - the flop count for Householder QR is $2n^2(m - n/3)$ (for $\mathbf{R}$ and for $m \geq n$)
  - the flop count for Givens QR is $3n^2(m - n/3)$
- can be faster than Householder QR if $\mathbf{A}$ has certain sparse structures and we exploit them

Method of Normal Equations vs. QR for LS

- In terms of complexity, method of normal equations only needs half of the arithmetic compared to QR decompostion when $m \gg n$.

- Method of normal equations can be easy for implementation, however, it is not recommended due to its numerical instability.
  - By forming the Gram matrix $\mathbf{A}^T\mathbf{A}$, we square the condition number of $\mathbf{A}$. (cf. SVD Topic)

- Thus, using the QR decomposition yields a better least-squares estimate than the normal equations in terms of solution quality because it avoids forming $\mathbf{A}^T\mathbf{A}$.

Solving Underdetermined Linear Systems by QR

For $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m < n$ and $\operatorname{rank}(\mathbf{A}) = m$, we have

$$\mathbf{A}^T = \mathbf{QR} = \begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix}\begin{bmatrix}\mathbf{R}_1 \\ \mathbf{0}\end{bmatrix} = \mathbf{Q}_1\mathbf{R}_1 + \mathbf{Q}_2\mathbf{0}$$

- note

  $$\mathbf{Ax} = \mathbf{R}_1^T\mathbf{Q}_1^T\mathbf{x} + \mathbf{0}^T\mathbf{Q}_2^T\mathbf{x} = \mathbf{b}$$

  which indicates $\mathbf{R}_1^T\mathbf{Q}_1^T\mathbf{x} = \mathbf{b}$ (solving via Linear System Topic) and $\mathbf{Q}_2^T\mathbf{x}$ can be anything, which we set to be $\mathbf{d}$. Then we have

  $$\begin{bmatrix}\mathbf{Q}_1^T\mathbf{x} \\ \mathbf{Q}_2^T\mathbf{x}\end{bmatrix} = \mathbf{Q}^T\mathbf{x} = \begin{bmatrix}\mathbf{R}_1^{-T}\mathbf{b} \\ \mathbf{d}\end{bmatrix}$$

- the solution is

  $$\mathbf{x} = \mathbf{Q}\begin{bmatrix}\mathbf{R}_1^{-T}\mathbf{b} \\ \mathbf{d}\end{bmatrix} = \mathbf{Q}_1\mathbf{R}_1^{-T}\mathbf{b} + \mathbf{Q}_2\mathbf{d}$$

  where to get the minimum norm solution, we can set $\mathbf{d} = \mathbf{0}$.

QR with Column Pivoting

- QR with column pivoting for $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(\mathbf{A}) = r \leq \min\{m, n\}$ (cf. Section 5.4.2 in **[Golub-Van Loan'13]**)

  $$\mathbf{AP} = \mathbf{QR} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \mathbf{Q}_1 \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \end{bmatrix}$$

  or

  $$\mathbf{A} = \mathbf{QRP}^T$$

  where $\mathbf{P}$ is a permutation matrix for the columns of $\mathbf{A}$ and $\mathbf{Q}$ is orthogonal
  - $\mathbf{Q}_1 \in \mathbb{R}^{m \times r}$, $\mathbf{R}_1 \in \mathbb{R}^{r \times r}$ with $r_{ii} > 0$ and $r_{11} \geq r_{22} \geq \cdots \geq r_{rr} > 0$
  - $\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{Q}_1)$, $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{R}) = \operatorname{rank}(\mathbf{R}_1)$
  - can be used to find the (numerical) rank of $\mathbf{A}$ at lower computational cost than a singular value decomposition
  - Gram-Schmidt with column pivoting, Householder QR with column pivoting

- more sophisticated pivoting schemes than QR with column pivoting are rank-revealing QR algorithms

Linear Systems: Solution via QR with Column Pivoting

- Problem: given *general* $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{y} \in \mathbb{R}^m$, determine
  - whether $\mathbf{y} = \mathbf{A}\mathbf{x}$ has a solution
  - what is the solution

- by QR with column pivoting, it can be shown that

$$
\begin{aligned}
\mathbf{y} = \mathbf{A}\mathbf{x} \quad &\Longleftrightarrow \quad \mathbf{y} = \mathbf{Q}_1 \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \end{bmatrix} \mathbf{P}^T\mathbf{x} \; (\text{define } \mathbf{P}^T\mathbf{x} = \mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix}) \\
&\Longleftrightarrow \quad \mathbf{Q}_1^T\mathbf{y} = \mathbf{R}_1\mathbf{z}_1 + \mathbf{R}_2\mathbf{z}_2, \; \mathbf{Q}_2^T\mathbf{y} = \mathbf{0} \\
&\Longleftrightarrow \quad \mathbf{z}_1 = \mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y} - \mathbf{R}_1^{-1}\mathbf{R}_2\mathbf{z}_2, \; \mathbf{Q}_2^T\mathbf{y} = \mathbf{0} \\
&\Longleftrightarrow \quad \begin{array}{l} \mathbf{x} = \mathbf{P}\mathbf{z} = \mathbf{P}\begin{bmatrix} \mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y} - \mathbf{R}_1^{-1}\mathbf{R}_2\mathbf{z}_2 \\ \mathbf{z}_2 \end{bmatrix}, \; \text{for any } \mathbf{z}_2 \in \mathbb{R}^{n-r}, \\ \mathbf{Q}_2^T\mathbf{y} = \mathbf{0} \end{array}
\end{aligned}
$$

- a linear system $\mathbf{y} = \mathbf{A}\mathbf{x}$ is said to be consistent if $\mathbf{Q}_2^T\mathbf{y} = \mathbf{0}$, i.e., $\mathbf{y} \in \mathcal{R}(\mathbf{A})$

Linear Systems: Solution via QR with Column Pivoting

- Case (a): full-column rank $\mathbf{A}$, i.e., $r = n \leq m$
  - there is no $\mathbf{R}_2$, and $\mathbf{Q}_2^T\mathbf{y} = \mathbf{0}$ is equivalent to $\mathbf{y} \in \mathcal{R}(\mathbf{Q}_1) = \mathcal{R}(\mathbf{A})$
  - Result: the linear system has a solution if and only if $\mathbf{y} \in \mathcal{R}(\mathbf{A})$, and the solution, if exists, is uniquely given by $\mathbf{x} = \mathbf{P}\mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y}$
- Case (b): full-row rank $\mathbf{A}$, i.e., $r = m \leq n$
  - there is no $\mathbf{Q}_2$
  - Result: the linear system always has a solution, and the solution is given by $\mathbf{x} = \mathbf{P}\begin{bmatrix} \mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y} - \mathbf{R}_1^{-1}\mathbf{R}_2\mathbf{z}_2 \\ \mathbf{z}_2 \end{bmatrix}$, for any $\mathbf{z}_2 \in \mathbb{R}^{n-r}$
- Case (c): square and full rank $\mathbf{A}$, i.e., $r = m = n$
  - there is no $\mathbf{R}_2$ and no $\mathbf{Q}_2$
  - Result: the linear system always has a solution, and the solution is given by $\mathbf{x} = \mathbf{P}\mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y}$

- we can find the minimum norm solution; $\|\mathbf{x}\|_2$ given by

$$\|\mathbf{x}\|_2^2 = \|\mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y} - \mathbf{R}_1^{-1}\mathbf{R}_2\mathbf{z}_2\|_2^2 + \|\mathbf{z}_2\|_2^2$$

is minimized when $\mathbf{z}_2 = (\mathbf{R}_2^T\mathbf{R}_1^{-T}\mathbf{R}_1^{-1}\mathbf{R}_2 + \mathbf{I})^{-1}\mathbf{R}_2^T\mathbf{R}_1^{-T}\mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y}$

Least Squares: Solution via QR with Column Pivoting

- consider the LS problem
$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2$$
for *general* $\mathbf{A}\in\mathbb{R}^{m\times n}$
- we have, for any $\mathbf{x}\in\mathbb{R}^n$,
$$\begin{aligned}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 = \|\mathbf{y}-\mathbf{Q}\mathbf{R}\mathbf{P}^T\mathbf{x}\|_2^2 &= \|\mathbf{Q}^T\mathbf{y}-\mathbf{R}\mathbf{P}^T\mathbf{x}\|_2^2 \\ &= \left\| \begin{bmatrix}\mathbf{Q}_1^T\\ \mathbf{Q}_2^T\end{bmatrix}\mathbf{y} - \begin{bmatrix}\mathbf{R}_1 & \mathbf{R}_2\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}^T\mathbf{x}\right\|_2^2 \text{ (define } \mathbf{P}^T\mathbf{x}=\mathbf{z}=\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix}) \\ &= \|\mathbf{Q}_1^T\mathbf{y}-\mathbf{R}_1\mathbf{z}_1-\mathbf{R}_2\mathbf{z}_2\|_2^2 + \|\mathbf{Q}_2^T\mathbf{y}\|_2^2 \\ &\geq \|\mathbf{Q}_2^T\mathbf{y}\|_2^2\end{aligned}$$
- the equality above is attained if $\mathbf{z}$ satisfies $\mathbf{Q}_1^T\mathbf{y}=\mathbf{R}_1\mathbf{z}_1+\mathbf{R}_2\mathbf{z}_2$, and that leads to an LS solution
$$\begin{aligned}\mathbf{Q}_1^T\mathbf{y}=\mathbf{R}_1\mathbf{z}_1+\mathbf{R}_2\mathbf{z}_2 \quad &\Longleftrightarrow \quad \mathbf{z}_1=\mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y}-\mathbf{R}_1^{-1}\mathbf{R}_2\mathbf{z}_2 \\ &\Longleftrightarrow \quad \mathbf{x}_{\mathsf{LS}}=\mathbf{P}\begin{bmatrix}\mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y}-\mathbf{R}_1^{-1}\mathbf{R}_2\mathbf{z}_2\\ \mathbf{z}_2\end{bmatrix}, \text{ for any } \mathbf{z}_2\in\mathbb{R}^{n-r}\end{aligned}$$
- it becomes $\mathbf{x}_{\mathsf{LS}}=\mathbf{P}\mathbf{R}_1^{-1}\mathbf{Q}_1^T\mathbf{y}$ for full-column rank $\mathbf{A}$

QR with Column Pivoting

- columns of $\mathbf{AP} = \mathbf{QR}$ are the columns of $\mathbf{A}$ in a different order
- the columns are divided in two groups:

$$\mathbf{AP} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} = \mathbf{Q}_1 \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \end{bmatrix}$$

  where
  - $\mathbf{A}_1 \in \mathbb{R}^{m \times r}$ with linearly independent columns and QR factorization $\mathbf{Q}_1\mathbf{R}_1$
  - $\mathbf{A}_1 \in \mathbb{R}^{m \times (n-r)}$ with columns that are linear combinations of columns of $\mathbf{A}_1$

$$\mathbf{A}_2 = \mathbf{Q}_2\mathbf{R}_2 = \mathbf{A}_1\mathbf{R}_1^{-1}\mathbf{R}_2$$

- the QR factorization with column pivoting provides two useful bases for $\mathcal{R}(\mathbf{A})$
  - columns of $\mathbf{Q}_1$ are an orthonormal basis
  - columns of $\mathbf{A}_1$ are a basis selected from the columns of $\mathbf{A}$

Other Contents on QR

- generalized QR decomposition for pair of matrices $(\mathbf{A}, \mathbf{B})$
- QR algorithm for computing eigenvalues (cf. Eigendecomposition Topic)
- QR algorithm for computing SVD (cf. SVD Topic)

References

[Golub-Van Loan'13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th edition, JHU Press, 2013.