St. Xavier's College (Autonomous), Kolkata

MSc in Data Science

MDTS4113

Linear Algebra

Notes on QR Factorization

Reduced QR Factorization

Consider an mxn matrix **A** = $[\underline{a}_1 \quad \underline{a}_2 \ldots \underline{a}_{n-1} \quad \underline{a}_n]$

successive spaces spanned by the columns of **A**:

$$\langle \underline{a}_1 \rangle \subseteq \langle \underline{a}_1, \quad \underline{a}_2 \rangle \subseteq \langle \underline{a}_1, \quad \underline{a}_2, \quad \underline{a}_3 \rangle \subseteq \ldots$$

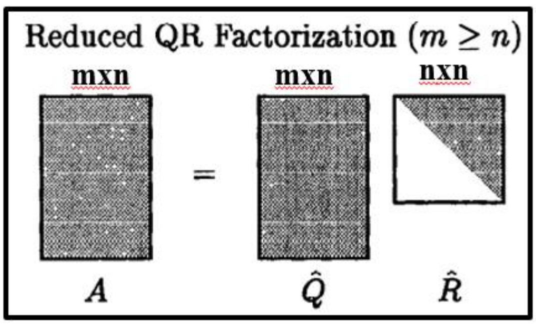
The idea of **QR factorization** is the construction of a sequence of orthonormal vectors $\underline{q}_1, \underline{q}_2, \ldots$ that span these successive spaces.

Assume that A $\epsilon\, \boldsymbol{R}^{mxn}$(m > n) has full rank n (All n columns are independent)

We want the sequence $\underline{q}_1, \underline{q}_2, \ldots$ to have the property

$$\langle \underline{a}_1, \quad \underline{a}_2, \ldots, \quad \underline{a}_j \rangle = \left\langle \underline{q}_1, \underline{q}_2, \ldots \underline{q}_j \right\rangle \quad j=1,\ldots,n$$

$$\left[\begin{array}{c|c|c|c} \\ \underline{a}_1 & \underline{a}_2 & \cdots & \underline{a}_n \\ \\ \end{array} \right] = \left[\begin{array}{c|c|c|c} \\ \underline{q}_1 & \underline{q}_2 & \cdots & \underline{q}_n \\ \\ \end{array} \right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix} \quad \text{----}(*)$$



- As A $\epsilon\, \boldsymbol{R}^{mxn}$(m > n) is assumed to be of full rank n, the diagonal entries $\boldsymbol{r_{kk}}$ are nonzero.
- If (*) holds, then $\underline{a}_1, \quad \underline{a}_2, \ldots, \quad \underline{a}_k$ can be expressed as linear combinations of $\underline{q}_1, \underline{q}_2, \cdots \underline{q}_k$.

$\underline{\boldsymbol{a}}_{\mathbf{1}} = r_{11}\underline{\boldsymbol{q}}_{\mathbf{1}}$

$\underline{\boldsymbol{a}}_2 = r_{12}\underline{\boldsymbol{q}}_{\mathbf{1}} + r_{22}\underline{\boldsymbol{q}}_2$

$\underline{\boldsymbol{a}}_2 = r_{13}\underline{\boldsymbol{q}}_{\mathbf{1}} + r_{23}\underline{\boldsymbol{q}}_2 + r_{33}\underline{\boldsymbol{q}}_3$

$\vdots$

$\underline{\boldsymbol{a}}_n = r_{1n}\underline{\boldsymbol{q}}_{\mathbf{1}} + r_{2n}\underline{\boldsymbol{q}}_2 + \cdots + r_{nn}\underline{\boldsymbol{q}}_n$

As a matrix formula, we have **A** = $\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{R}}$

where $\widehat{\boldsymbol{Q}}$ is **m x n** with orthonormal columns ($\widehat{\boldsymbol{Q}}'\ \widehat{\boldsymbol{Q}} = \boldsymbol{I}$). If A is square, then $\widehat{\boldsymbol{Q}}$ is orthogonal.

and $\widehat{\boldsymbol{R}}$ is **n x n** non-singular, upper triangular matrix (r_{kk}> 0).

Such a factorization is called a ***Reduced QR factorization*** of A.

The invertibility of the upper-left k x k block of the triangular matrix implies that, conversely, $\underline{\boldsymbol{q}}_{\mathbf{1}}, \underline{\boldsymbol{q}}_{\mathbf{2}}, \ldots \underline{\boldsymbol{q}}_k$ can be expressed as linear combinations of $\underline{\boldsymbol{a}}_{\mathbf{1}}, \quad \underline{\boldsymbol{a}}_{\mathbf{2}}, \ldots, \quad \underline{\boldsymbol{a}}_k$.

- Range of a matrix **A** $\boldsymbol{\epsilon}$ $\boldsymbol{R}^{mxn}$, R(A) = {A$\underline{x}$|$\underline{x} \in R^n$}
  suppose A has linearly independent columns with QR factors Q, R.
  Q has the same range as A:
  $\underline{y} \in$ range(A)
  $\Longleftrightarrow \underline{y} = A\underline{x}$ for some $\underline{x}$
  $\Longleftrightarrow \underline{y} = QR\underline{x}$ for some $\underline{x}$
  $\Longleftrightarrow \underline{y} = Q\underline{z}$ for some $\underline{z}$
  $\Longleftrightarrow \underline{y} \in$ range(Q)
  columns of Q are an orthonormal basis for range(A)

Ex:

$$A = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix} = \begin{bmatrix} -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 2 & 8 \\ 0 & 0 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix} = \widehat{\boldsymbol{Q}}\widehat{\boldsymbol{R}}$$

Full QR Factorization

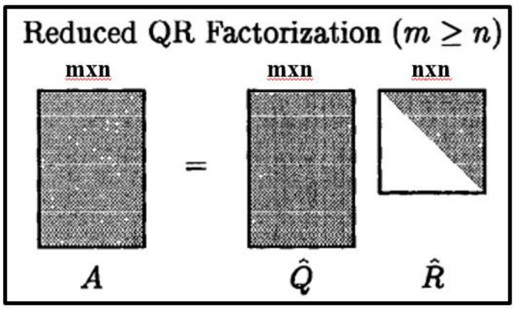
For full QR factorization of **A** $\boldsymbol{\epsilon}$ $\boldsymbol{R}^{mxn}$(m ≥ n) , an additional (m – n) orthonormal columns are appended to $\widehat{\boldsymbol{Q}}$ so that it becomes an m x m unitary matrix **Q**.

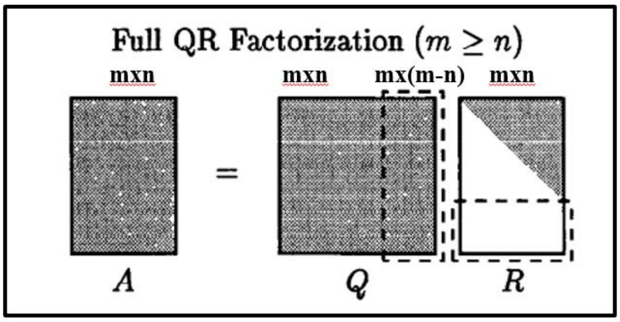
$$A = [\hat{Q} \quad \tilde{Q}]\begin{bmatrix}\hat{R}\\0\end{bmatrix} = QR$$

In the process, rows of zeros are appended to R so that it becomes an m x n matrix R, still upper-triangular.

In the full QR factorization, Q is m x m, R is m x n, and the last (m-n) columns of Q are multiplied by zeros in R.

In the reduced QR factorization, the silent columns and rows are removed. Now Q is m x n, R is n x n, and none of the rows of R are necessarily zero.





Note that in the full QR factorization, the columns $\underline{q_j}$ for j > n are orthogonal to range(A).

Assuming A is of full rank n, they constitute an orthonormal basis for range$A^{\perp}$ (the space orthogonal to range(A)), i.e, for null (A').

Applications

The QR factorization to solve

• the pseudo-inverse of a matrix with linearly independent columns

• the inverse of a nonsingular matrix

• projection on the range of a matrix with linearly independent columns

• linear equations

• least squares problems

• constrained least squares problems

Pseudo-inverse of matrix with independent columns

Suppose **A** $\boldsymbol{\epsilon}$ $\boldsymbol{R}^{mxn}$ has linearly independent columns.

This implies that A is tall or square $(m \geq n)$

The pseudo-inverse of A is defined as $A^{\dagger} = (A'\ A)^{-1}A'$

This matrix exists, because the Gram matrix $(A'\ A)$ is non-singular.

$A^{\dagger}$ is a left inverse of A: $A^{\dagger}A = (A'\ A)^{-1}\ (A'\ A) = I$.

Pseudo-inverse of matrix in terms of QR Factors

$A^{\dagger} = (A'\ A)^{-1}A'$

$A^{\dagger} = ((QR)'\ (QR))^{-1}(QR)'$

$= (R'Q'\ QR)^{-1}R'Q'$

$= (R'R)^{-1}R'Q'$

$= R^{-1}(R')^{-1}R'Q'$

$= R^{-1}Q'$

For square nonsingular A this is the inverse: $A^{-1} = R^{-1}Q'$

Solution of Ax = b by QR Factorization

To solve $A\underline{x} = \underline{b}$ for $\underline{x}$, where $\mathbf{A} \in \boldsymbol{R}^{mxn}$.

If A = QR is a QR factorization, then we can write $QR\underline{x} = \underline{b}$

$$\Rightarrow R\underline{x} = Q'\underline{b}$$

The right-hand side of this equation is easy to compute, if Q is known,

The system of linear equations implicit in the left-hand side is also easy to solve because it is triangular.

This suggests the following method for computing the solution to Ax = b:

1. Compute a QR factorization A = QR.
2. Compute $\underline{y} = Q'\underline{b}$.
3. Solve $R\underline{x} = \underline{y}$ for $\underline{x}$.

However, it is not the standard method for such problems. Gaussian elimination is the algorithm generally used in practice, since it requires only half as many numerical operations.

The LS Problem

The ***QR*** matrix decomposition allows us to *compute* the solution to the Least Squares problem.

OLS gives us the closed from solution in the form of the normal equations.

But when we want to find the actual numerical solution they aren't really useful.

In Least Squares problem we want to solve the following equation: X$\underline{\beta} = \underline{y}$

Usually $\underline{\hat{\beta}} = X^{-1}\underline{y}$ does not exist as we might have more observations than variables and X^{-1} might not exist.

Instead, we try to find some $\underline{\hat{\beta}}$ that minimizes the objective function $\sum(\underline{y} - X\underline{\beta})^2$.

Taking derivatives with respect to $\underline{\hat{\beta}}$ and setting to zero will lead us to the normal equations and provides a closed-form solution.

Alternative way: QR Factorization!

Solving the LS problem Using the QR factorization

X$\underline{\beta} = \underline{y}$

Assuming the columns of X to be Linearly independent, X=QR

⇒ (QR) $\underline{\beta} = \underline{y}$

⇒ $\underline{\hat{\beta}} = \mathrm{R}^{-1}Q'\underline{y}$ [Existence of Pseudo Inverse of X]

This means that all we need to do is find an inverse of ***R***, transpose ***Q***, and take the product. That will produce the OLS coefficients.

We don't even need to compute the Variance-Covariance matrix and its inverse $(X'X)^{-1}$which is how OLS solutions are usually presented.

QR using Gram-Schmidt Orthogonalisation

Let A be an m x n matrix with n independent columns.

To find the QR factors of A.

Apply Gram Schmidt orthogonalisation to obtain Q from the columns of A.

⇒ Columns of Q will span the Column Space of A.

Once we have Q we can solve for R easily from QR = A ⇒ R = $Q'A$