Thin QR Decomposition via Gram-Schmidt

From Gram-Schmidt,

$$\mathbf{a}_i = \sum_{j=1}^{i} r_{ji}\mathbf{q}_j, \quad i = 1, \ldots, n$$

where

$$r_{ii} = \|\tilde{\mathbf{q}}_i\|_2, \quad r_{ji} = \mathbf{q}_j^T \mathbf{a}_i, \ j = 1, \ldots, i-1$$

Equivalently,

$$\mathbf{A} = \mathbf{Q}_1 \mathbf{R}_1$$

where

$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ full column rank

$\mathbf{Q}_1 = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$ semi-orthogonal

$\mathbf{R}_1$ is upper triangular with $[\mathbf{R}_1]_{ij} = r_{ij}$ for $i \leq j$

$\tilde{q}_i \neq 0 \quad \forall i$

- $\mathbf{R}_1$ is nonsingular because $\det(\mathbf{R}) = \Pi_{i=1}^n r_{ii} \neq 0$

General Gram-Schmidt Procedure

Extension to the case where $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$ may not have full column rank

Observation from Gram-Schmidt:

- If $\mathbf{a}_j$ is linearly dependent of $\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}$, then $\tilde{\mathbf{q}}_j = 0$
- The number of nonzero $\tilde{\mathbf{q}}_i$'s is rank(A)

Idea: If $\tilde{\mathbf{q}}_j = 0$, skip to $j+1$ without computing $\mathbf{q}_j$

All the $\mathbf{q}_i$'s form an orthonormal basis for $\mathcal{R}(\mathbf{A})$

Algorithm: General Gram-Schmidt

input: a collection of possibly linearly dependent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$

k=0

for $i = 1, \ldots, n$

  $\tilde{\mathbf{q}}_i = \mathbf{a}_i - \sum_{j=1}^{k} (\mathbf{q}_j^T \mathbf{a}_i)\mathbf{q}_j$

  if $\tilde{\mathbf{q}}_i \neq 0$

    $k \leftarrow k + 1$

    $\mathbf{q}_k = \tilde{\mathbf{q}}_i / \|\tilde{\mathbf{q}}_i\|_2$

  end % $\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_i\} = \mathrm{span}\{\mathbf{q}_1 \ldots, \mathbf{q}_k\}$

end

output: $\mathbf{q}_1, \ldots, \mathbf{q}_k$ % $k = \mathrm{rank}(A)$

$i=1$: $\sum_{j=1}^{0} \boxed{} = 0$

General Gram-Schmidt Procedure (cont'd)

Example: Let $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{a}_5 \end{bmatrix} \in \mathbb{R}^{6\times 5}$
Suppose $\mathbf{a}_1 \neq \mathbf{0}$; $\mathbf{a}_2$ is linearly independent from $\mathbf{a}_1$; $\mathbf{a}_3$ is linearly dependent of $\mathbf{a}_1$ and $\mathbf{a}_2$; $\mathbf{a}_4$ is linearly independent from $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$; $\mathbf{a}_5$ is linearly dependent of $\mathbf{a}_2$ only

$\tilde{q}_1 = a_1 \neq 0 \qquad k=1 \qquad q_1 = \tilde{q}_1 / \|\tilde{q}_1\|_2$

$\tilde{q}_2 = a_2 - (q_1^T a_2) q_1 \neq 0 \qquad k=2 \qquad q_2 = \tilde{q}_2 / \|\tilde{q}_2\|_2$

$\tilde{q}_3 = a_3 - (q_1^T a_3) q_1 - (q_2^T a_3) q_2 = 0 \qquad k=2$

$\text{span}\{q_1, q_2\} = \text{span}\{a_1, a_2\} = \text{span}\{a_1, a_2, a_3\}$

$\tilde{q}_4 = a_4 - (q_1^T a_4) q_1 - (q_2^T a_4) q_2 \neq 0 \qquad k=3 \qquad q_3 = \tilde{q}_4 / \|\tilde{q}_4\|_2$

$\tilde{q}_5 = a_5 - (q_1^T a_5) q_1 - (q_2^T a_5) q_2 - (q_3^T a_5) q_3 = 0 \qquad k=3$

$\text{span}\{q_1, q_2, q_3\} = \text{span}\{a_1, \ldots, a_5\}$

General Gram-Schmidt Procedure (cont'd)

Example: Let $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{a}_5 \end{bmatrix} \in \mathbb{R}^{6\times 5}$

Suppose $\mathbf{a}_1 \neq \mathbf{0}$; $\mathbf{a}_2$ is linearly independent from $\mathbf{a}_1$; $\mathbf{a}_3$ is linearly dependent of $\mathbf{a}_1$ and $\mathbf{a}_2$; $\mathbf{a}_4$ is linearly independent from $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$; $\mathbf{a}_5$ is linearly dependent of $\mathbf{a}_2$ only

$$a_1 = \|\tilde{q}_1\|_2 \cdot q_1, \quad a_2 = \underbrace{(q_1^T a_2)}_{0 \text{ if } a_1 \perp a_2} q_1 + \underbrace{\|\tilde{q}_2\|_2}_{\neq 0} q_2$$

$$a_3 = (q_1^T a_3) q_1 + (q_2^T a_3) q_2$$

$$a_4 = (q_1^T a_4) q_1 + (q_2^T a_4) q_2 + \|\tilde{q}_4\|_2 \cdot q_3$$

$$a_5 = (q_1^T a_5) q_1 + (q_2^T a_5) q_2 + \underbrace{(q_3^T a_5)}_{=0} q_3$$

$$a_5 \in \operatorname{span}\{a_2\} \subset \operatorname{span}\{q_1, q_2\}, \quad a_5 \perp q_3$$

$$A = [a_1 \; a_2 \; a_3 \; a_4 \; a_5] = [q_1 \; q_2 \; q_3] \begin{bmatrix} \|\tilde{q}_1\|_2 & q_1^T a_2 & q_1^T a_3 & q_1^T a_4 & q_1^T a_5 \\ 0 & \|\tilde{q}_2\|_2 & q_2^T a_3 & q_2^T a_4 & q_2^T a_5 \\ 0 & 0 & 0 & \|\tilde{q}_4\|_2 & 0 \end{bmatrix}$$

General Gram-Schmidt Procedure (cont'd)

Using General Gram-Schmidt, $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\text{rank}(\mathbf{A}) = k \leq n$ can be decomposed as

$$\mathbf{A} = \mathbf{Q}_1 \mathbf{R}_1$$

where

$\mathbf{Q}_1 = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_k \end{bmatrix} \in \mathbb{R}^{m \times k}$ is semi-orthogonal

$\mathbf{R}_1 \in \mathbb{R}^{k \times n}$ is in an upper staircase form, where each staircase corresponds to a column of $\mathbf{A}$ that is independent from previous columns

$\mathbf{R}_1 \in \mathbb{R}^{k \times n}$ is upper triangular[1]

Applications:

- Obtain an orthonormal basis for $\mathcal{R}(\mathbf{A})$
- Check whether $\mathbf{b} \in \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ by applying general Gram-Schmidt to $\{\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{b}\}$
- The staircase pattern of $\mathbf{R}_1$ indicates the dependence of each column of $\mathbf{A}$ on previous columns

[1] From now on, we say a rectangular matrix is upper triangular if its (i, j)-entry is zero for all $i > j$

QR Decomposition

Theorem

Any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ *admits a decomposition*

$$\mathbf{A} = \mathbf{QR} \qquad \textit{(QR Decomposition)}$$

where $\mathbf{Q} \in \mathbb{R}^{m \times m}$ *is an orthogonal matrix and* $\mathbf{R} \in \mathbb{R}^{m \times n}$ *is an upper triangular matrix.*
In addition, when $m = n$ *and* $\mathbf{A}$ *has full rank,* $(\mathbf{Q}, \mathbf{R})$ *is unique if we restrict* $r_{ii} > 0$ *for all* i.

Finding QR Decomposition via General Gram-Schmidt

1. Find any matrix $\tilde{\mathbf{A}}$ s.t. the matrix $\begin{bmatrix}\mathbf{A} & \tilde{\mathbf{A}}\end{bmatrix}$ has full row rank   rank = m
   - We may simply let $\tilde{\mathbf{A}} = \mathbf{I}_m$
2. Applying General Gram-Schmidt gives

$$\begin{bmatrix}\mathbf{A} & \tilde{\mathbf{A}}\end{bmatrix} = \mathbf{Q}\bar{\mathbf{R}}, \quad \mathbf{Q} \in \mathbb{R}^{m \times m} \text{ orthogonal}$$

3. Write $\mathbf{Q} = \begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix}$ where
   - $\mathbf{Q}_1 \in \mathbb{R}^{m \times k}$ with $k = \text{rank}(\mathbf{A})$ provides an orthonormal basis for $\mathcal{R}(A)$
   - $\mathbf{Q}_2 \in \mathbb{R}^{m \times (m-k)}$ provides an orthonormal basis for $\mathcal{R}(\tilde{\mathbf{A}})$
4. Note that

$$\mathbf{A} = \underbrace{\mathbf{Q}_1}_{m \times k} \underbrace{\mathbf{R}_1}_{k \times n} = \underbrace{\begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix}\mathbf{R}_1 \\ \mathbf{0}_{(m-k) \times n}\end{bmatrix}}_{\mathbf{R}}$$

Discussions

Thin QR Decomposition for general $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m \geq n$:

$$\mathbf{A} = \tilde{\mathbf{Q}}_1 \underbrace{\begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0}_{(n-k)\times n} \end{bmatrix}}_{\tilde{\mathbf{R}}_1}$$

where

$\tilde{Q}_1(:, 1:k) = Q_1$

$\tilde{\mathbf{Q}}_1 \in \mathbb{R}^{m \times n}$ is semi-orthogonal

$\tilde{\mathbf{R}}_1 \in \mathbb{R}^{n \times n}$ is upper triangular

When $\mathbf{A}$ has full column rank, then $\tilde{\mathbf{Q}}_1 = \mathbf{Q}_1$ and $\tilde{\mathbf{R}}_1 = \mathbf{R}_1$

$\mathbf{A}$ has full column rank if and only if $[\mathbf{R}_1]_{ii} \neq 0$ for all i

Discussions (cont'd)

Since $\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \in \mathbb{R}^{m \times m}$ is orthogonal,

- $\mathcal{R}(\mathbf{Q}_1)$ and $\mathcal{R}(\mathbf{Q}_2)$ are orthogonal
- $\mathcal{R}(\mathbf{Q}_1)$ and $\mathcal{R}(\mathbf{Q}_2)$ spans $\mathbb{R}^m$ $= \mathcal{R}(Q)$

Therefore, $\mathcal{R}(\mathbf{Q}_1)$ and $\mathcal{R}(\mathbf{Q}_2)$ are orthogonal complements of each other, i.e.,

$$\mathcal{R}(\mathbf{Q}_1)^{\perp} = \mathcal{R}(\mathbf{Q}_2)$$

It follows that

$$\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{Q}_1), \quad \mathcal{R}(\mathbf{A})^{\perp} = \mathcal{R}(\mathbf{Q}_2)$$

- The columns of $\mathbf{Q}_1$ form an orthonormal basis for $\mathcal{R}(\mathbf{A})$
- The columns of $\mathbf{Q}_2$ form an orthonormal basis for $\mathcal{R}(\mathbf{A})^{\perp} = \mathcal{N}(\mathbf{A}^T)$

LS via QR

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\text{rank}(\mathbf{A}) = k$

$$\mathbf{A} = \begin{bmatrix} \underbrace{\mathbf{Q}_1}_{m\times k} & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \underbrace{\mathbf{R}_1}_{k\times n} \\ \mathbf{0}_{(m-k)\times n} \end{bmatrix}$$

$$\|Q^T z\|_2^2 = (Q^T z)^T (Q^T z) = z^T \underbrace{Q Q^T}_{= I} z$$

for orthogonal Q

Using the QR decomposition,

$$\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 = \|\mathbf{Q}^T\mathbf{A}\mathbf{x} - \mathbf{Q}^T\mathbf{y}\|_2^2 \qquad \text{because orthogonal } Q \text{ preserves norm}$$

$$= \left\| \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \mathbf{y} \right\|_2^2$$

Q_1, Q_2 semi-orthogonal

$$= \left\| \begin{bmatrix} \underbrace{\mathbf{Q}_1^T\mathbf{Q}_1}_{\mathbf{I}} & \underbrace{\mathbf{Q}_1^T\mathbf{Q}_2}_{\mathbf{0}} \\ \underbrace{\mathbf{Q}_2^T\mathbf{Q}_1}_{\mathbf{0}} & \underbrace{\mathbf{Q}_2^T\mathbf{Q}_2}_{\mathbf{I}} \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{Q}_1^T\mathbf{y} \\ \mathbf{Q}_2^T\mathbf{y} \end{bmatrix} \right\|_2^2$$

$$= \left\| \begin{bmatrix} \mathbf{R}_1\mathbf{x} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{Q}_1^T\mathbf{y} \\ \mathbf{Q}_2^T\mathbf{y} \end{bmatrix} \right\|_2^2 = \|\mathbf{R}_1\mathbf{x} - \mathbf{Q}_1^T\mathbf{y}\|_2^2 + \|\mathbf{Q}_2^T\mathbf{y}\|_2^2$$

LS via QR (cont'd)

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

Conclusion: $\mathbf{x}_{\mathrm{LS}}$ is a least-squares solution to $\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{Ax}\|_2^2$ if and only if it is a least-squares solution to $\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b}_1 - \mathbf{R}_1\mathbf{x}\|_2^2$ where $\mathbf{b}_1 = \mathbf{Q}_1^T\mathbf{y}$

Suppose $\mathbf{A}$ has full column rank, i.e., $k = n$

Then, $\mathbf{R}_1$ is nonsingular and the unique least-squares solution is

$$\mathbf{x}_{\mathrm{LS}} = \mathbf{R}_1^{-1}\mathbf{b}_1$$

$O(n^2)$

We may solve the triangular system $\mathbf{R}_1\mathbf{x} = \mathbf{b}_1$ by backward substitution

In this case, the optimal residual $\|\mathbf{Ax}_{\mathrm{LS}} - \mathbf{y}\|_2$ is

$$\|\mathbf{Q}_2^T\mathbf{y}\|_2 = \|\mathbf{Q}_2\mathbf{Q}_2^T\mathbf{y}\|_2$$

Note that $\mathbf{Q}_2\mathbf{Q}_2^T\mathbf{y} \in \mathcal{R}(\mathbf{Q}_2) = \mathcal{R}(\mathbf{Q}_1)^\perp = \mathcal{R}(\mathbf{A})^\perp$

$\mathbf{Q}_2\mathbf{Q}_2^T\mathbf{y}$ is the component of $\mathbf{y}$ orthogonal to $\mathcal{R}(\mathbf{A})$

Numerical Error Issue of Gram-Schmidt

Gram-Schmidt is numerically unstable due to propagation of numerical errors

Example: Given
$\mathbf{a}_1 = \begin{bmatrix} 1 & \epsilon & 0 & 0 \end{bmatrix}^T, \mathbf{a}_2 = \begin{bmatrix} 1 & 0 & \epsilon & 0 \end{bmatrix}^T, \mathbf{a}_3 = \begin{bmatrix} 1 & 0 & 0 & \epsilon \end{bmatrix}^T$ with tiny ϵ so that the approximation $1 + \epsilon^2 \approx 1$ can be made

Applying Gram-Schmidt with the above approximation yields

- $\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|_2} = \begin{bmatrix} 1 & \epsilon & 0 & 0 \end{bmatrix}^T$ $\quad \|a_1\|_2 = \sqrt{1+\epsilon^2} \approx 1$
- $\tilde{\mathbf{q}}_2 = \mathbf{a}_2 - \mathbf{q}_1^T \mathbf{a}_2 \mathbf{q}_1 = \begin{bmatrix} 0 & -\epsilon & \epsilon & 0 \end{bmatrix}^T$ no round-off error
  $\mathbf{q}_2 = \frac{\tilde{\mathbf{q}}_2}{\|\tilde{\mathbf{q}}_2\|_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 & 1 & 0 \end{bmatrix}^T$
- $\tilde{\mathbf{q}}_3 = \mathbf{a}_3 - \mathbf{q}_1^T \mathbf{a}_3 \mathbf{q}_1 - \mathbf{q}_2^T \mathbf{a}_3 \mathbf{q}_2 = \begin{bmatrix} 0 & -\epsilon & 0 & \epsilon \end{bmatrix}^T$ no error
  $\mathbf{q}_3 = \frac{\tilde{\mathbf{q}}_3}{\|\tilde{\mathbf{q}}_3\|_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 & 0 & 1 \end{bmatrix}^T$

  $q_1^T q_2 \approx 0$

  Orthogonality is lost!

  $q_2^T q_3 = \frac{1}{2}$

Modified Gram-Schmidt

The ith iteration

Instead of computing $\tilde{\mathbf{q}}_i = \mathbf{a}_i - (\mathbf{q}_1^T \mathbf{a}_i)\mathbf{q}_1 - (\mathbf{q}_2^T \mathbf{a}_i)\mathbf{q}_2 - \cdots - (\mathbf{q}_{i-1}^T \mathbf{a}_i)\mathbf{q}_{i-1}$ in Gram-Schmidt (full column rank case), compute

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

Complexity: $\mathcal{O}(mn^2)$

Modified Gram-Schmidt (cont'd)

Example (revisit): Given
$\mathbf{a}_1 = \begin{bmatrix}1 & \epsilon & 0 & 0\end{bmatrix}^T, \mathbf{a}_2 = \begin{bmatrix}1 & 0 & \epsilon & 0\end{bmatrix}^T, \mathbf{a}_3 = \begin{bmatrix}1 & 0 & 0 & \epsilon\end{bmatrix}^T$ with tiny ϵ so that the approximation $1+\epsilon^2 \approx 1$ can be made

Applying modified Gram-Schmidt with the above approximation yields

- $\tilde{\mathbf{q}}_1 = \begin{bmatrix}1 & \epsilon & 0 & 0\end{bmatrix}^T$
  $\mathbf{q}_1 = \begin{bmatrix}1 & \epsilon & 0 & 0\end{bmatrix}^T$ round-off error
- $\tilde{\mathbf{q}}_2 = \mathbf{a}_2 - \mathbf{q}_1^T \mathbf{a}_2 \mathbf{q}_1 = \begin{bmatrix}0 & -\epsilon & \epsilon & 0\end{bmatrix}^T$
  $\mathbf{q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix}0 & -1 & 1 & 0\end{bmatrix}^T$
- $\tilde{\mathbf{q}}_3^{(1)} = \mathbf{a}_3 - \mathbf{q}_1^T \mathbf{a}_3 \mathbf{q}_1 = \begin{bmatrix}0 & -\epsilon & 0 & \epsilon\end{bmatrix}^T$
  $\tilde{\mathbf{q}}_3 = \tilde{\mathbf{q}}_3^{(2)} = \tilde{\mathbf{q}}_3^{(1)} - \mathbf{q}_2^T \tilde{\mathbf{q}}_3^{(1)} \mathbf{q}_2 = \begin{bmatrix}0 & -\frac{\epsilon}{2} & -\frac{\epsilon}{2} & \epsilon\end{bmatrix}^T$
  $\mathbf{q}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix}0 & -1 & -1 & 2\end{bmatrix}^T$

  Orthogonality is preserved approximately

We may also compute QR using reflection and rotation approaches