Lesson 11: Gram-Schmidt process, $QR = A$

- New concepts:
  - Orthonormal vector set
  - Transforming a basis set into an orthonormal set by Gram-Schmidt
  - QR factorization of a matrix

Definition. *The Dirac delta symbol* δ_{ij} *is defined as*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Definition. *A set of vectors* $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ *is said to be orthonormal if*

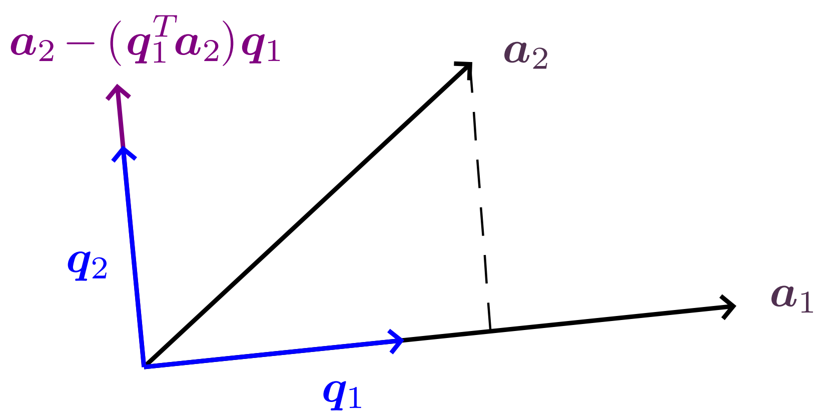
$$\boldsymbol{u}_i^T \boldsymbol{u}_j = \delta_{ij}$$

- The column vectors of the identity matrix are orthonormal

$$\boldsymbol{I} = \begin{pmatrix} \boldsymbol{e}_1 & \ldots & \boldsymbol{e}_m \end{pmatrix}$$

$$\boldsymbol{e}_i^T \boldsymbol{e}_j = \delta_{ij}$$

- An arbitrary vector set can be transformed into an orthonormal set by the Gram-Schmidt algorithm
- Idea:
  - Start with an arbitrary direction $\boldsymbol{a}_1$
  - Divide by its norm to obtain a unit-norm vector $\boldsymbol{q}_1 = \boldsymbol{a}_1 / \|\boldsymbol{a}_1\|$
  - Choose another direction $\boldsymbol{a}_2$
  - Subtract off its component along previous direction(s) $\boldsymbol{a}_2 - (\boldsymbol{q}_1^T \boldsymbol{a}_2)\boldsymbol{q}_1$
  - Divide by norm $\boldsymbol{q}_2 = (\boldsymbol{a}_2 - (\boldsymbol{q}_1^T \boldsymbol{a}_2)\boldsymbol{q}_1) / \|\boldsymbol{a}_2 - (\boldsymbol{q}_1^T \boldsymbol{a}_2)\boldsymbol{q}_1\|$
  - Repeat the above



- Consider $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ with linearly independent columns. By linear combinations of the columns of $\boldsymbol{A}$ a set of orthonormal vectors $\boldsymbol{q}_1$, ..., $\boldsymbol{q}_n$ will be obtained. This can be expressed as a matrix product

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1n} \\ 0 & r_{22} & r_{23} & \dots & r_{2n} \\ 0 & 0 & r_{33} & \dots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & r_{mn} \end{pmatrix} = \boldsymbol{Q}\boldsymbol{R}$$

  with $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$, $\boldsymbol{R} \in \mathbb{R}^{n\times n}$. The matrix $\boldsymbol{R}$ is upper-triangular (also referred to as right-triangular) since to find vector $\boldsymbol{q}_1$ only vector $\boldsymbol{a}_1$ is used, to find vector $\boldsymbol{q}_2$ only vectors $\boldsymbol{a}_1, \boldsymbol{a}_2$ are used

- The above is equivalent to the system

$$\begin{cases} \boldsymbol{a}_1 = r_{11}\boldsymbol{q}_1 \\ \boldsymbol{a}_2 = r_{12}\boldsymbol{q}_1 + r_{22}\boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{a}_n = r_{1n}\boldsymbol{q}_1 + r_{2n}\boldsymbol{q}_2 + \dots + r_{nn}\boldsymbol{q}_n \end{cases}$$

- The system can be solved to find $\boldsymbol{R}, \boldsymbol{Q}$ by:
  1. Imposing $\|\boldsymbol{q}_1\| = 1 \Rightarrow r_{11} = \|\boldsymbol{a}_1\|$, $\boldsymbol{q}_1 = a_1 / r_{11}$
  2. Computing projections of $\boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ along $\boldsymbol{q}_1$

     $$r_{12} = \boldsymbol{q}_1^T \boldsymbol{a}_2, \ldots, r_{1n} = \boldsymbol{q}_1^T \boldsymbol{a}_n$$

  3. Subtracting components along $\boldsymbol{q}_1$ from $\boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$

     $$\begin{cases} \boldsymbol{a}_2 - r_{12}\boldsymbol{q}_1 = r_{22}\boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{a}_n - r_{1n}\boldsymbol{q}_1 = r_{2n}\boldsymbol{q}_2 + \ldots + r_{nn}\boldsymbol{q}_n \end{cases}$$

  4. The above steps reduced the size of the system by 1. Repeating the steps completes the solution. The overall process is known as the Gram-Schmidt algorithm

Algorithm (Gram-Schmidt)

Given m vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$
Initialize $\boldsymbol{q}_1 = \boldsymbol{a}_1$,..,$\boldsymbol{q}_m = \boldsymbol{a}_m$, $\boldsymbol{R} = \boldsymbol{I}$
for $i = 1$ to m
  $r_{ii} = (\boldsymbol{q}_i^T \boldsymbol{q}_i)^{1/2}$; $\boldsymbol{q}_i = \boldsymbol{q}_i / r_{ii}$
  for $j = i$+1 to m
    $r_{ij} = \boldsymbol{q}_i^T \boldsymbol{a}_j$; $\boldsymbol{q}_j = \boldsymbol{q}_j - r_{ij} \boldsymbol{q}_i$
  end
end
return $\boldsymbol{Q}, \boldsymbol{R}$

$\boldsymbol{Q}\boldsymbol{R}$ factorization

- For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with linearly independent columns, the Gram-Schmidt algorithm furnishes a factorization

$$\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{A}$$

with $\boldsymbol{Q} \in \mathbb{R}^{m \times n}$ with orthonormal columns and $\boldsymbol{R} \in \mathbb{R}^{n \times n}$ an upper triangular matrix.

- Since the column vectors within $\boldsymbol{Q}$ were obtained through linear combinations of the column vectors of $\boldsymbol{A}$ we have

$$C(\boldsymbol{A}) = C(\boldsymbol{Q})$$