Orthogonality - The Gram-Schmidt algorithm

1. Two vectors \( u \) and \( v \) are orthogonal if \( u \cdot v = 0 \).
2. They are orthonormal if in addition \( \|u\| = \|v\| = 1 \).
3. In this case the matrix \( Q = [u, v] \) is such that \( Q^T Q = I \).

- We say that the system \( \{u, v\} \) is orthonormal.
- and that the matrix \( Q \) has orthonormal columns.
- or that it is orthogonal [Text reserves this term to \( n \times n \) case].

Example: An orthonormal system \( \{u, v\} \)

\[
\begin{align*}
u &= \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix} \\
v &= \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}
\end{align*}
\]

Generalization: (to \( n \) vectors)

- A system of vectors \( \{v_1, \ldots, v_n\} \) is orthogonal if \( v_i \cdot v_j = 0 \) for \( i \neq j \); and orthonormal if in addition \( \|v_i\| = 1 \) for \( i = 1, \ldots, n \).

A matrix is orthogonal if its columns are orthonormal.

Then: \( V = [v_1, \ldots, v_n] \) has orthonormal columns.

[Note: The term 'orthonormal matrix' is not used. 'orthogonal' is often used for square matrices only (textbook)].

Question: We are given the set \( \{u_1, u_2, \ldots, u_n\} \) which is not orthogonal. How do we get a set of vectors \( \{q_1, q_2, \ldots, q_n\} \) that is orthonormal and spans the same subspace as \( \{u_1, u_2, \ldots, u_n\} \)?

Rationale: Orthonormal systems are easier to use.

Answer: Gram-Schmidt process - to be described next.

*See section 6.4 of text – example 1 with 2 vectors.*
The Gram-Schmidt algorithm

Problem: Given a set \( \{u_1, u_2\} \) how can we generate another set \( \{q_1, q_2\} \) from linear combinations of \( u_1, u_2 \) so that \( \{q_1, q_2\} \) is orthonormal?

**Step 1** Define first vector: \( q_1 = \frac{u_1}{\|u_1\|} \) (‘Normalization’)

**Step 2:** Orthogonalize \( u_2 \) against \( q_1 \): \( \hat{q} = u_2 - (u_2.q_1)q_1 \)

**Step 3** Normalize to get second vector: \( q_2 = \hat{q}/\|\hat{q}\| \)

Result: \( \{q_1, q_2\} \) is an orthonormal set of vectors which spans the same space as \( \{u_1, u_2\} \).

**Example:**
\[
\begin{align*}
    u_1 &= \begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix} \\
    u_2 &= \begin{pmatrix} 2 \\ 0 \\ 0 \\ 2 \end{pmatrix}
\end{align*}
\]

**Step 1:** \( q_1 = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix} \)

**Step 2:** First compute \( u_2.q_1 = \ldots = 2 \). Then:
\[
\begin{align*}
    \hat{q} &= \begin{pmatrix} 2 \\ 0 \\ 0 \\ 2 \end{pmatrix} - 2 \times \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix} \\
    q_2 &= \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}
\end{align*}
\]

**Generalization: 3 vectors**

How to generalize to 3 or more vectors?

For 3 vectors: \( [u_1, u_2, u_3] \)

- First 2 steps are the same \( \rightarrow q_1, q_2 \)
- Then orthogonalize \( u_3 \) against \( q_1 \) and \( q_2 \):
\[
\begin{align*}
    \hat{q} &= u_3 - (u_3.q_1)q_1 - (u_3.q_2)q_2
\end{align*}
\]

- Finally, normalize:
\[
q_3 = \frac{\hat{q}}{\|\hat{q}\|}
\]

General problem: Given \( U = [u_1, \ldots, u_n] \), compute \( Q = [q_1, \ldots, q_n] \) which is orthonormal and s.t. \( \text{Col}(Q) = \text{Col}(U) \).
**ALGORITHM : 1. Classical Gram-Schmidt**

1. For \( j = 1 : n \) Do:
   2. \( \hat{q} = u_j \)
   3. For \( i = 1 : j - 1 \)
   4. \( \hat{q} := \hat{q} - (u_j \cdot q_i)q_i \) / set \( r_{ij} = (u_j \cdot q_i) \)
   5. End
   6. \( q_j := \hat{q}/\|\hat{q}\| \) / set \( r_{jj} = \|\hat{q}\| \)
   7. End

- All \( n \) steps can be completed iff \( u_1, u_2, \ldots, u_n \) are linearly independent.

- Define a matrix \( R \) as follows:
  
  \[
  r_{ij} = \begin{cases} 
  u_j \cdot q_i & \text{if } i < j \text{ (see line 4)} \\
  \|\hat{q}\| & \text{if } i = j \text{ (see line 6)} \\
  0 & \text{if } i > j \text{ (lower part)}
  \end{cases}
  \]

- We have from the algorithm: (For \( j = 1, 2, \ldots, n \))
  
  \[ u_j = r_{1j}q_1 + r_{2j}q_2 + \ldots + r_{jj}q_j \]

- If \( U = [u_1, u_2, \ldots, u_n] \), \( Q = [q_1, q_2, \ldots, q_n] \), and if \( R \) is the \( n \times n \) upper triangular matrix defined above:
  
  \[ R = \{r_{ij}\}_{i,j=1,\ldots,n} \]

  then the above relation can be written as
  
  \[ U = QR \]

- This is called the QR factorization of \( U \).

- \( Q \) has orthonormal columns. It satisfies:
  
  \[ QTQ = I \]

- It is said to be orthogonal

- \( R \) is upper triangular

- What is the inverse of an orthogonal \( n \times n \) matrix?

- Show that when \( U \in \mathbb{R}^{m \times n} \) the total cost of Gram-Schmidt is \( \approx 2mn^2 \).

**Another decomposition:**

A matrix \( U \), with linearly independent columns, is the product of an orthogonal matrix \( Q \) and a upper triangular matrix \( R \).

\[
U = \begin{bmatrix} \* \end{bmatrix} Q \begin{bmatrix} R \end{bmatrix}
\]

\( R \) is upper triangular

Original matrix \( Q \) is orthogonal \((QTQ = I)\)
Orthonormalize the system of vectors:

\[ U = [u_1, u_2, u_3] = \begin{pmatrix} 1 & -4 & 3 \\ -1 & 2 & -1 \\ 1 & 0 & 1 \\ 1 & -2 & -1 \end{pmatrix} \]

For this example:

1) what is \( Q \)? what is \( R \)?

2) Verify (matlab) that \( U = QR \)

3) Compute \( QTQ \). [Result should be the identity matrix]

\[ \text{Solution: [values for } R \text{ are in red]} \]

Step 1: \( q_1 = \frac{u_1}{\|u_1\|} = \frac{1}{\frac{\sqrt{2}}{2}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, r_{11} = \|u_1\| = 2 \]

Step 2: \( \hat{q}_2 = u_2 - (u_2.q_1)q_1 \)
\[
\begin{pmatrix} -4 \\ 0 \\ -2 \end{pmatrix} - \frac{-8}{2} \times \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix}, r_{12} = \frac{-8}{2} = -4
\]

\[ \rightarrow q_2 = \frac{\hat{q}_2}{\|\hat{q}_2\|} = \frac{1}{\sqrt{8}} \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, r_{22} = \sqrt{8} \]

\[ \text{Solving LS systems via QR factorization} \]

In practice: not a good idea to solve the system \( ATAx = ATb \).
Use the QR factorization instead. How?

Answer in the form of an exercise

Problem: \( Ax \approx b \) in least-squares sense

\( A \) is an \( m \times n \) (full-rank) matrix.
Consider the QR factorization of \( A \)

\[ A = QR \]

Approach 1: Write the normal equations – then 'simplify'

Approach 2: Write the condition \( b - Ax \perp \text{Col}(A) \) and recall that \( A \) and \( Q \) have the same column space.

Total cost?