Lecture 4
Orthonormal sets of vectors and $QR$ factorization

- orthonormal sets of vectors
- Gram-Schmidt procedure, $QR$ factorization
- orthogonal decomposition induced by a matrix
Orthonormal set of vectors

set of vectors $u_1, \ldots, u_k \in \mathbb{R}^n$ is

- **normalized** if $\|u_i\| = 1$, $i = 1, \ldots, k$
  ($u_i$ are called *unit vectors* or *direction vectors*)
- **orthogonal** if $u_i \perp u_j$ for $i \neq j$
- **orthonormal** if both

**slang:** we say ‘$u_1, \ldots, u_k$ are orthonormal vectors’ but orthonormality (like independence) is a property of a set of vectors, not vectors individually

in terms of $U = [u_1 \cdots u_k]$, orthonormal means

$$U^T U = I_k$$
• orthonormal vectors are independent
  (multiply $\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_k u_k = 0$ by $u_i^T$)

• hence $u_1, \ldots, u_k$ is an orthonormal basis for

$$\text{span}(u_1, \ldots, u_k) = \mathcal{R}(U)$$

• **warning**: if $k < n$ then $UU^T \neq I$ (since its rank is at most $k$)
  (more on this matrix later . . . )
**Geometric properties**

suppose columns of \( U = [u_1 \cdots u_k] \) are orthonormal

if \( w = Uz \), then \( ||w|| = ||z|| \)

- multiplication by \( U \) does not change norm
- mapping \( w = Uz \) is *isometric*: it preserves distances
- simple derivation using matrices:

\[
||w||^2 = ||Uz||^2 = (Uz)^T(Uz) = z^T U^T U z = z^T z = ||z||^2
\]
• *inner products* are also preserved: \( \langle U z, U \tilde{z} \rangle = \langle z, \tilde{z} \rangle \)

• if \( w = U z \) and \( \tilde{w} = U \tilde{z} \) then

\[
\langle w, \tilde{w} \rangle = \langle U z, U \tilde{z} \rangle = (U z)^T (U \tilde{z}) = z^T U^T U \tilde{z} = \langle z, \tilde{z} \rangle
\]

• norms and inner products preserved, so *angles* are preserved:

\[
\angle(U z, U \tilde{z}) = \angle(z, \tilde{z})
\]

• thus, multiplication by \( U \) preserves inner products, angles, and distances
Orthonormal basis for $\mathbb{R}^n$

- Suppose $u_1, \ldots, u_n$ is an orthonormal basis for $\mathbb{R}^n$
- Then $U = [u_1 \cdots u_n]$ is called **orthogonal**: it is square and satisfies $U^T U = I$

  (you’d think such matrices would be called *orthonormal*, not *orthogonal*)

- It follows that $U^{-1} = U^T$, and hence also $UU^T = I$, i.e.,

  $$\sum_{i=1}^{n} u_i u_i^T = I$$
suppose $U$ is orthogonal, so $x = UU^T x$, i.e.,

$$x = \sum_{i=1}^{n} (u_i^T x) u_i$$

- $u_i^T x$ is called the component of $x$ in the direction $u_i$

- $a = U^T x$ resolves $x$ into the vector of its $u_i$ components

- $x = U a$ reconstitutes $x$ from its $u_i$ components

- $x = U a = \sum_{i=1}^{n} a_i u_i$ is called the (orthonormal) expansion of $x$
the identity \( I = UU^T = \sum_{i=1}^{n} u_i u_i^T \) is sometimes written (in physics) as

\[
I = \sum_{i=1}^{n} |u_i\rangle\langle u_i|
\]

since

\[
x = \sum_{i=1}^{n} |u_i\rangle\langle u_i|x\rangle
\]

(but we won’t use this notation)
Geometric interpretation

if $U$ is orthogonal, then transformation $w = Uz$

- preserves norm of vectors, i.e., $\|Uz\| = \|z\|$

- preserves angles between vectors, i.e., $\angle(Uz, U\tilde{z}) = \angle(z, \tilde{z})$

examples:

- rotations (about some axis)

- reflections (through some plane)
**Example:** rotation by $\theta$ in $\mathbb{R}^2$ is given by

$$y = U_\theta x, \quad U_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

since $e_1 \rightarrow (\cos \theta, \sin \theta), \ e_2 \rightarrow (-\sin \theta, \cos \theta)$

reflection across line $x_2 = x_1 \tan(\theta/2)$ is given by

$$y = R_\theta x, \quad R_\theta = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix}$$

since $e_1 \rightarrow (\cos \theta, \sin \theta), \ e_2 \rightarrow (\sin \theta, -\cos \theta)$

Orthonormal sets of vectors and $QR$ factorization
can check that $U_\theta$ and $R_\theta$ are orthogonal
Gram-Schmidt procedure

- given independent vectors $a_1, \ldots, a_k \in \mathbb{R}^n$, G-S procedure finds orthonormal vectors $q_1, \ldots, q_k$ s.t.

$$\text{span}(a_1, \ldots, a_r) = \text{span}(q_1, \ldots, q_r) \quad \text{for} \quad r \leq k$$

- thus, $q_1, \ldots, q_r$ is an orthonormal basis for $\text{span}(a_1, \ldots, a_r)$

- rough idea of method: first orthogonalize each vector w.r.t. previous ones; then normalize result to have norm one
Gram-Schmidt procedure

- step 1a. $\tilde{q}_1 := a_1$

- step 1b. $q_1 := \tilde{q}_1/\|\tilde{q}_1\|$ (normalize)

- step 2a. $\tilde{q}_2 := a_2 - (q_1^T a_2)q_1$ (remove $q_1$ component from $a_2$)

- step 2b. $q_2 := \tilde{q}_2/\|\tilde{q}_2\|$ (normalize)

- step 3a. $\tilde{q}_3 := a_3 - (q_1^T a_3)q_1 - (q_2^T a_3)q_2$ (remove $q_1$, $q_2$ components)

- step 3b. $q_3 := \tilde{q}_3/\|\tilde{q}_3\|$ (normalize)

- etc.
for \( i = 1, 2, \ldots, k \) we have

\[
a_i = (q_1^T a_i)q_1 + (q_2^T a_i)q_2 + \cdots + (q_{i-1}^T a_i)q_{i-1} + \|\tilde{q}_i\|q_i = r_{1i}q_1 + r_{2i}q_2 + \cdots + r_{ii}q_i
\]

(note that the \( r_{ij} \)'s come right out of the G-S procedure, and \( r_{ii} \neq 0 \))
**QR decomposition**

written in matrix form: $A = QR$, where $A \in \mathbb{R}^{n \times k}$, $Q \in \mathbb{R}^{n \times k}$, $R \in \mathbb{R}^{k \times k}$:

\[
\begin{bmatrix}
a_1 & a_2 & \cdots & a_k
\end{bmatrix}_A = \begin{bmatrix}
q_1 & q_2 & \cdots & q_k
\end{bmatrix}_Q \begin{bmatrix}
r_{11} & r_{12} & \cdots & r_{1k} \\
0 & r_{22} & \cdots & r_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & r_{kk}
\end{bmatrix}_R
\]

- $Q^T Q = I_k$, and $R$ is upper triangular & invertible
- called **QR decomposition** (or factorization) of $A$
- usually computed using a variation on Gram-Schmidt procedure which is less sensitive to numerical (rounding) errors
- columns of $Q$ are orthonormal basis for $\mathcal{R}(A)$
General Gram-Schmidt procedure

• in basic G-S we assume \( a_1, \ldots, a_k \in \mathbb{R}^n \) are independent

• if \( a_1, \ldots, a_k \) are dependent, we find \( \tilde{q}_j = 0 \) for some \( j \), which means \( a_j \)
  is linearly dependent on \( a_1, \ldots, a_{j-1} \)

• modified algorithm: when we encounter \( \tilde{q}_j = 0 \), skip to next vector \( a_{j+1} \)
  and continue:

\[
\begin{align*}
r &= 0; \\
\text{for } i &= 1, \ldots, k \\
\{ & \\
\tilde{a} &= a_i - \sum_{j=1}^{r} q_j q_j^T a_i; \\
\text{if } \tilde{a} &\neq 0 \{ r = r + 1; q_r = \tilde{a}/\|\tilde{a}\|; \} \\
\} 
\end{align*}
\]
on exit,

- $q_1, \ldots, q_r$ is an orthonormal basis for $\mathcal{R}(A)$ (hence $r = \text{Rank}(A)$)
- each $a_i$ is linear combination of previously generated $q_j$'s

in matrix notation we have $A = QR$ with $Q^TQ = I_r$ and $R \in \mathbb{R}^{r \times k}$ in upper staircase form:

```
  X X X
  \downarrow \downarrow \downarrow
  X X X X X
  \downarrow \downarrow \downarrow \downarrow \downarrow
  X X X X X X

  possibly nonzero entries

  zero entries

  'corner' entries (shown as $\times$) are nonzero
```
can permute columns with $\times$ to front of matrix:

$$A = Q[\tilde{R} \ S]P$$

where:

- $Q^TQ = I_r$
- $\tilde{R} \in \mathbb{R}^{r \times r}$ is upper triangular and invertible
- $P \in \mathbb{R}^{k \times k}$ is a permutation matrix
  (which moves forward the columns of $a$ which generated a new $q$)
Applications

- directly yields orthonormal basis for $\mathcal{R}(A)$
- yields factorization $A = BC$ with $B \in \mathbb{R}^{n \times r}$, $C \in \mathbb{R}^{r \times k}$, $r = \text{Rank}(A)$
- to check if $b \in \text{span}(a_1, \ldots, a_k)$: apply Gram-Schmidt to $[a_1 \cdots a_k \ b]$
- staircase pattern in $R$ shows which columns of $A$ are dependent on previous ones

works incrementally: one G-S procedure yields $QR$ factorizations of $[a_1 \cdots a_p]$ for $p = 1, \ldots, k$:

$$
[a_1 \cdots a_p] = [q_1 \cdots q_s]R_p
$$

where $s = \text{Rank}([a_1 \cdots a_p])$ and $R_p$ is leading $s \times p$ submatrix of $R$
‘Full’ QR factorization

with \( A = Q_1 R_1 \) the QR factorization as above, write

\[
A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}
\]

where \([Q_1 \ Q_2]\) is orthogonal, \(i.e.,\) columns of \(Q_2 \in \mathbb{R}^{n \times (n-r)}\) are orthonormal, orthogonal to \(Q_1\)

to find \(Q_2\):

- find any matrix \(\tilde{A}\) s.t. \([A \ \tilde{A}]\) is full rank (\(e.g., \ \tilde{A} = I\))
- apply general Gram-Schmidt to \([A \ \tilde{A}]\)
- \(Q_1\) are orthonormal vectors obtained from columns of \(A\)
- \(Q_2\) are orthonormal vectors obtained from extra columns (\(\tilde{A}\))
i.e., any set of orthonormal vectors can be extended to an orthonormal basis for $\mathbb{R}^n$

$\mathcal{R}(Q_1)$ and $\mathcal{R}(Q_2)$ are called complementary subspaces since

- they are orthogonal (i.e., every vector in the first subspace is orthogonal to every vector in the second subspace)
- their sum is $\mathbb{R}^n$ (i.e., every vector in $\mathbb{R}^n$ can be expressed as a sum of two vectors, one from each subspace)

this is written

- $\mathcal{R}(Q_1) \perp + \mathcal{R}(Q_2) = \mathbb{R}^n$
- $\mathcal{R}(Q_2) = \mathcal{R}(Q_1)\perp$ (and $\mathcal{R}(Q_1) = \mathcal{R}(Q_2)\perp$)
  
  (each subspace is the orthogonal complement of the other)

we know $\mathcal{R}(Q_1) = \mathcal{R}(A)$; but what is its orthogonal complement $\mathcal{R}(Q_2)$?
Orthogonal decomposition induced by $A$

from $A^T = \begin{bmatrix} R_1^T & 0 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix}$ we see that

$$A^T z = 0 \iff Q_1^T z = 0 \iff z \in \mathcal{R}(Q_2)$$

so $\mathcal{R}(Q_2) = \mathcal{N}(A^T)$

(in fact the columns of $Q_2$ are an orthonormal basis for $\mathcal{N}(A^T)$)

we conclude: $\mathcal{R}(A)$ and $\mathcal{N}(A^T)$ are complementary subspaces:

- $\mathcal{R}(A) \perp \mathcal{N}(A^T) = \mathbb{R}^n$ (recall $A \in \mathbb{R}^{n \times k}$)

- $\mathcal{R}(A)^{\perp} = \mathcal{N}(A^T)$ (and $\mathcal{N}(A^T)^{\perp} = \mathcal{R}(A)$)

- called orthogonal decomposition (of $\mathbb{R}^n$) induced by $A \in \mathbb{R}^{n \times k}$
- every $y \in \mathbb{R}^n$ can be written uniquely as $y = z + w$, with $z \in \mathcal{R}(A)$, $w \in \mathcal{N}(A^T)$ (we’ll soon see what the vector $z$ is . . . )

- can now prove most of the assertions from the linear algebra review lecture

- switching $A \in \mathbb{R}^{n \times k}$ to $A^T \in \mathbb{R}^{k \times n}$ gives decomposition of $\mathbb{R}^k$:

$$\mathcal{N}(A) \perp \mathcal{R}(A^T) = \mathbb{R}^k$$