Outline

1. Inner Product and Orthogonality

2. Orthogonal Projection and Least Squares
Definition

A vector space \( \mathcal{V} \) over \( \mathbb{F} \) (where \( \mathbb{F} \) denotes either \( \mathbb{R} \) or \( \mathbb{C} \)) is said to be an inner-product space if it is equipped with an inner product. An inner product is a function \( \langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{F} \), which satisfies the following properties:

- \( \langle u + w, v \rangle = \langle u, v \rangle + \langle w, v \rangle, \forall u, w, v \in \mathcal{V} \).
- (Homogeneity in first argument): \( \langle \alpha \cdot v, w \rangle = \alpha \langle v, w \rangle, \forall \alpha \in \mathbb{F}, v, w \in \mathcal{V} \).
- \( \langle u, v \rangle = \langle v, u \rangle^*, \forall u, v \in \mathcal{V} \).
- \( \langle u, u \rangle \geq 0, \forall u \in \mathcal{V} \). Moreover, \( \langle u, u \rangle = 0 \) if and only if \( u = 0 \).
**Definition**

Let $F = \mathbb{C}$ or $\mathbb{R}$. A vector space $V$ over $F$ equipped with a norm is called a normed linear space. A norm is a function $\| \cdot \| : V \to \mathbb{R}$ such that

- $\| \alpha \cdot v \| = |\alpha| \| v \|$, $\forall \alpha \in F, v \in V$
- $\| v \| \geq 0$, $\forall v \in V$ and $\| v \| = 0$ iff $v = 0$.
- $\| u + v \| \leq \| u \| + \| v \|$, $\forall u, v \in V$ (*Triangle Inequality*)

Let $x \in V = \mathbb{C}^n$. For each $p \geq 1$, the function

$$\| x \|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

is a norm in $V$ and is denoted as the $p-$ norm of $x$.

The most frequently used norms in $\mathbb{C}^n$ are $\| x \|_1$ (1-norm), $\| x \|_2$ (2-norm) and $\| x \|_\infty \triangleq \max_{1 \leq i \leq n} |x_i|$ (max-norm).
The inner product in $V$ induces a norm $\| \cdot \|$ in $V$ defined as

$$\| \mathbf{x} \| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

1. $V = \mathbb{C}^n$, $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^H \mathbf{u}$ \footnote{In class, we defined inner product as $\mathbf{u}^H \mathbf{v}$. You can choose to follow either convention, as long as you stick to it. The definition used in the notes implies homogeneity in the first argument.} It induces the 2–norm, since $\langle \mathbf{x}, \mathbf{x} \rangle = \| \mathbf{x} \|^2_2$.

2. $V = \mathbb{C}^{m \times n}$, $\langle \mathbf{U}, \mathbf{V} \rangle = \text{Trace}(\mathbf{V}^H \mathbf{U})$. It induces the matrix Frobenius norm, defined as

$$\| \mathbf{X} \|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |X_{i,j}|^2}$$

You can verify that $\| \mathbf{X} \|^2_F = \text{Trace}(\mathbf{X}^H \mathbf{X})$.
1. Given a vector space $\mathcal{V}$ equipped with an inner-product $\langle \cdot, \cdot \rangle$, and corresponding induced norm $\| \cdot \|$,

$$|\langle x, y \rangle| \leq \|x\| \|y\| \quad \forall x, y \in \mathcal{V} \quad \text{(Cauchy-Schwarz Inequality)}$$

Equality holds if and only if $y = \frac{\langle x, y \rangle}{\|x\|^2} x$

2. Given $\mathcal{V} = \mathbb{C}^n$, if $p > 1$ and $q > 1$ are real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$, then

$$|\langle x, y \rangle| \leq \|x\|_p \|y\|_q \quad \forall x, y \in \mathcal{V}$$

3. Given any norm $\| \cdot \|$ in $\mathcal{V}$, $\|x - y\| \geq \|x\| - \|y\|$
Vectors \( \mathbf{x} \) and \( \mathbf{y} \) are orthogonal if \( \langle \mathbf{x}, \mathbf{y} \rangle = 0 \).

A set of non-zero vectors \( \{\mathbf{v}_1, \cdots, \mathbf{v}_K\}, \mathbf{v}_i \in \mathbb{C}^m \) are orthogonal if \( \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0, i \neq j \). In this case, the following are true:

- \( \{\mathbf{v}_1, \cdots, \mathbf{v}_K\} \) are linearly independent.
- \( \| \sum_{i=1}^{K} \mathbf{v}_i \|_2^2 = \sum_{i=1}^{K} \| \mathbf{v}_i \|_2^2 \) (Pythagorean Theorem)
- If \( K = m \), then \( \{\mathbf{v}_1, \cdots, \mathbf{v}_K\} \) constitute an orthogonal basis for \( \mathbb{C}^m \). In this case, any vector \( \mathbf{y} \in \mathbb{C}^m \) can be uniquely represented as a linear combination of \( \{\mathbf{v}_1, \cdots, \mathbf{v}_m\} \) as
  \[
  \mathbf{y} = \sum_{i=1}^{m} \alpha_i \mathbf{v}_i
  \]
  where
  \[
  \alpha_i = \frac{\langle \mathbf{y}, \mathbf{v}_i \rangle}{\| \mathbf{v}_i \|_2^2}
  \]
- If \( \| \mathbf{v}_i \|_2 = 1, \forall i = 1, \cdots K \), then \( \| \sum_{i=1}^{K} \alpha_i \mathbf{v}_i \|_2^2 = \sum_{i=1}^{K} |\alpha_i|^2 \). (Generalized form of Parseval’s Theorem)
Orthogonal Projection

Given an inner product space $\mathcal{V}$ and a subspace $\mathcal{W} \subseteq \mathcal{V}$, the orthogonal projection of a vector $y \in \mathcal{V}$ onto $\mathcal{W}$ is defined as a vector $\hat{y}_\mathcal{W} \in \mathcal{W}$ such that

$$\langle y - \hat{y}_\mathcal{W}, w \rangle = 0, \quad \forall w \in \mathcal{W}$$

1. The projection $\hat{y}_\mathcal{W}$ exists and is unique if $\mathcal{V}$ is a complete inner product space (also known as Hilbert Space). Easy to show this when $\mathcal{V} = \mathbb{C}^n$ or $\mathbb{R}^n$.
2. It can be shown that the projection of $y$ onto a vector $v_i$ is given by $\frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i$.
3. $\|y - \hat{y}_\mathcal{W}\|_2 \leq \|y - w\|_2 \quad \forall w \in \mathcal{W}$. 
Least Squares

Given $A = [a_1, \cdots a_n]$, consider the following minimization problem:

$$\min_{x \in \mathbb{C}^n} \| y - Ax \|_2 \quad \text{(Least Squares)} \quad (1)$$

**Orthogonality Principle:** $x^*$ is a minimizer of (1) if and only if it satisfies the following orthogonality principle:

$$\langle y - Ax^*, a_i \rangle = 0, \quad i = 1, 2, \cdots, n$$

- The orthogonality principle implies that $Ax^*$ is the orthogonal projection of $y$ onto $\mathcal{R}(A)$.
- There could be multiple minimizers $x^*$ of (1) but $Ax^*$ is unique.
Inner Product and Orthogonality

Orthogonal Projection and Least Squares

Normal Equations

The orthogonality principle can be succinctly represented in terms of the following system of equations, known as the normal equations:

\[
\begin{pmatrix}
A^H A
\end{pmatrix} x^* = A^H y
\]  

(2)

1. If \( A \) has full column rank of \( n \), then \( A^H A \) is invertible. In this case, the solution \( x^* \) to (1) is unique and is given by

\[
x^* = \left( A^H A \right)^{-1} A^H y
\]

implying

\[
\hat{y}_{R(A)} = A \left( A^H A \right)^{-1} A^H y
\]

The matrix \( A \left( A^H A \right)^{-1} A^H \) is known as the projection matrix (for \( R(A) \))

2. If \( \text{rank}(A) < n \), there are multiple \( x^* \) satisfying (2).
If $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n \in \mathbb{C}^m$ are orthogonal, the projection of $\mathbf{y} \in \mathbb{C}^m$ onto $\mathcal{W} \triangleq \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$ is given by

$$
\hat{\mathbf{y}}_\mathcal{W} = \sum_{i=1}^{n} \frac{\mathbf{a}_i^H \mathbf{y}}{\|\mathbf{a}_i\|^2} \mathbf{a}_i
$$

Additionally, if $\|\mathbf{a}_i\|_2 = 1$, $i = 1, 2, \cdots, n$, then

$$
\mathbf{A}^H \mathbf{A} = \mathbf{I}_{n \times n}
$$

In this case, the matrix $\mathbf{A} \mathbf{A}^H$ represents the projection matrix, i.e. $\hat{\mathbf{y}}_\mathcal{W} = \mathbf{A} \mathbf{A}^H \mathbf{y}$
Gram-Schmidt Orthogonalization

**Goal:** Given a set \( \{ \mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n \} \) of linearly independent vectors in an inner product space \( \mathcal{V} \), construct another set of vectors \( \{ \mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n \} \) in \( \mathcal{V} \) such that

- \( \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\} = \text{Span}\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n\} \)
- \( \langle \mathbf{q}_i, \mathbf{q}_j \rangle = 0, \quad i \neq j, \quad \| \mathbf{q}_i \| = 1 \)

In other words, \( \{ \mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n \} \) constitute an orthonormal basis for \( \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\} \).

Gram-Schmidt Orthogonalization is a systematic procedure to generate the orthonormal basis vectors *sequentially*. 
Gram-Schmidt Orthogonalization constructs \( \{q_1, q_2, \cdots, q_n\} \) sequentially as follows:

\[
q_1 = \frac{a_1}{\|a_1\|_2}
\]

\cdots

\[
q_k = \frac{a_k - \sum_{i=1}^{k-1} \langle a_k, q_i \rangle q_i}{\|a_k - \sum_{i=1}^{k-1} \langle a_k, q_i \rangle q_i\|_2}
\]

\cdots

\[
q_n = \frac{a_n - \sum_{i=1}^{n-1} \langle a_n, q_i \rangle q_i}{\|a_n - \sum_{i=1}^{k-1} \langle a_n, q_i \rangle q_i\|_2}
\]

For each integer \( k \), \( \{q_1, \cdots, q_k\} \) is an orthonormal basis for \( \text{Span}\{a_1, \cdots, a_k\} \).
Any matrix $A \in \mathbb{C}^{m \times n}$ with linearly independent columns can be uniquely factorized as

$$A = QR$$

where $Q \in \mathbb{C}^{m \times n}$ has orthonormal columns and $R$ is an upper-triangular matrix with positive diagonal elements.

- Columns of $Q$ are given by the orthonormal vectors $q_1, \ldots, q_n$ produced by the Gram-Schmidt Orthonogalization.
- The elements of $R$ are given by

$$R_{i,j} = \begin{cases} 
\mu_i & i = j \\
\langle a_j, q_i \rangle & j > i \\
0 & \text{else}
\end{cases}$$

where $\mu_1 = \|a_1\|$ and $\mu_j = \|a_j - \sum_{i=1}^{j-1} \langle a_j, q_i \rangle q_i\|_2, j > 1$. 
Consider the Least squares problem \( \min_{x \in \mathbb{R}^n} \| y_i - A_i x \|_2 \) where \( A_i \in \mathbb{R}^{i \times n} \) has \( n \) independent columns. The solution is given by

\[
x_i = (A_i^T A_i)^{-1} A_i^T y_i
\]

Suppose we solve a slightly larger least squares problem by appending one more row \( a_{i+1}^T \in \mathbb{R}^{1 \times n} \) to \( A_i \) and define

\[
A_{i+1} = [A_i^T, a_{i+1}]^T
\]
\[
y_{i+1} = [y_i^T, y_{i+1}]^T
\]

Consider solving the following Least Squares problem

\[
x_{i+1} = \arg \min_{x \in \mathbb{R}^n} \| y_{i+1} - A_{i+1} x \|_2
\]
Can the computation of $x_{i+1}$ benefit from the solution of
$\min_{x \in \mathbb{R}^n} \|y_i - A_i x\|_2$?

The most computationally intensive step is to compute the inverse
$P_{i+1} = (A_{i+1}^T A_{i+1})^{-1}$ which can be simplified by using the
pre-computed inverse $P_i = (A_i^T A_i)^{-1}$ and using the Matrix
Inversion Lemma. Notice that

$$P_{i+1} = \left( A_i^T A_i + a_{i+1} a_{i+1}^T \right)^{-1}$$

Apply the Matrix Inversion Lemma to compute $P_{i+1}$ using $P_i$. 