Chapter 6. Orthogonality and Least Squares

6.1 Inner Product, Length, and Orthogonality
Conjugate transpose. If $A \in \mathbb{C}^{m \times n}$, then we define

$$A^* = (\bar{A})^T \in \mathbb{C}^{n \times m}.$$  

We call $A^*$ the conjugate transpose or the adjoint of $A$. If $A \in \mathbb{R}^{m \times n}$, then $A^* = A^T$.

Note that

- $(\alpha A + \beta B)^* = \bar{\alpha} A^* + \bar{\beta} B^*$
- $(AC)^* = C^* A^*$.

Definition. $A \in \mathbb{C}^{n \times n}$ is hermitian if $A^* = A$.

Note that $A \in \mathbb{R}^{n \times n}$ is hermitian if and only if it is symmetric, i.e. $A = A^T$.

Definition. If $u = (a_1, \ldots, a_n) \in \mathbb{C}^n$ and $v = (b_1, \ldots, b_n) \in \mathbb{C}^n$, then we define the inner product of $u$ and $v$ by

$$u \cdot v = \bar{a}_1 b_1 + \cdots + \bar{a}_n b_n \in \mathbb{C}.$$  

Note that if we regard $u, v$ as $n \times 1$ matrices, then $u \cdot v = u^* v$.

Note also that for $A = [u_1 \ldots u_k] \in \mathbb{C}^{n \times k}$ and $B = [v_1 \ldots v_\ell] \in \mathbb{C}^{n \times \ell}$, then $A^* B \in \mathbb{C}^{k \times \ell}$,

$$A^* B = [u_i v_j] \quad i = 1, \ldots, k, \quad j = 1, \ldots, \ell.$$  

Properties of the inner product. For $u, v, w \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$:

- $u \cdot v = v \cdot u$
- $u \cdot (v + w) = u \cdot v + u \cdot w$
- $\alpha (u \cdot v) = (\overline{\alpha}u) \cdot v = u \cdot (\alpha v)$
- If $u = (a_1, \cdots, a_n) \in \mathbb{C}^n$, then
  \[ u \cdot u = |a_1|^2 + \cdots + |a_n|^2 \geq 0, \]
  and $u \cdot u = 0$ only if $u = 0$.

Definition. If $u = (a_1, \ldots, a_n) \in \mathbb{C}^n$, then the norm of $u$ is given by

\[ \|u\| = \sqrt{u \cdot \overline{u}} = \sqrt{|a_1|^2 + \cdots + |a_n|^2}. \]

Properties. For $u, v \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$,

- $\|\alpha u\| = |\alpha| \|u\|
- $|u \cdot v| \leq \|u\| \|v\|$ (Cauchy–Schwarz inequality)
- $\|u + v\| \leq \|u\| + \|v\|$ (triangle inequality)

The norm measures length; $\|u - v\|$ measures the distance between $u$ and $v$.

A vector $u \in \mathbb{C}^n$ is a unit vector if $\|u\| = 1$. 
Example. Let \( A = [v_1 \ v_2] = \begin{bmatrix} 1 & i \\ 3 + 8i & 2i \end{bmatrix} \).

Then \( A^* = \begin{bmatrix} 1 & 3 - 8i \\ -i & -2i \end{bmatrix} \). So \( A \) is not hermitian.

We have \( v_1 \cdot v_2 = 1 \cdot i + (3 - 8i) \cdot 2i = 16 + 7i \).

Note \( \|v_1\|^2 = 1 \cdot 1 + (3 - 8i)(3 + 8i) = 74 \).

Consequently, \( \frac{1}{\sqrt{74}} v_1 \) is a unit vector.

Definition. Two vectors \( u, v \in \mathbb{C}^n \) are orthogonal if \( u \cdot v = 0 \). We write \( u \perp v \).

A set \( \{v_1, \ldots, v_k\} \subset \mathbb{C}^n \) is an orthogonal set if \( v_i \cdot v_j = 0 \) for each \( i, j = 1, \ldots, k \) (with \( i \neq j \)).

A set \( \{v_1, \ldots, v_k\} \subset \mathbb{C}^n \) is an orthonormal set if it is orthogonal and each \( v_i \) is a unit vector.

Remark. In general, we have
\[
\|u + v\| \leq \|u\| + \|v\|.
\]

However, we have
\[
u \perp v \implies \|u + v\|^2 = \|u\|^2 + \|v\|^2.
\]

This is the Pythagorean theorem.
**Definition.** Let \( W \subset \mathbb{C}^n \). The **orthogonal complement** of \( W \), denoted \( W^\perp \), is defined by

\[
W^\perp = \{ v \in \mathbb{C}^n : v \cdot w = 0 \text{ for every } w \in W \}.
\]

**Subspace property.** \( W^\perp \) is a subspace of \( \mathbb{C}^n \) satisfying \( W \cap W^\perp = \{ 0 \} \).

Indeed, \( W^\perp \) is closed under addition and scalar multiplication, and \( w \cdot w = 0 \implies w = 0 \).

Suppose \( A = [v_1 \cdots v_k] \in \mathbb{C}^{n \times k} \). Then

\[
[\text{col}(A)]^\perp = \text{nul}(A^*)
\]

Indeed

\[
0 = A^* x = \begin{bmatrix} v_1^* \cdots v_k^* \end{bmatrix} \begin{bmatrix} x \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} v_1^* x \\ \vdots \\ v_k^* x \end{bmatrix} = \begin{bmatrix} v_1 \cdot x \\ \vdots \\ v_k \cdot x \end{bmatrix}
\]

if and only if

\[
v_1 \cdot x = \cdots = v_k \cdot x = 0.
\]
Example 1. Let \( \mathbf{v}_1 = [1, -1, 2]^T \) and \( \mathbf{v}_2 = [0, 2, 1]^T \). Note that \( \mathbf{v}_1 \perp \mathbf{v}_2 \).

Let \( W = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \) and \( A = [\mathbf{v}_1 \mathbf{v}_2] \). Note that

\[
W^\perp = [\text{col}(A)]^\perp = \text{nul}(A^*) = \text{nul}(A^T).
\]

We have

\[
A^T = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & \frac{5}{2} \\ 0 & 1 & \frac{1}{2} \end{bmatrix},
\]

and thus \( \text{nul}(A^T) = \text{span}\{\mathbf{v}_3\} = \text{span}\{[-\frac{5}{2}, -\frac{1}{2}, 1]^T\} \).

Note \( \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \) is an orthogonal set, and \( W^\perp \) is a line perpendicular to the plane \( W \).

Example 2. Let \( \mathbf{v}_1 = [1, -1, 1, -1]^T \) and \( \mathbf{v}_2 = [1, 1, 1, 1]^T \). Again, \( \mathbf{v}_1 \cdot \mathbf{v}_2 = 0 \).

Let \( W = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \) and \( A = [\mathbf{v}_1 \mathbf{v}_2] \) as before. Then \( W^\perp = \text{nul}(A^T) \), with

\[
A^T = \begin{bmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.
\]

In particular, \( W^\perp = \text{span}\{\mathbf{v}_3, \mathbf{v}_4\} \), with

\[
\mathbf{v}_3 = [-1, 0, 1, 0]^T \quad \mathbf{v}_4 = [0, -1, 0, 1]^T.
\]

Again, \( \{\mathbf{v}_1, \ldots, \mathbf{v}_4\} \) is an orthogonal set. This time \( W \) and \( W^\perp \) are planes in \( \mathbb{R}^4 \), with \( W \cap W^\perp = \{0\} \).
Chapter 6. Orthogonality and Least Squares

6.2 Orthogonal Sets

Definition. If \( S \) is an orthogonal set that is linearly independent, then we call \( S \) an **orthogonal basis** for \( \text{span}(S) \).

Similarly, a linearly independent orthonormal set \( S \) is a **orthonormal basis** for \( \text{span}(S) \).

Example. Let

\[
\begin{bmatrix}
1 \\
0 \\
1
\end{bmatrix}, \quad \begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}, \quad \begin{bmatrix}
-1 \\
0 \\
1
\end{bmatrix}.
\]

\( S = \{v_1, v_2, 0\} \) is an orthogonal set, but not a basis for \( \mathbb{C}^3 \)

\( S = \{v_1, v_3\} \) is an orthogonal basis for \( \mathbb{C}^3 \)

\( S = \{\frac{1}{\sqrt{2}}v_1, v_2, \frac{1}{\sqrt{2}}v_3\} \) is an orthonormal basis for \( \mathbb{C}^3 \)
**Test for orthogonality.** Let $A = [v_1 \cdots v_p] \in \mathbb{C}^{n \times p}$. Note that

$$A^*A = \begin{bmatrix} v_1 \cdot v_1 & \cdots & v_1 \cdot v_p \\ \vdots & \ddots & \vdots \\ v_p \cdot v_1 & \cdots & v_p v_p \end{bmatrix} \in \mathbb{C}^{p \times p}.$$

Thus $A^*A$ is diagonal precisely when $\{v_1, \ldots, v_p\}$ is orthogonal.

Furthermore, $\{v_1, \ldots, v_p\}$ is orthonormal precisely when $A^*A = I_p$.

**Definition.** A matrix $A \in \mathbb{C}^{n \times n}$ is **unitary** if $A^*A = I_n$.

The following conditions are equivalent:

- $A \in \mathbb{C}^{n \times n}$ is unitary
- $A \in \mathbb{C}^{n \times n}$ satisfies $A^{-1} = A^*$
- the columns of $A$ are an orthonormal basis for $\mathbb{C}^n$
- $A \in \mathbb{C}^{n \times n}$ satisfies $AA^* = I_n$
- the rows of $A$ are an orthonormal basis for $\mathbb{C}^n$
**Theorem.** (Independence) If $S = \{v_1, \ldots, v_p\}$ is an orthogonal set of non-zero vectors, then $S$ is independent and $S$ is a basis for $\text{span}(S)$.

Indeed, suppose  

$$c_1v_1 + \cdots + c_pv_p = 0.$$ 

Now take an inner product with $v_j$:

$$0 = c_1v_1 \cdot v_j + \cdots + c_jv_j \cdot v_j + \cdots + c_pv_p \cdot v_j$$

$$= 0 + \cdots + c_j\|v_j\|^2 + \cdots + 0.$$ 

Thus $c_j = 0$ for any $j = 1, \ldots, p$.

**Theorem.** If $S = \{w_1, \ldots, w_p\} \subset W$ are independent and $T = \{v_1, \ldots, v_q\} \subset W^\perp$ are independent, then $S \cup T$ is independent.

**Theorem.** Suppose $W \subset \mathbb{C}^n$ has dimension $p$. Then $\dim(W^\perp) = n - p$.

Let $A = [w_1 \cdots w_p]$, where $\{w_1, \ldots, w_p\}$ is a basis for $W$. Note

$$W^\perp = [\text{col}(A)]^\perp = \text{nul}(A^*) \subset \mathbb{C}^n.$$ 

Thus

$$\dim(W^\perp) = \dim(\text{nul}(A^*)) = n - \text{rank}(A^*) = n - \text{rank}(A) = n - p.$$ 

In particular, we find

$$\dim(W) + \dim(W^\perp) = \dim(\mathbb{C}^n) = n.$$ 

**Remark.** If $B = \{w_1, \ldots, w_p\}$ is a basis for $W$ and $C = \{v_1, \ldots, v_{n-p}\}$ is a basis for $W^\perp$, then $B \cup C$ is a basis for $\mathbb{C}^n$. 
**Theorem.** (Orthogonal decomposition) Let $W$ be a subspace of $\mathbb{C}^n$. For every $x \in \mathbb{C}^n$ there exist unique $y \in W$ and $z \in W^\perp$ such that $x = y + z$.

Indeed, let $B = \{w_1, \ldots, w_p\}$ be a basis for $W$ and $C = \{v_1, \ldots, v_{n-p}\}$ a basis for $W^\perp$. Then $B \cup C$ is a basis for $\mathbb{C}^n$, and so every $x$ has a unique representation $x = y + z$, where $y \in \text{span}(B)$ and $z \in \text{span}(C)$.

Uniqueness can also be deduced from the fact that $W \cap W^\perp = \{0\}$.

**Remark.** Suppose $B$ is an orthogonal basis of $W$. Then

$$x = \alpha_1 w_1 + \cdots + \alpha_p w_p + z, \quad z \in W^\perp.$$ 

One can compute $\alpha_j$ via

$$w_j \cdot x = \alpha_j w_j \cdot w_j \implies \alpha_j = \frac{w_j \cdot x}{\|w_j\|^2}.$$ 

**Projection.** Let $W$ be a subspace of $\mathbb{C}^n$. As above, for each $x \in \mathbb{C}^n$ there exists a unique $y \in W$ and $z \in W^\perp$ so that $x = y + z$. We define

$$\text{proj}_W : \mathbb{C}^n \to W \subset \mathbb{C}^n \quad \text{by} \quad \text{proj}_W x = y.$$ 

We call $\text{proj}_W$ the (orthogonal) projection of $\mathbb{C}^n$ onto $W$.

**Example.** Suppose $W = \text{span}\{w_1\}$. Then

$$\text{proj}_W x = \frac{w_1 \cdot x}{\|w_1\|^2} w_1.$$ 

Note that $\text{proj}_W$ is a linear transformation, with matrix representation given by

$$[\text{proj}_W] = \frac{1}{\|w_1\|^2} w_1 w_1^* \in \mathbb{C}^{n \times n}.$$
Example. Let $w_1 = [1, 0, 1]^T$ and $v = [-1, 2, 2]^T$, with $W = \text{span}\{w_1\}$. Then

$$\text{proj}_W(v) = \frac{w_1 \cdot v}{\|w_1\|^2} w_1 = \frac{1}{2} w_1.$$ 

In fact,

$$[\text{proj}_W]_E = \frac{1}{\|w_1\|^2} w_1 w_1^* = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$ 

Thus

$$\text{proj}_W(x) = \frac{1}{2} \begin{bmatrix} x_1 + x_3 \\ 0 \\ x_1 + x_3 \end{bmatrix}.$$
Orthogonal projections. Let $W$ be a subspace of $\mathbb{C}^n$. Recall that
\[ \operatorname{proj}_W x = y, \quad \text{where} \quad x = y + z, \quad y \in W, \quad z \in W^\perp. \]

Let $B = \{w_1, \ldots, w_p\}$ be a basis for $W \subset \mathbb{C}^n$. We wish to find the $B$-coordinates of $\operatorname{proj}_W x$, i.e. to write
\[ x = \alpha_1 w_1 + \cdots + \alpha_p w_p + z, \quad z \in W^\perp. \]

This yields a system of $p$ equations and $p$ unknowns:
\[ w_1 \cdot x = \alpha_1 w_1 \cdot w_1 + \cdots + \alpha_p w_1 \cdot w_p \]
\[ \vdots \quad \vdots \quad \vdots \]
\[ w_p \cdot x = \alpha_1 w_p \cdot w_1 + \cdots + \alpha_p w_p \cdot w_p. \]

Normal system. Write $A = [w_1 \cdots w_p] \in \mathbb{C}^{n \times p}$. The system
\[ w_1^* x = \alpha_1 w_1^* w_1 + \cdots + \alpha_p w_1^* w_p \]
\[ \vdots \quad \vdots \quad \vdots \]
\[ w_p^* x = \alpha_1 w_p^* w_1 + \cdots + \alpha_p w_p^* w_p \]

may be written as the normal system $A^* A \hat{x} = A^* x$, where
\[ \hat{x} = [\operatorname{proj}_W(x)]_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}. \]

One calls $A^* A \in \mathbb{C}^{p \times p}$ the Gram matrix.

- The normal system has at least one solution, namely $[\operatorname{proj}_W(x)]_B$.
- If the normal system has a unique solution, it is $[\operatorname{proj}_W(x)]_B$. 

Theorem. (Null space and rank of $A^*A$) If $A \in \mathbb{C}^{n \times p}$, then $A^*A \in \mathbb{C}^{p \times p}$ satisfies

$$\text{nul}(A^*A) = \text{nul}(A) \quad \text{and} \quad \text{rank}(A^*A) = \text{rank}(A).$$

- First note $\text{nul}(A) \subset \text{nul}(A^*A)$.
- If instead $Ax \in \text{nul}(A^*) = [\text{col}(A)]^\perp$, then $Ax \in \text{col}(A) \cap \text{col}(A)^\perp$ and hence $Ax = 0$.
  Thus $\text{nul}(A^*A) \subset \text{nul}(A)$.
- Thus

$$\text{rank}(A^*A) = p - \dim(\text{nul}(A^*A)) = p - \dim(\text{nul}(A)) = \text{rank}(A).$$

**Solving the normal system.** If the columns of $A$ are independent, then $A^*A \in \mathbb{C}^{p \times p}$ and $\text{rank}(A^*A) = \text{rank}(A) = p$ and hence $A^*A$ is invertible.

---

**Solving the normal system.** Suppose $B = \{w_1, \ldots, w_p\}$ is a basis for a subspace $W \subset \mathbb{C}^n$. Writing $A = [w_1, \ldots, w_p]$, we have that $A^*A$ is invertible and the normal system

$$A^*A\hat{x} = A^*x$$

has a unique solution

$$\hat{x} = [\text{proj}_W(x)]_B = (A^*A)^{-1}A^*x.$$

We can then obtain $\text{proj}_W(x)$ via

$$\text{proj}_W(x) = A[\text{proj}_W(x)]_B = A\hat{x} = A(A^*A)^{-1}A^*x.$$
Example 1. If \( p = 1 \) (so \( W \) is a line), then

\[
A^*A\hat{x} = A^*x \implies [w_1 \cdot w_1]\hat{x} = w_1 \cdot x,
\]

leading again to

\[
\text{proj}_W(x) = \frac{w_1 \cdot x}{\|w_1\|^2}w_1.
\]

Example 2. If \( p > 1 \) and \( B = \{w_1, \ldots, w_p\} \) is an orthogonal basis then

\[
A^*A = \text{diag}\{\|w_1\|^2, \ldots, \|w_p\|^2\}.
\]

Recalling that \( \hat{x} = (A^*A)^{-1}A^*x \), we find

\[
[\text{proj}_W(x)]_B = \begin{bmatrix}
\frac{w_1 \cdot x}{\|w_1\|^2} \\
\frac{w_2 \cdot x}{\|w_2\|^2} \\
\vdots \\
\frac{w_p \cdot x}{\|w_p\|^2}
\end{bmatrix},
\]

Thus

\[
\text{proj}_W(x) = \frac{w_1 \cdot x}{\|w_1\|^2}w_1 + \cdots + \frac{w_p \cdot x}{\|w_p\|^2}w_p.
\]

Example. Let

\[
w_1 = \begin{bmatrix}
1 \\
2 \\
1
\end{bmatrix}, \quad w_2 = \begin{bmatrix}
-2 \\
1 \\
-1
\end{bmatrix}, \quad x = \begin{bmatrix}
4 \\
5 \\
-3 \\
3
\end{bmatrix}.
\]

Set \( B = \{w_1, w_2\} \) and \( A = [w_1 \ w_2] \). Then \( A^*A = \text{diag}\{7, 7\} \). Thus

\[
\hat{x} = [\text{proj}_W(x)]_B = (A^*A)^{-1}A^*x = \begin{bmatrix}
2 \\
3
\end{bmatrix},
\]

\[
\text{proj}_W(x) = 2w_1 + \frac{3}{7}w_2.
\]

The projection of \( x \) onto \( W^\perp \) is simply

\[
\text{proj}_{W^\perp}(x) = x - \text{proj}_W(x).
\]
Example 2. If $A^*A$ is not diagonal, then the columns of $A$ are not an orthogonal basis for $\text{col}(A)$.

One can still compute the projection via

$$\text{proj}_W(x) = A(A^*A)^{-1}A^*x.$$ 

Distance minimization. Orthogonal projection is related to minimizing a distance. To see this, suppose $w \in W$ and $x \in \mathbb{C}^n$. By the Pythagorean theorem,

$$\|x - w\|^2 = \|\text{proj}_{W^\perp}(x)\|^2 + \|\text{proj}_W(x) - w\|^2,$$

and thus

$$\min_{w \in W} \|x - w\| = \|x - \text{proj}_W(x)\| = \|\text{proj}_{W^\perp}(x)\|.$$

Example. Let $W = \text{span}\{w_1, w_2\} \subset \mathbb{R}^3$. Then $\|\text{proj}_{W^\perp}(x)\|$ is the distance from $x$ to the plane spanned by $w_1$ and $w_2$. 

Conclusion. Let $B = \{w_1, \ldots, w_p\}$ be a basis for $W \subset \mathbb{C}^n$, 
$A = [w_1 \cdots w_p]$, and $x \in \mathbb{C}^n$.

- $\text{nul}(A^*A) = \text{nul}(A)$, $\text{rank}(A^*A) = \text{rank}(A) = p$, and so $A^*A$ is invertible.
- The solution to $A^*Ax = A^*x$ is $\hat{x} = (A^*A)^{-1}A^*x$.
- $\hat{x} = (A^*A)^{-1}A^*x = [\text{proj}_W(x)]_B$.
- $\text{proj}_W : \mathbb{C}^n \rightarrow W$ is given by $\text{proj}_W(x) = A(A^*A)^{-1}A^*x$.
- $\text{proj}_W(x) = x - \text{proj}_W(x)$.
- $x = \text{proj}_W(x) + \text{proj}_{W^\perp}(x)$.
- If $B$ is orthogonal, $\text{proj}_W(x) = \frac{w_1 \cdot x}{\|w_1\|^2}w_1 + \cdots + \frac{w_p \cdot x}{\|w_p\|^2}w_p$.
- $\min_{w \in W} \|x - w\| = \|x - \text{proj}_W(x)\|$.

Chapter 6. Orthogonality and Least Squares

6.4 The Gram–Schmidt Process
**Orthogonal projections.** Recall that if $B = \{w_1, \ldots, w_p\}$ is an independent set and $A = [w_1 \cdots w_p]$, then
\[
\text{proj}_W(x) = A(A^* A)^{-1} A^* x, \quad W = \text{span}(B)
\]
If $B$ is orthogonal, then
\[
\text{proj}_W(x) = \frac{w_1 \cdot x}{\|w_1\|^2} w_1 + \cdots + \frac{w_p \cdot x}{\|w_p\|^2} w_p.
\]
This may be written
\[
\text{proj}_W(x) = \text{proj}_{W_1}(x) + \cdots + \text{proj}_{W_p}(x), \quad W_j = \text{span}\{w_j\}.
\]
If $B$ is not orthogonal, then we may apply an algorithm to $B$ to obtain an orthogonal basis for $W$.

**Gram-Schmidt algorithm.** Let $A = \{w_1, \ldots, w_p\}$.

Let $v_1 := w_1$ and $\Omega_1 := \text{span}\{v_1\}$.

Let $v_2 = \text{proj}_{\Omega_1^\perp}(w_2) = w_2 - \text{proj}_{\Omega_1}(w_2)$, $\Omega_2 := \text{span}\{v_1, v_2\}$

\[\ldots\]

Let $v_{j+1} = \text{proj}_{\Omega_j^\perp}(w_{j+1})$, $\Omega_{j+1} := \text{span}\{v_1, \ldots, v_{j+1}\}$

Here $j = 1, \ldots, p - 1$. This generates a pairwise orthogonal set $B = \{v_1, \ldots, v_p\}$ with $\text{span}(B) = \text{span}(A)$. Note that
\[
v_{j+1} = 0 \iff w_{j+1} \in \Omega_j.
\]
Matrix representation. Write $V_i = \text{span} \{ v_i \}$. Since $\{ v_i \}$ are orthogonal, we can write

$$\text{proj}_{\Omega_j}(w_{j+1}) = \sum_{k=1}^{j} \text{proj}_{V_k}(w_{j+1}) = \sum_{k=1}^{j} r_{k,j+1} v_k,$$

where $r_{k,j+1} = \frac{v_k \cdot w_{j+1}}{\|v_k\|^2}$ if $v_k \neq 0$ and $r_{k,j+1}$ can be anything if $v_k = 0$.

Thus, using $v_{j+1} = w_{j+1} - \sum_{k=1}^{j} r_{k,j+1} v_k$, we find

$$w_{j+1} = [v_1 \cdots v_{j+1}] \begin{bmatrix} r_{1,j+1} \\ \vdots \\ r_{j,j+1} \\ 1 \end{bmatrix}, \quad j = 1, \ldots, p - 1.$$

Matrix representation (continued). The Gram-Schmidt algorithm therefore has the matrix representation

$$[w_1 \cdots w_p] = [v_1 \cdots v_p] R,$$

where

$$R = \begin{bmatrix} 1 & r_{1,2} & \cdots & r_{1,p} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_{p-1,p} \\ 0 & \cdots & \cdots & 1 \end{bmatrix}$$

This shows that any matrix $A = [w_1 \cdots w_p] \in \mathbb{C}^{n \times p}$ may be factored as $A = QR$, where the columns of $Q$ are orthogonal and $R \in \mathbb{C}^{p \times p}$ is an invertible upper triangular matrix.

The non-zero vectors in $\{ v_1, \ldots, v_j \}$ form an orthogonal basis for $\Omega_j$.

$R$ is unique when each $v_j$ is non-zero.
Example 1. Let
\[
\begin{align*}
\mathbf{w}_1 &= [1, 0, 1, 0]^T, \quad \mathbf{w}_2 = [1, 1, 1, 1]^T, \\
\mathbf{w}_3 &= [1, -1, 1, -1]^T, \quad \mathbf{w}_4 = [0, 0, 1, 1]^T.
\end{align*}
\]
We apply Gram–Schmidt:
\[
\begin{align*}
\mathbf{v}_1 &= \mathbf{w}_1, \quad \Omega_1 = \text{span}\{\mathbf{v}_1\} \\
\mathbf{v}_2 &= \mathbf{w}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 = [0, 1, 0, 1]^T, \quad \Omega_2 = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \\
\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\|\mathbf{v}_2\|^2} \mathbf{v}_2 = 0, \quad \Omega_3 = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \\
\mathbf{v}_4 &= \mathbf{w}_4 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_4}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_4}{\|\mathbf{v}_2\|^2} \mathbf{v}_2 = \frac{1}{2} [-1, -1, 1, 1]^T \\
\Omega_4 &= \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}.
\end{align*}
\]
In particular, \(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}\) is an orthogonal basis for \(\text{span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\}\).

Example 1. (cont.) Let
\[
\mathbf{A} = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4]\text{ and } \mathbf{R} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4].
\]
Then we can write \(\mathbf{A} = \mathbf{QR}\), where
\[
\mathbf{R} = \begin{bmatrix}
1 & 1 & 1 & 1/2 \\
0 & 1 & -1 & 1/2 \\
0 & 0 & 1 & c \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
for any \(c\). These coefficients are determined by evaluating the inner products above, cf.
\[
r_{k,j+1} = \frac{\mathbf{v}_k \cdot \mathbf{w}_{j+1}}{\|\mathbf{v}_k\|^2} \quad \text{if } \mathbf{v}_k \neq 0.
\]
Example 2. Let

\[ \mathbf{w}_1 = [1, 1, 1, 0]^T, \quad \mathbf{w}_2 = [0, 1, -1, 1]^T. \]

Note \( \mathbf{w}_1 \perp \mathbf{w}_2 \). Extend \( \{\mathbf{w}_1, \mathbf{w}_2\} \) to an orthogonal basis for \( \mathbb{C}^4 \).

Write \( A = [\mathbf{w}_1 \mathbf{w}_2] \) and \( W = \text{col}(A) \).

First want a basis for \( W^\perp = \text{nul}(A^\ast) \). Since

\[ A^\ast = A^T = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 1 \end{bmatrix}, \]

we get \( \text{nul}(A^\ast) = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\} \), where

\[ \mathbf{x}_1 = [-2, 1, 1, 0]^T, \quad \mathbf{x}_2 = [1, -1, 0, 1]^T. \]

Now \( \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{x}_1, \mathbf{x}_2\} \) is a basis for \( \mathbb{C}^4 \), but not orthogonal.

---

Example 2. (Cont.) We now apply Gram–Schmidt to \( \{\mathbf{x}_1, \mathbf{x}_2\} \).

- \( \mathbf{v}_1 = \mathbf{x}_1 = [-2, 1, 1, 0]^T \)
- \( \mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{||\mathbf{v}_1||^2} \mathbf{v}_1 = [0, -\frac{1}{2}, \frac{1}{2}, 1]^T \)

Thus \( B = \{\mathbf{w}_1, \mathbf{w}_2\} \) is an orthogonal basis for \( W = \text{col}(A) \), \( C = \{\mathbf{v}_1, \mathbf{v}_2\} \) is an orthogonal basis for \( W^\perp \), and \( B \cup C \) is an orthogonal basis for \( \mathbb{C}^4 \).
Chapter 6. Orthogonality and Least Squares

6.5 Least-Squares Problems

The normal system. For $A \in \mathbb{C}^{n \times p}$ and $b \in \mathbb{C}^n$, the equation

$$A^* A x = A^* b$$

is called the normal system for $Ax = b$.

- The normal system arose when computing the orthogonal projection onto a subspace, where the columns of $A$ were assumed to be a basis $B = \{w_1, \ldots, w_p\}$ for $W = \text{col}(A)$.
- The system was $A^* A \hat{x} = A^* x$, with solution

$$\hat{x} = [\text{proj}_W(x)]_B = (A^* A)^{-1} A^* x, \quad A \hat{x} = \text{proj}_W(x).$$

- Invertibility of $A^* A \in \mathbb{C}^{p \times p}$ followed from

$$\text{rank}(A^* A) = \text{rank}(A) = p.$$ 

In general, we need not assume that $\text{rank}(A) = p$...
Claim. $A^*Ax = A^*b$ is consistent for every $b \in \mathbb{C}^n$.

To see this we first show $\text{col}(A^*A) = \text{col}(A^*)$.

- Indeed, if $y \in \text{col}(A^*A)$, then we may write $y = A^*[Ax]$, so that $y \in \text{col}(A^*)$.

- On the other hand, we have previously shown that $\text{rank}(A^*A) = \text{rank}(A^*)$. Thus $\text{col}(A^*A) = \text{col}(A^*)$.

Since $A^*b \in \text{col}(A^*)$, the claim follows.

If $\hat{x}$ is a solution to the normal system, then

$$A^*(b - A\hat{x}) = 0 \implies b - A\hat{x} \in [\text{col}(A)]^\perp.$$  

On the other hand, $A\hat{x} \in \text{col}(A)$, which shows

$$A\hat{x} = \text{proj}_W(b), \text{ where } W := \text{col}(A).$$

I.e. solutions to the normal system give combinations of the columns of $A$ equal to $\text{proj}_W(b)$.

Least squares solutions of $Ax = b$. We have just seen that $A^*Ax = A^*b$ is always consistent, even if $Ax = b$ is not!

We saw that the solution set of $A^*Ax = A^*b$ is equivalent to the solution set of

$$Ax = \text{proj}_W(b), \text{ where } W = \text{col}(A). \quad (\ast)$$

Indeed, we just saw that any solution to the normal system satisfies $(\ast)$, while applying $A^*$ to $(\ast)$ gives

$$A^*Ax = A^*\text{proj}_W(b) = A^*b \quad (\text{cf. } \text{col}(A)^\perp = \text{nul}A^*)$$

Note that $Ax = b$ is consistent precisely when $b \in \text{col}(A)$, i.e. when $b = \text{proj}_W(b)$.

Thus the normal system is equivalent to the original system precisely when the original system is consistent.
Least squares solutions of $Ax = b$ There is a clear geometric interpretation of the solution set to the normal system: let $\hat{x}$ be a solution to the normal system $A^*Ax = A^*b$. Then, with $W = \text{col}(A)$,

$$
\|b - A\hat{x}\| = \|b - \text{proj}_W(b)\| = \min_{w \in W} \|b - w\| = \min_{x \in \mathbb{C}^n} \|b - Ax\|.
$$

Thus $\hat{x}$ minimizes $\|b - Ax\|$ over all $x \in \mathbb{C}^n$.

The solution to the normal system for $Ax = b$ is called the least squares solution of $Ax = b$.

Example. Let

$$
A = [w_1 w_2 w_3] = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.
$$

The system $Ax = b$ is inconsistent, since

$$
[A|b] \sim \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
$$

The normal system $A^*Ax = A^*b$ is consistent, since

$$
[A^*A|A^*b] \sim \begin{bmatrix} 1 & 0 & 1 & 1/3 \\ 0 & 1 & 1 & -1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.
$$

The least squares solutions to $Ax = b$ are therefore

$$
\hat{x} = \begin{bmatrix} 1/3 \\ -1/3 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \quad z \in \mathbb{C}.
$$
Example. (cont.) We can also compute that
\[
\text{proj}_W b = A\hat{x} = \frac{1}{3}w_1 - \frac{1}{3}w_2 + 0 = \frac{1}{3} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix},
\]
where \( W = \text{col}(A) \).

The least squares error for \( Ax = b \) is defined by

\[
\min_{x \in \mathbb{C}^n} \| b - Ax \| = \| b - A\hat{x} \|.
\]

In this case, one can check that \( \| b - A\hat{x} \| = \frac{2}{3}\sqrt{3} \).

\[\blacktriangleright\] This is a measurement of the smallest error possible when approximating \( b \) by a vector in \( \text{col}(A) \).
**Linear models.** Suppose you have a collection of data from an experiment, given by

\[
\{(x_j, y_j) : j = 1, \ldots, n\}.
\]

You believe there is an underlying relationship describing this data of the form

\[
\beta_1 f(x) + \beta_2 g(x) + \beta_3 = h(y),
\]

where \(f, g, h\) are known but \(\beta = (\beta_1, \beta_2, \beta_3)^T\) is not.

Assuming a relation of this form and accounting for experimental error, we have

\[
\beta_1 f(x_j) + \beta_2 g(x_j) + \beta_3 = h(y_j) + \varepsilon_j
\]

for \(j = 1, \ldots, n\) and some small \(\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T\).

---

**Linear models.** (Cont.) In matrix form, we have

\[
X\beta - y = \varepsilon, \quad X = \begin{bmatrix} f(x_1) & g(x_1) & 1 \\ \vdots & \vdots & \vdots \\ f(x_n) & g(x_n) & 1 \end{bmatrix}, \quad y = \begin{bmatrix} h(y_1) \\ \vdots \\ h(y_n) \end{bmatrix}.
\]

Terminology:
- \(X\) is the design matrix,
- \(\beta\) is the parameter vector,
- \(y\) is the observation vector,
- \(\varepsilon\) is the residual vector.

The goal is to find \(\beta\) to minimize \(\|X\beta - y\|^2\).

To this end, we solve the normal system \(X^*X\beta = X^*y\). This solution gives the least squares best fit.
Example 1. (Fitting to a quadratic polynomial). Find a least squares best fit to the data

\((-1, 0),\) \((0, 1),\) \((1, 2),\) \((2, 4)\)

for the model given by \(y = \beta_1 x^2 + \beta_2 x + \beta_3.\) The associated linear model is

\[
X\beta = \begin{bmatrix}
1 & -1 & 1 \\
0 & 0 & 1 \\
1 & 1 & 1 \\
4 & 2 & 1
\end{bmatrix} \beta = \begin{bmatrix}
0 \\
1 \\
2 \\
4
\end{bmatrix} + \varepsilon = y + \varepsilon.
\]

Example 1. (cont.) The normal system \(X^*X\beta = X^*y\) has solution \(\hat{\beta} = [.25, 1.05, .85]^T,\) which implies the least squares best fit to the data is

\[y = .25x^2 + 1.05x + .85.\]

The least squares error is \(\|X\hat{\beta} - y\| = .0224.\)
Example 2. Kepler’s first law asserts that the orbit of a comet (parametrized by \((r, \theta)\)) is described by
\[ r = \beta + e(r \cos \theta), \]
where \(\beta, e\) are to be determined.

The orbit is elliptical when \(0 < e < 1\), parabolic when \(e = 1\), and hyperbolic when \(e > 1\).

Given observational data
\[(\theta, r) = \{(0.88, 3), (1.1, 2.3), (1.42, 1.65), (1.77, 1.25), (2.14, 1.01)\},\]
what is the nature of the orbit?
Example 2. (cont.) The associated linear model is

\[
\begin{bmatrix}
1 & r_1 \cos \theta_1 \\
\vdots & \vdots \\
1 & r_5 \cos \theta_5
\end{bmatrix}
\begin{bmatrix}
\beta \\
e
\end{bmatrix}
= 
\begin{bmatrix}
r_1 \\
\vdots \\
r_5
\end{bmatrix} + \varepsilon.
\]

We can rewrite this as \( X\beta = y + \varepsilon \). The solution to the normal system \( X^*X\beta = X^*y \) is given by \([\hat{\beta}, \hat{e}] = [1.45, .81]\).

We conclude that the orbit is most likely elliptical.