Last time: Orthogonality

1) Projections onto lines and planes in $\mathbb{R}^3$

Def: The projection of a vector $x$ onto $L = \text{span}\{u\}$ is the vector in $L$ that is closest to $x$.
We denote this vector by $P_u x$ or just $P x$.

Fact: The projection of $x$ onto $\text{span}\{u\}$ is given by $P_u x = \frac{\langle y, x \rangle}{\|y\|^2} y$.

Def: The matrix $P_u = \frac{uu^T}{\|u\|^2}$ is the orthogonal projector onto $\text{span}\{u\}$.

Facts:
1) $P_u^2 = P_u$
2) $P_u^T = P_u$ 

Consequences:
1) $\langle y, P_u z \rangle = \langle P_u y, z \rangle$ for all $y, z$
2) $\langle y, P_u z \rangle = \langle P_u y, z \rangle = \langle P_y, z \rangle$
3) $\text{R}(P_u) = \text{span}\{u\}$
4) $\text{N}(P_u) = [\text{R}(P_u)]^\perp$

Def: The projection onto $[\text{R}(P_u)]^\perp$ is given by $Q = I - P_u$ where $I$ is the identity matrix.

2) Orthogonal projection matrix

Def: An $n \times n$ matrix $P$ is an orthogonal projection matrix if and only if
1) $P^2 = P$
2) $P^T = P$

Facts:
1) If $P$ is an orthogonal projection matrix, then $Q = I - P$ is also an orthogonal projection matrix.
2) $P + Q = I$
3) $PQ = 0$
4) $P$ projects orthogonally onto $\text{R}(P)$
5) $Q$ projects orthogonally onto $\text{N}(P) = [\text{R}(P)]^\perp$
Proof of facts continued:

3. Want to find the closest vector in \( R(P) \) to \( x \).
   Thus we need to find \( y \) such that \( \|Py - x\|^2 \) is as small as possible.
   
   \[
   \|Py - x\|^2 = \|P(y - x) - Qx\|^2 = \langle P(y - x) - Qx, P(y - x) - Qx \rangle \\
   = \|P(y - x)\|^2 + \|Qx\|^2 - 2 \langle P(y - x), Qx \rangle \\
   = \|P(y - x)\|^2 + \|Qx\|^2
   \]

   \[\Rightarrow \text{This minimized if } y = x.\]
   Therefore \( Px \) is the closest vector in \( R(P) \) to \( x \).

4. We know from 1 & 3 that \( Q \) projects orthogonally on \( R(Q) \).
   Since \( R(P)^\perp = \text{N}(P) \) it remains to show that \( R(Q) = \text{N}(P) \).

   Claim: \( x \in R(Q) \iff Qx = x \)

   Proof: \( \Rightarrow \): Obvious
   \( \Leftarrow \): Let \( x \in R(Q) \), then for some \( y \) we have \( x = Qy \)
   Thus \( Qx = Q(Qy) = Q^2y = Qy = x \)

   We use the claim to prove 4:
   \( x \in R(Q) \iff Qx = x \iff (I - P)x = x \iff Px = 0 \iff x \in \text{N}(P) \)
We can also prove the Pythagorean theorem:

**Pythagorean theorem:** Let \( x \in \mathbb{R}^n \), \( P \) be an orthogonal projection matrix and \( Q = I - P \). Then

\[
\|x\|_2^2 = \|P x\|_2^2 + \|Q x\|_2^2
\]

**Proof:**

\[
\|x\|_2^2 = \langle x, x \rangle = \langle I x, I x \rangle = \langle (P+Q) x, (P+Q) x \rangle = \langle P x, P x \rangle + \langle P x, Q x \rangle + \langle Q x, P x \rangle + \langle Q x, Q x \rangle = \|P x\|_2^2 + \|Q x\|_2^2 = \|x\|_2^2
\]

We are done if \( \langle P x, Q x \rangle = 0 \)

\[
\langle P x, Q x \rangle = (P x)^T Q x = x^T P^T Q x
\]

\[
\langle P^T, P \rangle = x^T P Q x = 0
\]

\[
\Rightarrow \|x\|_2^2 = \|P x\|_2^2 + \|Q x\|_2^2
\]
Least squares and the projection onto $\text{R}(A)$

Recall $Ax = b$ has a solution if $b \in \text{R}(A)$

What if $b \notin \text{R}(A)$

We can find a vector $x$ such that $Ax \in \text{R}(A)$ and $Ax$ is the best match for $b$ in $\text{R}(A)$.

In other words, we want to find a vector $x_0$ such that $Ax_0 \in \text{R}(A)$ and $\|Ax_0 - b\|$ is as small as possible.

This will happen if $Ax_0$ is the projection of $b$ onto $\text{R}(A)$, this means

$$Ax_0 = \mathbf{P}_{\text{R}(A)} b$$

where $\mathbf{P}_{\text{R}(A)}$ is the projection matrix onto $\text{R}(A)$.

In the following I will write $P$ for $\mathbf{P}_{\text{R}(A)}$. 
From the theory we did before we know
\[ Qb = (I - P)b \perp R(A) \]
\[ = b - Ax_0 \]

\[ \iff Ax_0 - b \perp R(A) \]

However \( R(A) \perp = W(AT) \) \( \Rightarrow Ax_0 - b \in R(A) \perp = W(AT) \)

\[ \Rightarrow A^T(Ax_0 - b) = 0 \]

\[ \Rightarrow A^TAx_0 = A^Tb \] (Least square equation)

Every \( x_0 \) satisfying this equation is called a least square solution and \( Ax_0 \) is the projection of \( b \) onto \( R(A) \).

Properties of the least square equation
1. The least square equation always has a solution.
2. If \( A^TA \) is invertible, then the least square solution is unique.
"Proofs" for (1) & (2)

1. Homework. Hint: \( \mathbf{A} \mathbf{x} = \mathbf{A}^t \mathbf{b} \) has a solution iff \( \mathbf{A}^t \mathbf{b} \in \mathbf{R} (\mathbf{A}^t \mathbf{A}) \).
   Obviously \( \mathbf{A}^t \mathbf{b} \in \mathbf{R} (\mathbf{A} \mathbf{A}^t) \). So if \( \mathbf{R} (\mathbf{A}^t) \subseteq \mathbf{R} (\mathbf{A} \mathbf{A}^t) \) then we are done. It will turn out that
   \[ \mathbf{R} (\mathbf{A}^t) = \mathbf{R} (\mathbf{A} \mathbf{A}^t) \]
   which is equivalent to
   \[ \mathbf{N} (\mathbf{A}) = \mathbf{N} (\mathbf{A}^t \mathbf{A}) \]
   (Show this to prove 1)

2. If \( \mathbf{A}^t \mathbf{A} \) is invertible then
   \[ \mathbf{A}^t \mathbf{A} \mathbf{x}_0 = \mathbf{A}^t \mathbf{b} \iff \mathbf{x}_0 = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{b} \]
   So we have a unique solution.

Moreover, if \( \mathbf{x}_{LS} \) is the unique solution of the optimization problem:
Minimize \( \| \mathbf{A} \mathbf{x} - \mathbf{b} \|_2^2 \) for \( \mathbf{x} \in \mathbb{R}^n \)
and since \( \mathbf{A} \mathbf{x}_{LS} = \mathbf{P} \mathbf{b} \) where \( \mathbf{P} \) is the orthogonal projection onto \( \mathbf{R} (\mathbf{A}) \) we have
   \[ \mathbf{A} \mathbf{x}_{LS} = \mathbf{A} (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{b} = \mathbf{P} \mathbf{b} \iff \mathbf{P} = \mathbf{A} (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t. \]
i) $A^TA$ is invertible iff $\text{W}(A^TA) = \mathbb{R}^n$

This is equivalent to

$\text{W}(A) = \{0\} \iff$ the columns of $A$ are linearly independent

$\iff \text{rank}(A) = \# \text{ of columns of } A$.

ii) Let $A$ be an orthogonal projection matrix. ($A^2 = A$, $A^T = A$)

Consider the least square equation

$A^TAx = A^Tb \iff Ax = A^Tb$

$A^TA = A^T \Rightarrow A^T = A$

$A^2 = A \Rightarrow A^2x = Ab$

$A^2 = A \Rightarrow A^T = A$

$Ax = Ab$

$\Rightarrow x = Ab$ is solution of the least square equation!
Example:

\[
A = \begin{bmatrix}
1 & 1 \\
0 & -1 \\
0 & 0
\end{bmatrix}
\quad b = \begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix}
\]

Question: Is \( b \in \mathbb{R}(A) \)?

No there is no solution for \( Ax = b \).

So let do the next best thing:

Find a least square solution for \( Ax = b \).

That is a solution of \( A^TAx = A^Tb \)

\[
A^TA = \begin{bmatrix}
1 & 0 & 0 \\
1 & -1 & 0 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
1 & 1 \\
0 & -1 \\
0 & 0
\end{bmatrix} = \begin{bmatrix}
1 & 1 \\
1 & 2 \\
0 & 0
\end{bmatrix}
\]

\[
A^Tb = \begin{bmatrix}
1 & 0 & 0 \\
1 & -1 & 0 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} = \begin{bmatrix}
1 \\
-1 \\
3
\end{bmatrix}
\]

Thus

\[
A^TAx = A^Tb \iff \begin{bmatrix}
1 & 1 \\
1 & 2 \\
0 & 0
\end{bmatrix}x = \begin{bmatrix}
1 \\
-1
\end{bmatrix}
\]
Since \( \det \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = 2-1 \neq 0 \), \( X \) is invertible.

Hence

\[
X_{LS} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{2-1} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}
\]

is the unique solution for \( A^T A x = A^T b \).

Let's check the \( \ell_2 \)-distance between \( A x_{LS} \) and \( b \).

\[
A x_{LS} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}
\]

\[
\| \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \|_2 = \sqrt{9} = 3
\]

So the minimal \( \ell_2 \)-distance between \( b \) and \( R(A) \) is 3.
The orthogonal projection onto $\text{R}(A)$ is given by

$$P = A (A^T A)^{-1} A^T$$

$$= \begin{bmatrix} 1 & 1 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus the orthogonal projection of $b$ onto $\text{R}(A)$ is

$$P b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \ (= A x_{LS})$$