6.4 Best approximations; least squares

Exercise #2) b)

Solve the system \( \mathbf{A} \mathbf{x} = \mathbf{b} \) (1)

\[
\begin{bmatrix}
2 & -2 \\
1 & 1 \\
3 & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix}
= 
\begin{bmatrix}
2 \\
-1 \\
1
\end{bmatrix}
\]

\[
\begin{bmatrix}
2 & -2 & 2 \\
1 & 1 & -1 \\
3 & 1 & 1
\end{bmatrix}
\sim 
\begin{bmatrix}
1 & 1 & -1 \\
0 & -4 & 4 \\
0 & -2 & 4
\end{bmatrix}
\sim 
\begin{bmatrix}
1 & 1 & -1 \\
0 & 1 & 1 \\
0 & 0 & 0
\end{bmatrix}
\rightarrow \text{inconsistent!}
\]

This means that system (1) does not have a solution.

Then, the question or the vector \( \mathbf{b} \) is not in \( \text{Col}(\mathbf{A}) = \text{Span} \left\{ \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\} = W \)
Therefore, for any $\hat{x} \in \mathbb{R}^2$

$$\|A\hat{x} - \hat{b}\| > 0$$

The question is if there is $\hat{x}^* \in \mathbb{R}^2$ such that

$$\|A\hat{x}^* - \hat{b}\|$$

is minimum

or $0 < \|A\hat{x}^* - \hat{b}\| \leq \|A\hat{x} - \hat{b}\|$, for $\hat{x} \neq \hat{x}^*$ (2.1)

The answer is "yes" and the vector $\hat{x}^*$ happens to be a solution of

$$A\hat{x} = \text{proj}_W \hat{b}$$

where $W = \text{col}(A)$

This is the statement of next theorem.

**Thm 6.4.1** (Best approx. thm)

- $W \subset V$ finite-dimensional subspace of $\text{I.P.S. } V$
- $\hat{b} \in V$

Then, $\text{proj}_W \hat{b}$ is the best approximation to $\hat{b}$ from $W$

or $\|\hat{b} - \text{proj}_W \hat{b}\| < \|\hat{b} - \hat{w}\|$, for any $\hat{w} \in W$ such that $\hat{w} \neq \text{proj}_W \hat{b}$

(2.2)

Since $A\hat{x}^* = \text{proj}_W \hat{b}$ and for each $\hat{w} \in W$ there is at least one $\hat{x}_w$ such that $A\hat{x}_w = \hat{w}$, then (2.1) and (2.2) are equivalent.
Proof: the proof is based on the orthogonality of the vectors \( \vec{b} - \text{proj}_W \vec{b} \) with any vector \( \vec{w} \in W \). (thm 6.3.3)

In fact,
\[ \vec{b} - \vec{w} = (\vec{b} - \text{proj}_W \vec{b}) + (\text{proj}_W \vec{b} - \vec{w}) \]

\( \text{Now, } (\text{proj}_W \vec{b} - \vec{w}) \in W \). Therefore, \( (\vec{b} - \text{proj}_W \vec{b}) \perp (\text{proj}_W \vec{b} - \vec{w}) \)

Applying Pythagoras theorem
\[
\| \vec{b} - \vec{w} \|^2 = \| \vec{b} - \text{proj}_W \vec{b} \|^2 + \| \text{proj}_W \vec{b} - \vec{w} \|^2.
\]

\[ \Rightarrow \| \vec{b} - \text{proj}_W \vec{b} \|^2 < \| \vec{b} - \vec{w} \|^2 \quad \text{for any } \vec{w} = \text{proj}_W \vec{b} \]
Back to exercise # 2) b)

\[ \hat{b} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \quad \text{and} \quad W = \text{Col}(A) = \text{Span} \left\{ \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \right\} \]

It can be shown (verify it!) that

\[ \text{proj}_W \hat{b} = \begin{bmatrix} 10/21 \\ -5/21 \\ 13/21 \end{bmatrix} \quad \text{(It requires finding an orthogonal basis for Col}(A) \text{ using G-S process and then using the formula for proj}_W \hat{b} \text{ given by Thm 6.3.4)} \]

Therefore, a vector \( \hat{x}^* \in \mathbb{R}^2 \) that minimizes

\[ 0 < \| A \hat{x} - \hat{b} \| \]

Can be obtained by solving the system of equations

\[
\begin{bmatrix}
2 & -2 \\
1 & 1 \\
3 & 1
\end{bmatrix}
\begin{bmatrix}
x_1^* \\
x_2^*
\end{bmatrix} =
\begin{bmatrix}
10/21 \\
-5/21 \\
13/21
\end{bmatrix}
\]

\[ \Rightarrow \quad x_1^* = \frac{3}{7}, \quad x_2^* = -\frac{2}{3}, \quad \text{unique soln.} \]

Least squares problem:

Given a linear system \( A\hat{x} = \hat{b} \), a \( m \times n \) find \( \hat{x}^* \) if possible that minimizes \( \| A\hat{x} - \hat{b} \| \)

with respect to the Euclidean inner product.

Such \( \hat{x}^* \) is called a least squares solution of the system and \( \| b - A\hat{x}^* \| \) is called the least squares error.
In general, consider the system

\[ A\hat{x} = \hat{b} \]

If \[ W = \text{col}(A) \]

Then, from the best approx. thru the closest vector in \[ W \] to \[ \hat{b} \] is

\[ \text{proj}_W \hat{b} \]

Therefore, any least squares solution \( \hat{x}^* \) of \[ A\hat{x} = \hat{b} \] must satisfy

\[ A\hat{x}^* = \text{proj}_W \hat{b} \] \hspace{1cm} (5.0)

Notice that

\[ \| \hat{b} - \text{proj}_W \hat{b} \| < \| \hat{b} - \hat{w} \| \] \hspace{1cm} (5.1)

is equivalent to

\[ \| \hat{b} - A\hat{x}^* \| \leq \| \hat{b} - A\hat{x} \|, \quad \text{for all } \hat{x} \] \hspace{1cm} (5.2)

The equality sign in (5.2) happens when there is more than one soln. to (5.0).
Alternative to Solve Least Squares problems

Given \( Ax = \hat{b} \) consistent or not? \( (6.1) \)

its least squares soln. is obtained by solving
the least squares problem: \( A \hat{x} = \text{proj}_W \hat{b}, \quad W = \text{col}(A). \) \( (6.2) \)

We can avoid the comp. of \( \text{proj}_W \hat{b} \). In fact, from (6.2)

\[ \hat{b} - A \hat{x} = \hat{b} - \text{proj}_W \hat{b} \]

Applying \( A^T \) to both sides

\[ A^T (\hat{b} - A \hat{x}) = A^T (\hat{b} - \text{proj}_W \hat{b}) \]

Since \( (\hat{b} - \text{proj}_W \hat{b}) \in \text{Col}(A) = \text{Null}(A^T) \)

\[ A^T (\hat{b} - \text{proj}_W \hat{b}) = 0 \]

\[ A^T (\hat{b} - A \hat{x}) = 0 \iff A^T A \hat{x} = A^T \hat{b} \]

\( (6.3) \)

Normal e.qs. or
Assoc. normal syst. to
\( A \hat{x} = \hat{b}. \)
Therefore, any solution $\tilde{x}$ of

$$A\tilde{x} = \text{proj}_W \tilde{b} \tag{7.1}$$

is also a soln. of

$$A^T A \tilde{x} = A^T \tilde{b} \tag{7.2}$$

Conversely, if $\tilde{x}$ is a soln. of (7.2), then

$$A^T (\tilde{b} - A\tilde{x}) = 0$$

then, $(\tilde{b} - A\tilde{x}) \in \text{Null} (A^T) = [\text{Col} (A)]^\perp$

If we call $\tilde{w}_2 = \tilde{b} - A\tilde{x}$

then

$$\tilde{b} = \tilde{w}_2 + A\tilde{x}$$

i.e., $\tilde{w}_2 \in [\text{Col}(A)]^\perp$ and $A\tilde{x} \in \text{Col}(A) = W$

Now, from the "projection thm"

$$\tilde{b} = \tilde{w}_1 + \tilde{w}_2$$

where $\tilde{w}_1 = \text{proj}_W \tilde{b}$ and $\tilde{w}_2 = \tilde{b} - \text{proj}_W \tilde{b}$

and this decomp. is unique.

then, $A\tilde{x} = \tilde{w}_1 = \text{proj}_W \tilde{b}$ and $\tilde{w}_2 = \tilde{w}_2$

Therefore, $\tilde{x}$ is a soln. of (7.1)
The above results can be summarized in the following theorem.

Theorem 6.4.2 (diff than base)

Given the linear system $A\tilde{x} = \tilde{b}$ and column

$W = \text{col}(A)$.

The two linear systems

i) $A^TA\tilde{x} = A^T\tilde{b}$ (Associated linear system)

(ii) $A\tilde{x} = \text{proj}_W \tilde{b}$

are equivalent. They are also consistent and

all solutions are least squares soln. of $A\tilde{x} = \tilde{b}$. 

The systems are consistent, because $\text{proj}_W \tilde{b} \in \text{col}(A) = W$.

Also, from best approx. theorem, any soln. of (7.1) is

a least squares soln. of $A\tilde{x} = \tilde{b}$. 
Back to our original problem #2.(b)

Solve the system

\[
\begin{bmatrix}
2 & -2 \\
1 & 1 \\
3 & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix}
=
\begin{bmatrix}
2 \\
-1 \\
1
\end{bmatrix},
\quad W = \text{Col}(A)
\]

\( A \hat{x} = \hat{b} \)

We found this system inconsistent!

We also found that there is a unique least squares soln. \( \hat{x}^* \) satisfying the least squares problem

\[ A \hat{x}^* = \text{proj}_{W} \hat{b} \]

or

\[
\begin{bmatrix}
2 & -2 \\
1 & 1 \\
3 & 1
\end{bmatrix}
\begin{bmatrix}
x_1^* \\
x_2^*
\end{bmatrix}
=
\begin{bmatrix}
10/31 \\
-5/31 \\
13/31
\end{bmatrix}
\]

\( \Rightarrow \quad x_1^* = 3/7, \quad x_2^* = -2/3. \)

Using thm 6.4.2, the soln. \( \hat{x}^* \) can also be obtained by solving the system

\[ A^TA\hat{x}^* = A^T\hat{b} \]

or

\[
\begin{bmatrix}
2 & 1 & 3 \\
-2 & 1 & 1 \\
3 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
x_1^* \\
x_2^* \\
x_3^*
\end{bmatrix}
=
\begin{bmatrix}
2 & 1 & 3 \\
-2 & 1 & 1 \\
3 & 1 & 3
\end{bmatrix}
\]
or

\[
\begin{bmatrix}
14 & 0 \\
0 & 6 \\
\end{bmatrix}
\begin{bmatrix}
x^*_1 \\
x^*_2 \\
\end{bmatrix} =
\begin{bmatrix}
6 \\
-4 \\
\end{bmatrix} \Rightarrow x^*_1 = \frac{6}{14} = \frac{3}{7}
\text{ and }
\]

\[
x^*_2 = \frac{-4}{6} = -\frac{2}{3}
\]

This procedure is much simpler than the previous one of solving \( A\hat{x}^* = \text{proj}_W \hat{b} \).

To obtain \( \text{proj}_W \hat{b} \), we compute

\[
A\hat{x}^* = \begin{bmatrix}
2 & -2 \\
1 & 1 \\
3 & 1 \\
\end{bmatrix}
\begin{bmatrix}
\frac{3}{7} \\
-\frac{2}{3} \\
\end{bmatrix} =
\begin{bmatrix}
10/21 \\
-5/21 \\
13/21 \\
\end{bmatrix}
\]

and now, we can evaluate the least squares error.

\[
\|A\hat{x}^* - \hat{b}\| = \left\| \begin{bmatrix}
10/21 - 2 \\
-5/21 + 1 \\
13/21 - 1 \\
\end{bmatrix} \right\| = \left\| \begin{bmatrix}
-32/21 \\
16/21 \\
-8/21 \\
\end{bmatrix} \right\|
\]

\[
= \sqrt{\left(\frac{32}{21}\right)^2 + \left(\frac{16}{21}\right)^2 + \left(\frac{8}{21}\right)^2}
\]
Exercise #9 @

\[ \vec{u} = (6, 3, 7, 6), \quad \vec{v}_1 = (2, 1, 1, 1), \quad \vec{v}_2 = (1, 0, 1, 1) \]
\[ \vec{v}_3 = (-2, -1, 0, -1) \]

Find orthogonal projection of \( \vec{u} \) on \( W = \text{span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \).

---

If we define

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

then if \( \hat{x}^* \) is a least squares solution of

\[ A \hat{x} = \vec{u} \]

the projection can be computed as

\[ A \hat{x}^* = \text{proj}_W \vec{u} \]

In order to find \( \hat{x}^* \), we solve the normal system

\[ A^T A \hat{x} = A^T \vec{u} \]

Notice

\[ A^T A = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ -2 & -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 & -2 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 7 & 4 & -6 \\ 4 & 3 & -3 \\ -4 & 1 & 6 \end{bmatrix} \]

And

\[ A^T \vec{u} = \begin{bmatrix} 2 & 1 & 1 & 1 \\ -2 & -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \\ 9 \end{bmatrix} = \begin{bmatrix} 30 \\ 21 \end{bmatrix} \]
Then, we need to find $\mathbf{x}^*$ satisfying

$$
\begin{bmatrix}
7 & 4 & -6 \\
4 & 3 & -3 \\
-4 & -1 & 6
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
=
\begin{bmatrix}
30 \\
21 \\
21
\end{bmatrix}
$$

**Sln:** $x_1 = -6$, $x_2 = 9$, $x_3 = -6$

Therefore, $\text{proj}_{\mathbf{w}} \mathbf{u} =
\begin{bmatrix}
2 & 1 & -2 \\
1 & 0 & -1 \\
1 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
-6 \\
9 \\
6
\end{bmatrix}
=
\begin{bmatrix}
9 \\
0 \\
3
\end{bmatrix}$
**Thm 6.4.3** A mxn matrix. The following statements are equivalent.

a) A has linearly independent column vectors.

b) $A^TA$ is invertible

**Proof:**

a) $\Rightarrow$ b)

To show that $A^TA$ is invertible, we will prove the equivalent statement

$$A^TA \tilde{x} = \tilde{0}$$

only has the trivial solution.

In fact,

If $A^TA \tilde{x} = \tilde{0} \Rightarrow A\tilde{x} \in \text{Null}(A^T)$

but also $A\tilde{x} \in \text{Col}(A)$

from thm. 6.2.4 $\text{Null}(A^T) \cap \text{Col}(A) = \{\tilde{0}\}$

because they are orthogonal complements.

$\Rightarrow A\tilde{x} = \tilde{0}$, but A has n lin. indep. column vectors, then $\tilde{x} = \tilde{0}$ is the only possible solution of $A^TA \tilde{x} = \tilde{0}$.

b) $\Rightarrow$ a) If $\tilde{x}$ is a solution of $A\tilde{x} = \tilde{0}$ then $A^TA\tilde{x} = A^T(A\tilde{x}) = A\tilde{0} = \tilde{0}$.

Now, $A^TA$ is invertible, then only possible $\tilde{x}$ is $\tilde{x} = \tilde{0}$.

$\tilde{x} = \tilde{0} \Rightarrow \{c_1, c_2, ..., c_m\}$ is lin. indep.
Thm 6.4.4 - A m x n matrix
- A has n linearly indep columns. \( W = \text{col}(A) \)

then, the System \( A\hat{x} = \hat{b} \)
has a unique least squares soln.

\[ \hat{x} = (A^TA)^{-1} A^T \hat{b} \]

and \( \text{proj}_W \hat{b} = A\hat{x} = A (A^TA)^{-1} A^T \hat{b} \)

Proof: Follows from 6.4.3.
6.5 Least Squares Fitting Data (Lab)

Consider a set of points:

\[(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^2\]

Suppose we want to fit a line to these points.

**Equation of the Line:**

\[y = a + bx\]

\[
\begin{align*}
  y_1 &= a + bx_1 \\
  y_2 &= a + bx_2 \\
  \vdots \\
  y_n &= a + bx_n
\end{align*}
\]

\[
\begin{bmatrix}
  1 & x_1 \\
  1 & x_2 \\
  \vdots & \vdots \\
  1 & x_n
\end{bmatrix}
\begin{bmatrix}
  a \\
  b
\end{bmatrix}
=
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n
\end{bmatrix}
\]

\[M \hat{\mathbf{v}} = \hat{\mathbf{y}}\]

For noncollinear points, this system is inconsistent! It's possible to show that the two columns of \(M\) are linearly independent, except when \((x_1, y_1), \ldots, (x_n, y_n)\) lie on a vertical line. Therefore, the least squares solution \(\hat{\mathbf{v}} = (a^*, b^*)\) is unique and satisfies

\[M^TM \hat{\mathbf{v}}^* = M^T \hat{\mathbf{y}}\]

The sought line is given by

\[\hat{y} = a^* + b^*x\]