Sec 6.5 Least-Squares problems

Motivation:

- For applications, we often work with systems

$$A\vec{x} = \vec{b}$$

  that have no solutions (i.e. are inconsistent)

- We can approximate a solution by finding an $\vec{x}$ that makes $A\vec{x}$ as close as possible to $\vec{b}$,

  i.e. by finding an $\vec{x}$ that makes

$$\|\vec{b} - A\vec{x}\| \text{ as small as possible}$$

  This process is called the <u>general least-squares problem</u>

- "least-squares" comes from the fact that

$$\|\vec{b} - A\vec{x}\| = \sqrt{\text{sum of squares}}$$

Recall (Sec 6.3 Thm 9) "The best approximation theorem":

> Let W be a subspace of $\mathbb{R}^n$, let $\vec{b}$ be a vector in $\mathbb{R}^n$.
>
> Then the closest point in W to $\vec{b}$ is
>
> $$\hat{b} = \mathrm{proj}_W \vec{b}, \text{ the orthogonal projection of } \vec{b} \text{ onto } W,$$
>
> i.e.
>
> $$\|\vec{b} - \hat{b}\| < \|\vec{b} - \vec{v}\| \text{ for any } \vec{v} \text{ in } W \text{ that's not equal to } \hat{b}.$$

How to solve the general least-squares problem?

Let A be an $m \times n$ matrix and $\vec{b}$ in $\mathbb{R}^m$

Consider the system $A\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$

Let W denote $\mathrm{Col}\, A = \mathrm{Span}\left\{ \text{columns of } A \right\}$

(see Sec 4.2)

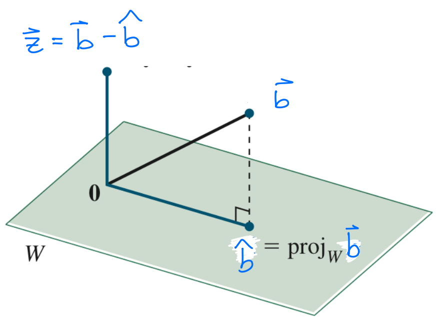
$= \left\{ \vec{v} : A\vec{x} = \vec{v} \text{ for some } \vec{x} \text{ in } \mathbb{R}^n \right\}$

Ex: (Example 1) $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ $\vec{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$

Note: The columns of A are linearly independent

Note $\left[\begin{array}{cc|c} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 1 & 11 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 4 & 0 & 2 \\ 0 & 2 & 0 \\ -4 & -4 & -44 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 4 & 0 & 2 \\ 0 & 4 & 0 \\ 0 & -4 & -42 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 4 & 0 & 2 \\ 0 & 4 & 0 \\ 0 & 0 & -42 \end{array}\right]$ so $A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ is inconsistent

Let $\hat{b}$ denote $\mathrm{proj}_W \vec{b}$, the orthogonal projection of $\vec{b}$ onto W.



Q: Is $\hat{b}$ in W?

Ans: Yes, by def

So $\hat{b}$ is in $\mathrm{Col}\, A$, so there is $\hat{x}$ in $\mathbb{R}^n$ such that $A\hat{x} = \hat{b}$.

Since $\hat{b} = \mathrm{proj}_W \vec{b}$ is the closest point in $W = \mathrm{Col}\, A$ to $\vec{b}$,

$A\hat{x} = \hat{b}$ is the closest point in $W = \mathrm{Col}\, A$ to $\vec{b}$.

That is, $\|\vec{b} - \hat{b}\| \leq \|\vec{b} - \vec{v}\|$ for all $\vec{v}$ in $W = \mathrm{Col}\, A$

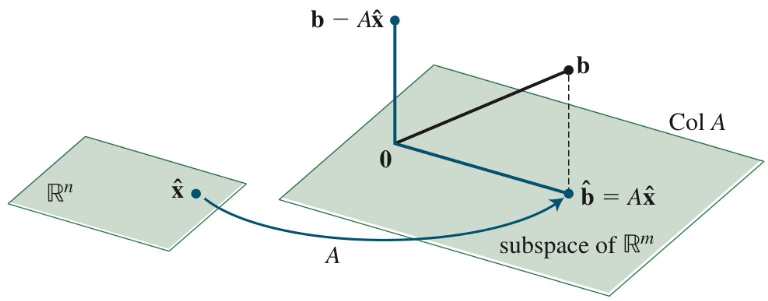
i.e. $\|\vec{b} - A\hat{x}\| \leq \|\vec{b} - A\vec{x}\|$ for all $\vec{x}$ in $\mathbb{R}^n$

Def If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

Note: $\hat{x}$ is a least-squares solution of $A\vec{x} = \vec{b}$ iff $A\hat{x} = \hat{b}$



$$\boxed{\text{How do we solve } A\hat{x} = \hat{b}\ ?}$$

Suppose $\hat{x}$ is a vector in $\mathbb{R}^n$ satisfying $A\hat{x} = \hat{b}$.

By the "Orthogonal Decomposition Theorem" (Sec 6.3),

$\vec{z} = \vec{b} - \hat{b}$ is orthogonal to $W = \text{Col } A = \text{Span}\{\text{cols of } A\}$
$= \vec{b} - A\hat{x}$,

so $\vec{b} - A\hat{x}$ is orthogonal to each column of A.

So... If $\vec{a}$ in $\mathbb{R}^m$ is any column of A,

then $0 = \vec{a} \cdot (\vec{b} - A\hat{x}) = \vec{a}^T (\vec{b} - A\hat{x})$

def of inner product

Ex: $\overbrace{[4\ \ 0\ \ 1]}^{\text{a column of } A} (\vec{b} - A\hat{x}) = 0$

Since $\vec{a}^T$ is a row of A^T, we have ..

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = A^T(\vec{b} - A\hat{x}) \qquad \text{Ex: } \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} (\vec{b} - A\hat{x})$$

$$\Downarrow$$

Thus $\vec{0} = A^T\vec{b} - A^TA\hat{x}$, i.e. $A^TA\hat{x} = A^T\vec{b}$.

Def The matrix equation $\boxed{A^T A \hat{x} \overset{(*)}{=} A^T \vec{b}}$ represents a system of equations called the <u>normal equations for $A\vec{x} = \vec{b}$</u>

Ex: The normal equations for our $A\vec{x} = \vec{b}$ are

$$A^T A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^T \vec{b}$$

$$\begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Thm The set of least-squares solutions of $A\vec{x} = \vec{b}$ (Thm 13) is equal to the (nonempty) set $\{$solutions of the normal equations $A^T A \vec{x} = A^T \vec{b}\}$

Ex $\left[\begin{array}{cc|c} 17 & 1 & 19 \\ 1 & 5 & 11 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right]$

For this example of A and $\vec{b}$, the equation $A\vec{x} = \vec{b}$ has a unique least-squares solution, $\hat{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

Note:

1st col of A is a linear Combination of cols 2,3,4

EXAMPLE 2 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$$

SOLUTION Compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

The augmented matrix for $A^TA\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The general solution is $x_1 = 3 - x_4$, $x_2 = -5 + x_4$, $x_3 = -2 + x_4$, and x_4 is free. So the general least-squares solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$ ■

The next theorem gives useful criteria for determining when there is only one least-squares solution of $A\mathbf{x} = \mathbf{b}$. (Of course, the orthogonal projection $\hat{\mathbf{b}}$ is always unique.)

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

Ex 1 has a unique least-squares solution since A^TA is invertible
(also the columns of A are linearly independent)

Ex 2 does not since A^TA is not invertible.
(also the columns of A are linearly dependent)

Def

When a least-squares solution $\hat{\mathbf{x}}$ is used to produce $A\hat{\mathbf{x}}$ as an approximation to $\mathbf{b}$, the distance from $\mathbf{b}$ to $A\hat{\mathbf{x}}$ is called the **least-squares error** of this approximation.

EXAMPLE 3 Given A and $\mathbf{b}$ as in Example 1, determine the least-squares error in the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

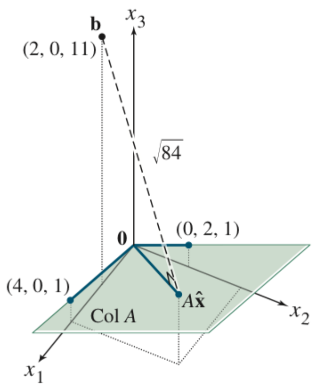


FIGURE 3

SOLUTION From Example 1,

$$\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \quad \text{and} \quad A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

Hence

$$\mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$$

and

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84}$$

The least-squares error is $\sqrt{84}$. For any $\mathbf{x}$ in $\mathbb{R}^2$, the distance between $\mathbf{b}$ and the vector $A\mathbf{x}$ is at least $\sqrt{84}$. See Figure 3. Note that the least-squares solution $\hat{\mathbf{x}}$ itself does not appear in the figure. ■

— end of PDF —