Moore-Penrose Pseudoinverse

Classical least squares problem

J. Antonio Ortega

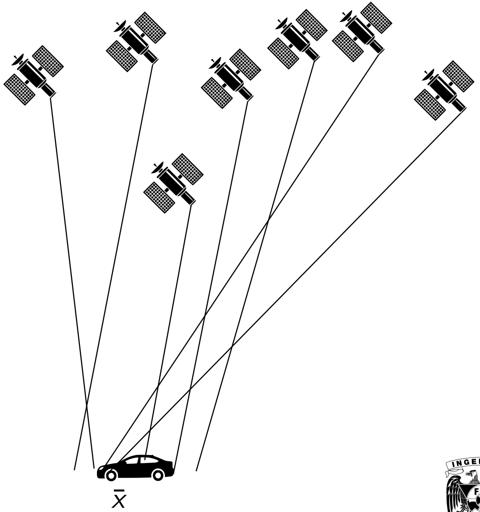
UNAM

November 15, 2021

Motivational Example

Consider the following equation:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix}}_{A} \begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 8 \end{bmatrix}$$



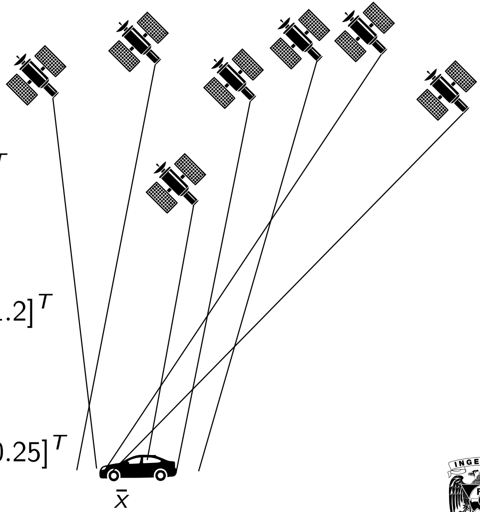
Motivational Example

It can be seen as tree different subsystems:

$$\Sigma_1 := \left\{ \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right. \implies \quad [i,\ j]^T = [1,\ 1]^T$$

$$\Sigma_2 := \left\{ \begin{bmatrix} 1 & 1 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \end{bmatrix} \right. \implies \quad [i,\ j]^T = [0.8,\ 1.2]^T$$

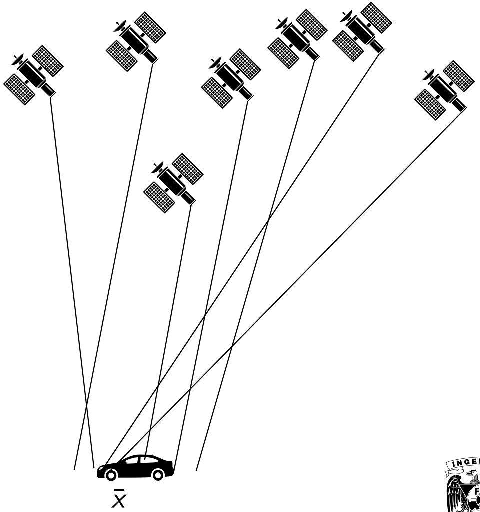
$$\Sigma_3 := \left\{ \begin{bmatrix} 1 & 2 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \end{bmatrix} \right. \implies \quad [i,\ j]^T = [0.5,\ 0.25]^T$$



Motivational Example

Consider the following equation:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix}}_{A} \begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

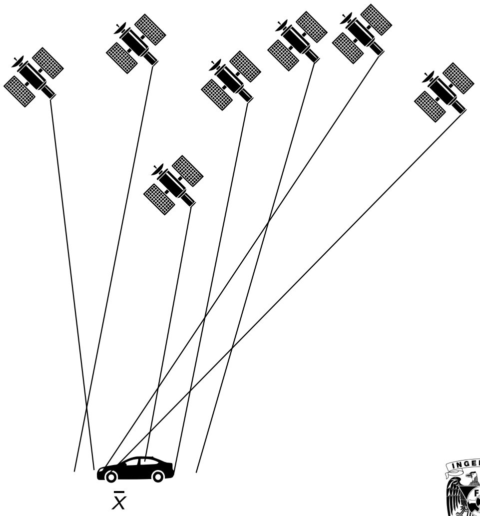


Motivational Example

Consider the following equation:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix}}_{A} \begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Is there solution for this system?



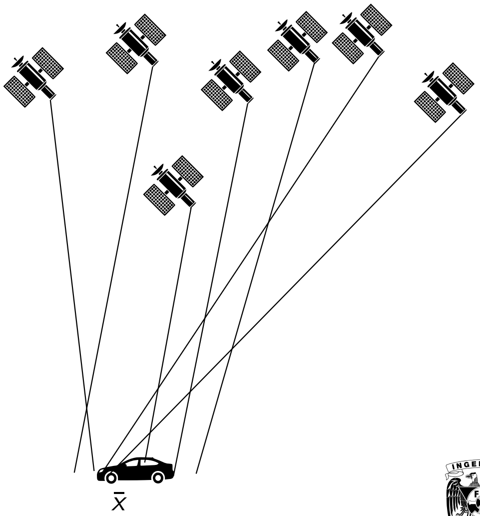
Motivational Example

Consider the following equation:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix}}_{A} \begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Is there solution for this system?

There will be solution iff
$b = [b_1\ b_2\ b_3]^T \in \mathcal{C}(A)$ *column space* of A
i.e., b is a linear combination of the columns of A



How to solve it in the *best possible way*?

- $Ax = b$ may have no solution

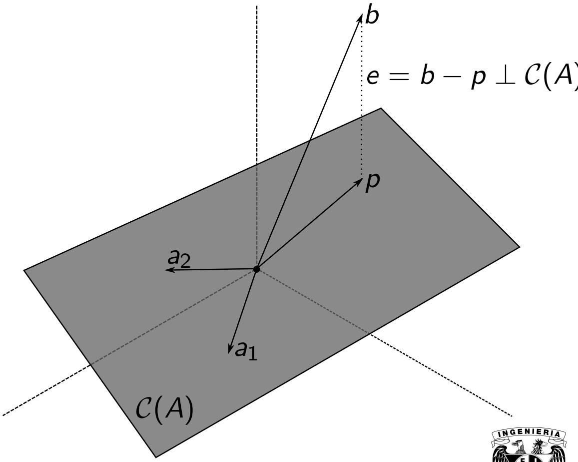
Denote a_1 and a_2 as the columns of the matrix A, thus,

$$\mathcal{C}(A) := \{b \in \mathbb{R}^3 | Ax = b,\ x \in \mathbb{R}^2\}$$

that is, all the linear combinations of the columns of A.
There exists a unique vector on $\mathcal{C}(A)$ nearest to b

$$p = a_1\hat{x}_1 + a_2\hat{x}_2 = A\hat{x}$$

where p is the projection of b onto the column space!

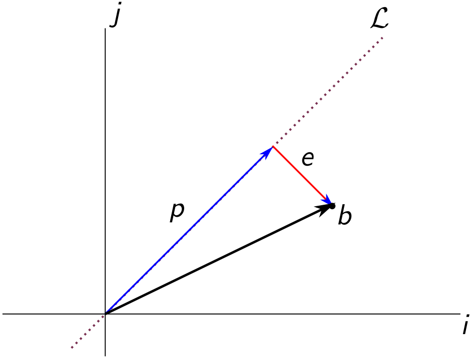


- A possible alternative $A\hat{x} = p$

Lemma: Let x be a vector and $\mathcal{L}$ is a linear manifold in $\mathbb{R}^n$ (*i.e.*, if p, $q \in \mathcal{L}$, then $\alpha i + \beta j \in \mathcal{L}$ for any scalars α, β). Then if

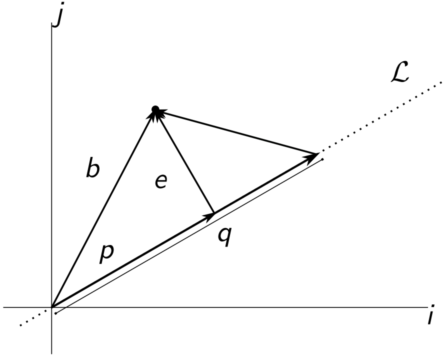
$$b = p + e$$

where $p \in \mathcal{L}$ and $e \perp \mathcal{L}$, then p is *"nearest"* to b, or, in other words, it is the projection of b to the manifold $\mathcal{L}$.



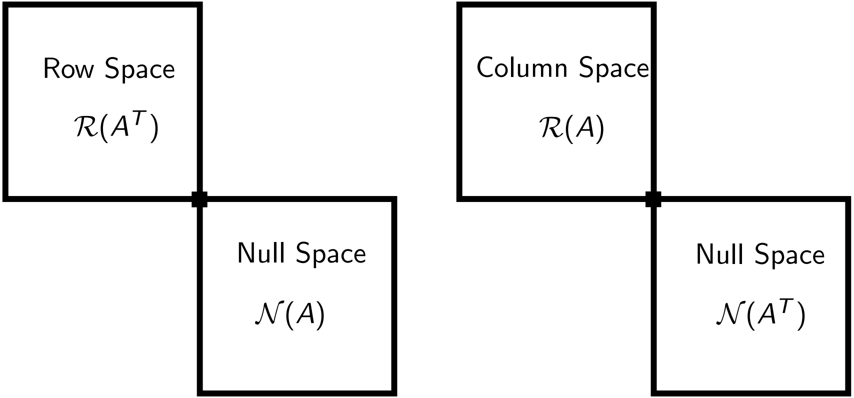
Proof.: For any $q \in \mathcal{L}$ we have

$$\begin{aligned}
\|b - q\|^2 &= \|p + e - q\|^2 \\
&= \|(p - q) + e\|^2 \\
&= \|(p - q)\|^2 + 2(p - q)^T e + \|e\|^2 \\
&= \|(p - q)\|^2 + \|e\|^2 \\
&\geq \|e\|^2
\end{aligned}$$



A little recall

The four subspaces of linear algebra. Lecture of Dr. Strang [Link]



Theorem: Let b be an $n-$dimensional real vector and $A \in \mathbb{R}^{n\times m}$.

- 1. There is always a vector, in fact a unique vector $\hat{x}$ of minimal (Euclidian norm), which minimizes

$$\hat{x} = argmin\left(\|b - Ax\|^2\right).$$

- 2. The vector $\hat{x}$ is the unique vector in the range

$$\mathcal{R}(A^T) := \{x : x = A^T b, b \in \mathbb{R}^n\}$$

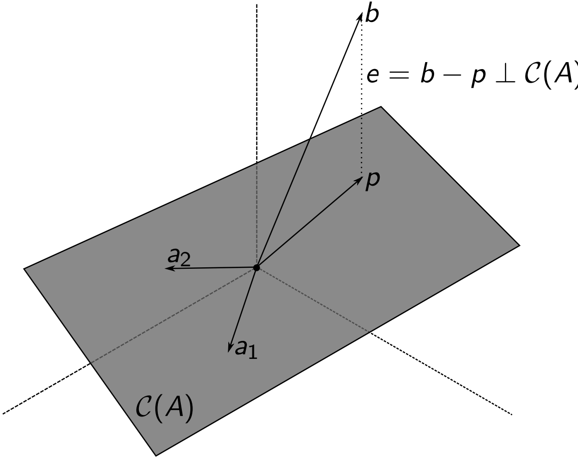
which satisfies the equation

$$A\hat{x} = p$$

where p is the projection of b on $\mathcal{R}(A) = \mathcal{C}(A)$.

1. There is always a vector, in fact a unique vector $\hat{x}$ of minimal (Euclidian norm), which minimizes

$$\|b - Ax\|^2.$$



$$e \perp \mathcal{C}(\mathcal{A})$$

$$b - p \perp \mathcal{C}(\mathcal{A})$$

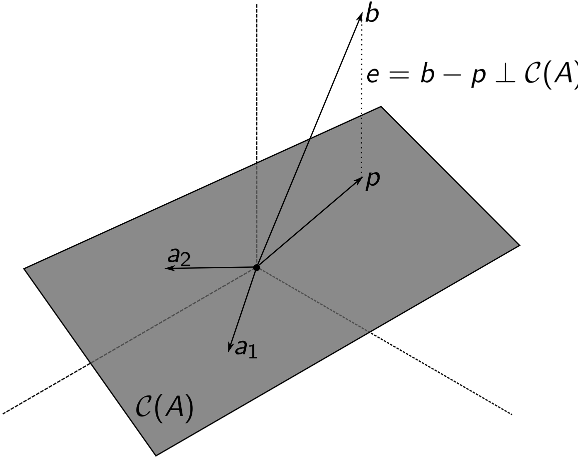
$$b - A\hat{x} \perp \mathcal{C}(\mathcal{A})$$

$$\begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix} (b - A\hat{x}) = 0$$

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

$$A^T e = 0 \implies e \in \mathcal{N}(A^T)$$

$$A^T A\hat{x} = A^T b$$



Proof.: We can write

$$b = \hat{b} + \tilde{b}$$

where $\hat{b}$ is the projection of b on the kernel (*null space*)

$$\mathcal{N}(A^T) := \{b \in \mathbb{R}^n | A^T b = 0\}.$$

Since $Ax \in \mathcal{R}(A)$ for any $x \in \mathbb{R}^m$, it follows that

$$\hat{b} - Ax \in \mathcal{R}(A)$$

and, since $\tilde{b} \in \mathcal{R}^{\perp}(A)$,

$$\tilde{b} \perp \hat{b} - Ax, \therefore$$

$$\begin{aligned}\|b - Ax\|^2 &= \|(\hat{b} - Ax) + \tilde{b}\|^2 \\ &= \|\hat{b} - Ax\|^2 + \|\tilde{b}\|^2 \geq \|\tilde{b}\|^2 = \|b - \hat{b}\|^2\end{aligned}$$

$\|b - Ax\|^2 \geq \|b - \hat{b}\|^2$ This low bound is attainable since $\hat{b}$, being the range of A, is the afterimage of some x^*, that is, $\hat{b} = Ax^*$.

1. Let us show that x^* has a minimal norm.

$$x^* = \hat{x}^* + \tilde{x}^*$$

where $\hat{x}^* \in \mathcal{R}(A^{\perp})$ and $\tilde{x}^* \in \mathcal{N}(A)$.

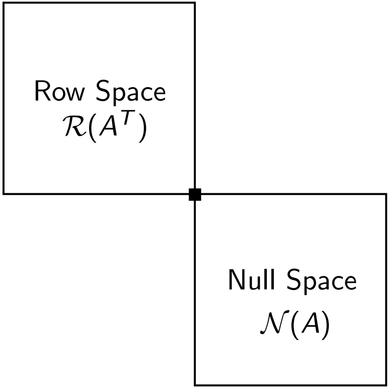
Thus, $Ax^* = A\hat{x}$ we have $\|b - Ax^*\|^2 = \|b - A\hat{x}\|^2$ and $\|x^*\|^2 \geq \|\hat{x}^*\|^2$.

So, x^* may be selected equal to $\hat{x}^*$.

2. Show that $x^* = \hat{x}^*$ is unique. Suppose that $Ax^* = Ax^{**} = p$. Then

$$(x^* - x^{**}) \in \mathcal{R}(A^T)$$

But, $A(x^* - x^{**}) = 0$, therefore $(x^* - x^{**}) \in \mathcal{N}(A) = \mathcal{C}^{\perp}(A^{\perp})$
Thus, $(x^* - x^{**})$ is orthogonal to itself! i.e., $\|x^* - x^{**}\|^2 = 0 \implies x^* = x^{**}$.



Theorem: Among those vectors x, which minimize $\|b - Ax\|^2$, $\hat{x}$, the one having minimal norm, is the unique vector of the form:

$$\hat{x} = A^T y,$$

satisfying

$$A^T A\hat{x} = A^T b.$$

Therefore,

$$\hat{x} = (A^T A)^{-1} A^T b.$$

The Moore-Penrose Pseudoinverse is defined by

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

Ps. Try to proof that if A is *full-column rank* the square matrix is invertible.

$A^T A$ is invertible if it's full-column rank

Consider that

$$A^T Ax = 0.$$

Thus, the column space $\mathcal{C}(A)$ is in the null space of A^T. But we know that $\mathcal{C}(A) \perp \mathcal{N}(A^T)$, which implies that $Ax = 0$.
If A is full-column rank, the columns are linearly independient, which implies that if $Ax = 0 \implies x = 0$.
Therefore $\mathcal{N}(A^T A) = \{0\}$ which implies that $A^T A$ is invertible.

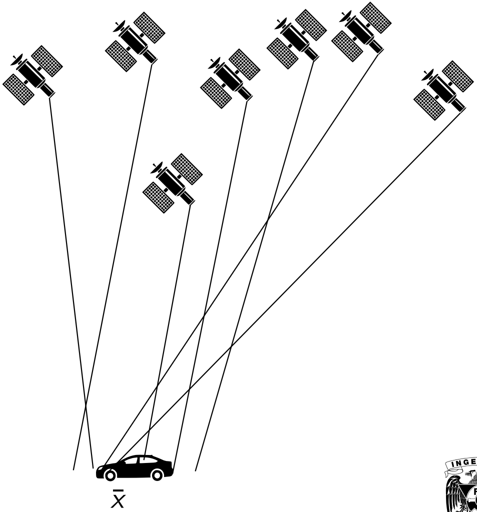
Motivational Example

Consider the following equation:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix}}_{A} \begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 8 \end{bmatrix}$$

We need to minimize $\|Ax - b\|^2$

$$\|Ax - b\|^2 = (i + j - 2)^2 + (i + 2j - 3)^2 + (i + 6j - 8)^2$$



Motivational Example

$$\|Ax - b\|^2 = (i + j - 2)^2 + (i + 2j - 3)^2 + (i + 6j - 8)^2$$

$$\frac{\partial}{\partial i}(\|Ax - b\|^2) = 0 \implies 3i + 9j = 13$$
$$\frac{\partial}{\partial j}(\|Ax - b\|^2) = 0 \implies 9i + 41j = 56$$

$$\Sigma_m := \begin{cases} 3i + 9j = 13 \\ 9i + 41j = 56 \end{cases}$$

$$[i, j]_{\Sigma m} = [0.6905, 1.2143]$$

Do you remember the expression $A^T A\hat{x} = A^T b$

$$A^T A = \begin{bmatrix} 3 & 9 \\ 9 & 41 \end{bmatrix} \text{ and } A^T b = \begin{bmatrix} 13 & 56 \end{bmatrix}$$

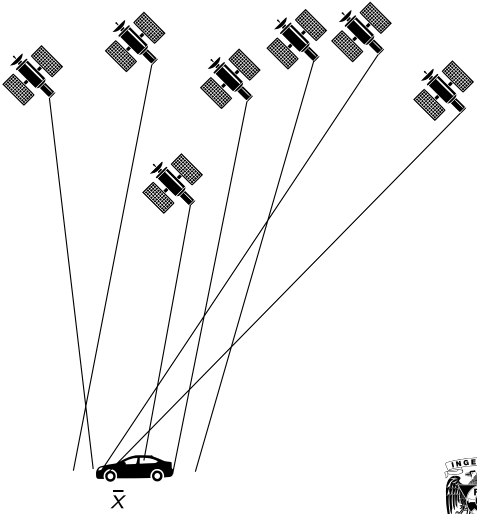
Motivational Example

Consider the following equation:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix}}_{A} \begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 8 \end{bmatrix}$$

So, the best possible solution of this system is:

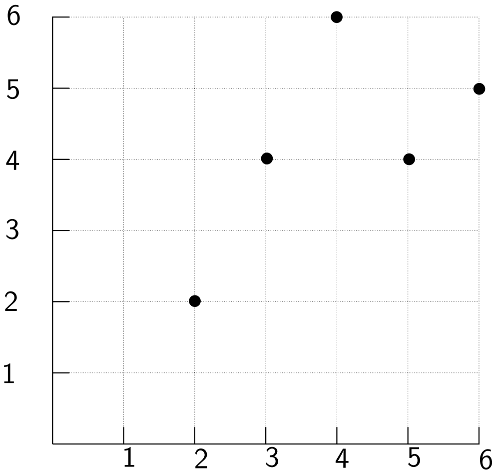
$$\begin{aligned} Ax &= b \\ A^{+}Ax &= A^{+}b \\ x &= A^{+}b \\ x &= [0.6905 \quad 1.2143]^{T} \end{aligned}$$



Example

Consider a classical least square problem, find the line that mimize the distance of all the points to the line.
We consider a line $y = \alpha x + \beta$, thus we can write the following system

$$
\begin{aligned}
\beta + 2\alpha &= 2 \\
\beta + 3\alpha &= 4 \\
\beta + 4\alpha &= 6 \\
\beta + 6\alpha &= 5
\end{aligned}
$$



Useful Resources

- **Advanced Mathematical Tools for Automatic Control Engineers**. *Alexander S. Poznyak* [Link]
- **Series of talks about Linear Algebra by 3blue1brown** [Link]
- **Video Lectures about Linear Algebra by Dr.Strang** [Link]