2. Pseudosolutions

Intuition behind pseudosolutions

Let's recall that the system of linear equations (SLE) can be represented using the following notation

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

where

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in M_{m\times n}(\mathbb{C}), \quad \vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_{,} \end{bmatrix} \in \mathbb{C}^m, \quad \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{C}^n.$$

Let's recall possible solution sets of SLE

Theorem: Solution sets of SLE

linear system of equations may behave in any one of three possible ways

1. The system is definite - has a single unique solution.
2. The system is indefinite - has infinitely many solutions.
3. The system is inconsistent - has no solution.

Example 1: Definite system of the square system.

If $A \in M_{n\times n}(\mathbb{C})$, $\operatorname{rank} A = n$, then we can easily obtain unique solution $\vec{x}$ by inverting the matrix of initial coefficients

$$\vec{x} = A^{-1}\vec{b}.$$

We want to generalize this result for any systems (definite, indefinite and inconsistent of any size). Let's outline intuition of this generalization in the following examples.

Definite system

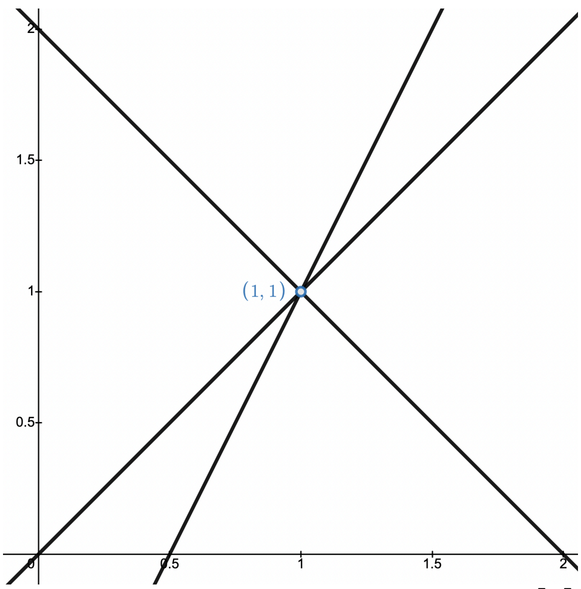
Example 2: If $A \in M_{m\times n}(\mathbb{C})$, $\operatorname{rank} A = n$, then there exist unique solution $\vec{x}$. For example, consider a system

$$\begin{cases} x - y = 0 \\ x + y = 2 \\ 2x - y = 1 \end{cases}.$$

The system has only one solution in the point $\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. However, we would like to generalize the method of obtaining a solution in such a way that it looks similar to the first example (definite system of square system)

$$\vec{x} = ? \cdot \vec{b}.$$

And looking ahead we can obtain such a factor to express solution that way. But now let's get a broader generalization.



Example of definite system.

Indefinite system

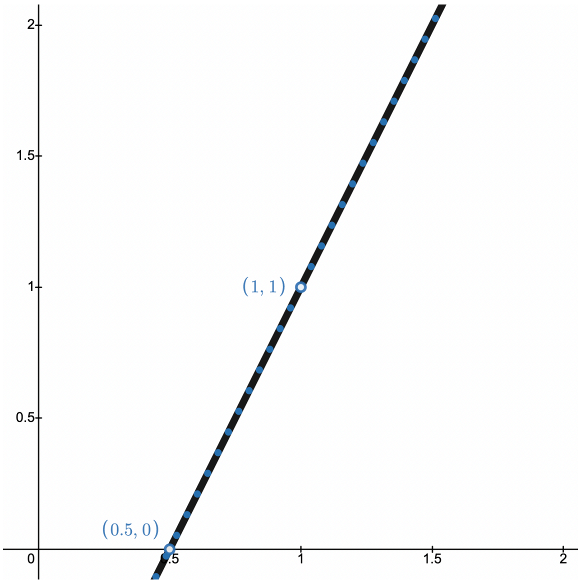
Example 3: Consider a system

$$\begin{cases} 2x - y = 1 \\ 4x - 2y = 2 \end{cases}.$$

It is not so obvious to choose a specific solution here because a whole family of solutions of the following form $\vec{x} = \begin{bmatrix} x \\ 1 - 2x \end{bmatrix}$ is suitable for us. At the same time, we would like express solution of the system in the following form

$$\vec{x} = ? \cdot \vec{b},$$

which assumes one solution, not infinite. For this reason, we need to choose the best solution out of infinite. In other words, we need to define "optimal"solution. Our intuition hints us to a solution point which is the closest to the origin.



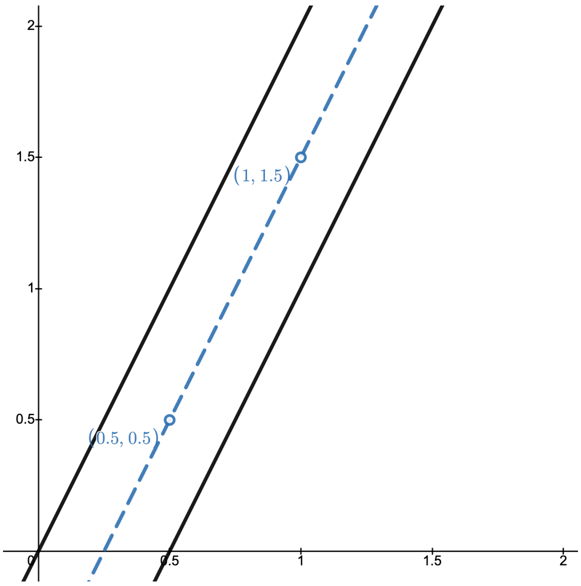
Example of indefinite system.

Inconsistent system

Example 4: Consider a system

$$\begin{cases} 2x - y = 0, \\ 2x - y = 1 \end{cases}.$$

The system does not have solution. Geometrically the system describes two parallel lines. Our intuition hints us to choose a point on a line right in middle of the parallel lines. However, it is still a whole family of solutions that can be the answer to the request of the product or business problem.



Example of inconsistent system.

Pseudosolution

Definition: Pseudosolution

Consider a system of a linear equations $A\vec{x} = \vec{b}$. A vector $\vec{u} \in \mathbb{C}^n$ is called a pseudosolution (or least square solution), if

$$\left|A\vec{x} - \vec{b}\right| \geqslant \left|A\vec{u} - \vec{b}\right|, \quad \forall \vec{x} \in \mathbb{C}^n.$$

That is, vector $f(\vec{x}) = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} = A\vec{x} - \vec{b}$ reaches it's minimum length at $\vec{x} = \vec{u}$. That is, $\vec{x} = \vec{u}$ is the solution of the following problem

$$|f(x)|^2 = |f_1|^2 + \ldots + |f_n|^2 \to \min_{\vec{x}}.$$

Theorem: $\vec{u} := A^{+}\vec{b}$

The vector $\vec{u} = A^{+}\vec{b}$ is a pseudosolution of the system of linear equations $A\vec{x} = \vec{b}$. Moreover, among all pseudosolutions, the vector $\vec{u}$ is the unique pseudosolution with minimal length.

Prop

If $\hat{x}$ is a solution, then it is a pseudosolution.

<u>Proof:</u> $A\hat{x} - \vec{b} = 0 \Longrightarrow \left|A\hat{x} - \vec{b}\right| = 0 = \min |f_x|^2.$ □

Prop: 1

Type of a system	Solution
definite	$\hat{x} = A^{+}\vec{b}$ is the solution
indefinite	$\hat{x} = A^{+}\vec{b}$ is the solution of minimal length
inconsistent	$\hat{x} = A^{+}\vec{b}$ is the pseudosolution of minimal length

Lemma: 1

$\mathrm{Im}(AA^{+} - I) \perp \mathrm{Im}(A)$.

<u>Proof:</u> Let us denote $AA^{+} - I$ by M. Let us denote vector columns that generate $\mathrm{Im}(M)$ and $\mathrm{Im}(A)$ by M^i and A^j respectively. Then the following holds

$$\begin{aligned} & \mathrm{Im}(M) \perp \mathrm{Im}(A) && (1) \\ \Leftrightarrow\ & M^i \perp A^j, \quad \forall i, j \\ \Leftrightarrow\ & (M^i, A^j) = 0, \quad \forall i, j \\ \Leftrightarrow\ & (M^i)^* \cdot A^j = 0, \quad \forall i, j \\ \Leftrightarrow\ & M^* A = 0, && (2) \end{aligned}$$

which means that by proving (2), we get (1). Let us prove (2).

$$M^*A = (AA^{+} - I)^*A = ((AA^{+})^* - I^*)A = (AA^{+})^*A - I^*A \overset{\text{axiom III}}{=} AA^{+}A - I \overset{\text{axiom I}}{=} A - A = 0.$$

□

Theorem: Pythagoras

Suppose $\vec{a} \perp \vec{b}$, then for $\vec{c} = \vec{a} + \vec{b}$, we have $|\vec{c}|^2 = |\vec{a}|^2 + |\vec{b}|^2$. In particular $|\vec{c}| \geq |\vec{b}|$. The equality holds only for $\vec{a} = \vec{0}$.

Proof: We leave as an exercise for the reader. □

Proof: (of Theorem $\vec{u} = A^+\vec{b}$) First, let us prove that $\vec{u} = A^+\vec{b}$ is a pseudosolution. Let us denote

$$\vec{A}_x := A(\vec{x} - \vec{u}) = A\vec{x} - A\vec{u} \in \mathbb{C}^n, \quad \vec{B} := (AA^+ - I)\vec{b} = A\vec{u} - \vec{b} \in \mathbb{C}^n, \quad \vec{C} := \vec{A}_x + \vec{B} = A\vec{x} - \vec{b}.$$

Using Lemma, we get $\vec{A}_x \perp \vec{B}$. Using Pythagoras theorem, we get

$$\begin{gathered} |C| \geqslant |B| \\ \Leftrightarrow \\ |A\vec{x} - \vec{b}| \geqslant |A\vec{u} - \vec{b}|, \quad \forall \vec{x} \in \mathbb{C}^n, \end{gathered}$$

which means that $\vec{u}$ is a pseudosolution.

Secondly, let us prove that $\vec{u} = A^+\vec{b}$ is the unique pseudosolution with minimal length. Let us denote $w := \vec{x} - \vec{u}$. Suppose $\vec{x}$ is another pseudosolution, which means that

$$A\vec{x} = A\vec{u} \Leftrightarrow A(\vec{x} - \vec{u}) = 0 \Leftrightarrow Aw = 0. \tag{3}$$

Let us consider the following

$$\begin{aligned} (\vec{u}, \vec{w}) &= \vec{u}^*\vec{w} = (A^+\vec{b})^*\vec{w} = \vec{b}^*A^{+^*}\vec{w} \\ &\overset{II}{=} \vec{b}^*\,(A^+AA^+)^*\,\vec{w} = \vec{b}^*A^{+^*}\,(A^+A)^*\,\vec{w} \\ &\overset{IV}{=} \vec{b}^*A^{+^*}A^+A\vec{w} \\ &= \vec{b}^*A^{+^*}A^+\underbrace{A\vec{w}}_{0} \overset{(3)}{=} 0, \end{aligned}$$

which means that $u \perp w$. Using Pythagoras theorem, we get $|\vec{x}| \geqslant |\vec{u}|$, which means that $\vec{u}$ is pseudosolution with minimal length. By Pythagoras theorem $|\vec{x}| = |\vec{u}| \Leftrightarrow \vec{w} = 0 \Leftrightarrow \vec{x} = \vec{u}$, which means that $\vec{u}$ is unique. □

Problem 1. Finding a pseudosolution

Find a pseudosolution of the following inconsistent system

$$A = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}.$$

Solution: Then pseudosolution can be found by the formula

$$\vec{u} = A^{+}\vec{b}.$$

Since $\operatorname{rank}(A) = 1$, we could apply following formula for pseudoinverse of matrix A

$$A^{+} = \frac{1}{\sum a_{ij}^2}A^{*} = \frac{1}{1^2 + 1^2 + 2^2 + 2^2}\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}^{*} = \frac{1}{10}\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}.$$

Finally, we get a pseudosolution of the system

$$\vec{u} = \frac{1}{10}\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 18 & 9 \end{bmatrix}.$$

■

Lemma: 2

$$A\vec{x} = \vec{b} \quad \Leftrightarrow \quad A\vec{x} = AA^{+}\vec{b}.$$

Prop

$\vec{x}$ is a pseudosolution ⇔ $\vec{x}$ is a solution of the "normal system", which means

$$A\vec{x} = \vec{b} \quad \Leftrightarrow \quad A^{*}A\vec{x} = A^{*}\vec{b}.$$

Proof: Let $\vec{x}$ be a pseudosolution of the system $A\vec{x} = \vec{b}$. Let us prove that pseudosolution $\vec{x}$ is also a solution the normal system $A^{*}A\vec{x} = A^{*}\vec{b}$

$$\begin{aligned} A^{*}A\vec{x} &\overset{\text{Lemma 2}}{=} A^{*}AA^{+}\vec{b} \\ &\overset{\text{axiom III}}{=} A^{*}(AA^{+})^{*}\vec{b} = (AA^{+}A)^{*}\vec{b} \\ &\overset{\text{axiom I}}{=} A^{*}\vec{b}. \end{aligned}$$

Let $\vec{x}$ be a solution of normal system $A^{*}A\vec{x} = A^{*}\vec{b}$. Let us prove that solution of normal system $\vec{x}$ is also a pseudosolution of the system $A\vec{x} = \vec{b}$

$$\begin{aligned} A\vec{x} &\overset{\text{axiom I}}{=} AA^{+}A\vec{x} \\ &\overset{\text{axiom III}}{=} (AA^{+})^{*}A\vec{b} = A^{+^{*}}A^{*}A\vec{x} \\ &\overset{\text{normal system}}{=} A^{+^{*}}A^{*}\vec{b} = (AA^{+})^{+}\vec{b} \\ &\overset{\text{axiom III}}{=} AA^{+}\vec{b}. \end{aligned}$$

□

Prop

All pseudosolutions of $A\vec{x} = \vec{b}$ are given by the formula

$$\vec{x} = A^{+}\vec{b} - (A^{+}A - I)\vec{y}, \tag{4}$$

where $\vec{y} \in \mathbb{C}^{n}$ is an arbitrary vector.

Proof: First, let us check that (4) is indeed a pseudosolution.

$$A\vec{x} \overset{(4)}{=} AA^{+}\vec{b} + \underbrace{(AA^{+}A - A)}_{0}\vec{y} \overset{\text{axiom I}}{=} AA^{+}\vec{b},$$

which by Lemma 2 is pseudosolution.

Now, let us prove that all pseudosolutions are in the form (4). Let $\vec{x}$ be a pseudosolutions. If we set $\vec{y} = \vec{x}$ in (4), we get

$$\begin{aligned}
\vec{x} &= A^{+}\vec{b} - (A^{+}A - I)\vec{y} \\
&= A^{+}\vec{b} - (A^{+}A - I)\vec{x} \\
&= A^{+}\vec{b} - A^{+}A\vec{x} + \vec{x} \\
&\overset{\text{Lemma 2}}{=} A^{+}\vec{b} - A^{+}AA^{+}\vec{b} + \vec{x} \\
&= \underbrace{(A^{+} - A^{+}AA^{+})}_{0}\vec{b} + \vec{x} \\
&\overset{\text{axiom II}}{=} \vec{x}.
\end{aligned}$$

□

Problem 2. Finding all pseudosolutions

Find all pseudosolutions of the system from problem 1.

Solution: We have already obtained one pseudosolution

$$\vec{u} = \frac{1}{10}\begin{bmatrix} 18 \\ 9 \end{bmatrix}.$$

Now we need to obtain $A^{+}A - I$

$$A^{+}A - I = \frac{1}{10}\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 8 & 4 \\ 4 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -0.2 & 0.4 \\ 0.4 & -0.8 \end{bmatrix}.$$

Finally, all pseudosolutions are given by the formula (4)

$$\vec{x} = \frac{1}{10}\begin{bmatrix} 18 \\ 9 \end{bmatrix} - \begin{bmatrix} -0.2 & 0.4 \\ 0.4 & -0.8 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1.8 + 0.2y_1 - 0.4y_2 \\ 0.9 - 0.4y_1 + 0.8y_2 \end{bmatrix}.$$

■

Linear regression problem

Let $y, x_1, x_2, \dots, x_n$ be variables. We believe that dependent variable y could be approximated by linear combination of independent variables $x_1, x_2, \dots, x_n$. That is,

$$y \approx a_1 x_1 + a_2 x_2 + \dots + a_n x_n.$$

The problem is to find coefficients $a_1, a_2, \dots, a_n$. In order to find them we conduct m experiments (samples) and get following dataset of observations

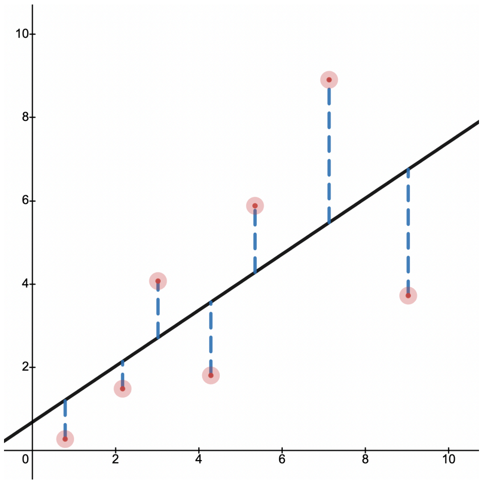
$$\begin{cases} y_1 \approx a_1 x_{11} + \ldots + a_n x_{n1} \\ y_2 \approx a_2 x_{12} + \ldots + a_n x_{n2} \\ \qquad\qquad \vdots \\ y_m \approx a_1 x_{1m} + \ldots + a_n x_{nm} \end{cases}$$

where $x_{1i}, x_{2i}, \dots, x_{ni}, y_i$ is a result of i-th experiment. Clearly, the problem could is equivalent to matrix equation of the form

$$X\vec{a} \approx \vec{y},$$

where

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{nm} \end{bmatrix}, \quad \vec{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$



In linear regression, the observations (red) are assumed to be the result of random deviations (blue) from an underlying relationship (black) between a dependent variable (y) and an independent variable (x).

The solution of the problem is least square solution (pseudosolution)

$$\vec{a} = X^{+}\vec{y}.$$

Problem 3. Linear regression

Find linear approximation of temperature using following dataset of temperature observations. Find temperature on day 4.

#	Day	Temperature
1	1	19°
2	2	18°
3	2	20°
4	3	15°

<u>Solution:</u> Let us denote day and temperature of the row i by x_i and y_i respectively. We assume that temperature ($y(x)$) could be approximated by linear model

$$y(x) = ax + b \in \mathbb{R},$$

where $x \in \mathbb{R}$ is number of the day and $a, b \in \mathbb{R}$ are unknown coefficients. This means that we need to find pseudosolution of the following system

$$X\vec{a} = \vec{y},$$

where

$$X = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 19 \\ 18 \\ 20 \\ 15 \end{bmatrix}, \quad \vec{a} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

So we need to evaluate the following $\vec{a} = X^{+}\vec{y}$. Since X is full column rank, we get

$$\begin{aligned} X^{+} &= (X^{*}X)^{-1}X^{*} = \left(2\begin{bmatrix} 9 & 4 \\ 4 & 2 \end{bmatrix}\right)^{-1} X^{*} \\ &= \frac{1}{2}\frac{1}{2}\begin{bmatrix} 2 & -4 \\ -4 & 9 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix} -2 & 0 & 0 & 2 \\ 5 & 1 & 1 & -3 \end{bmatrix} \end{aligned}$$

Hence, we get

$$\vec{a} = X^{+}\vec{y} = \frac{1}{4}\begin{bmatrix} -2 & 0 & 0 & 2 \\ 5 & 1 & 1 & -3 \end{bmatrix}\begin{bmatrix} 19 \\ 18 \\ 20 \\ 15 \end{bmatrix} = \begin{bmatrix} -2 \\ 22 \end{bmatrix}.$$

This means that linear model $y(x)$ is

$$y(x) = -2x + 22.$$

According to linear model $y(x)$, on day 4 temperature will be $y(4) = 14^{\circ}$. ■