Method of least squares

A ball rolls along a narrow straight trough (Fig.1). The distance S traversed is a variable depending on the elapsed time t.

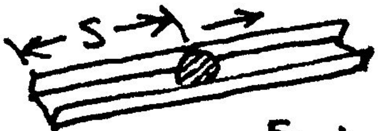

Fig.1

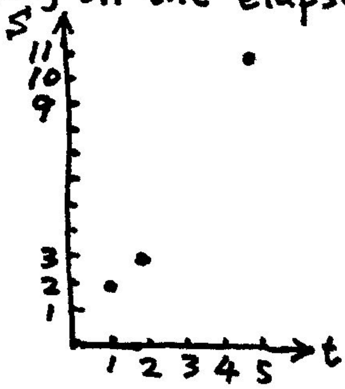

Fig.2

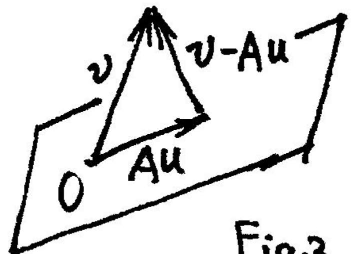

Fig.3

By measuring the variables $\binom{t}{S}$ ——— see Fig.2, we want to find the velocity of the ball. Errors always exist in measuring, thus in Fig.2, the three points of measured data do not situate exactly on a straight line. We thus need to find a best-fitting straight line for the three points $\binom{1}{2}$, $\binom{2}{3}$, $\binom{5}{11}$, in Fig.2.

We expect a straight line $a\cdot t+b=S$ which makes $\begin{cases} 1a+b=2 \\ 2a+b=3 \\ 5a+b=11 \end{cases}$ i.e. $\begin{pmatrix}1&1\\2&1\\5&1\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}=\begin{pmatrix}2\\3\\11\end{pmatrix}$

This is a linear equation system with contradiction. There are no $\binom{a}{b}$ which makes left side = right side. However, we could find $\binom{a}{b}$ such that $\begin{pmatrix}1&1\\2&1\\5&1\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}$ is the closest to $\begin{pmatrix}2\\3\\11\end{pmatrix}$, i.e. the two vectors in $\mathbb{R}^3$ have the minimum distance.

i.e. the modulus of the difference vector $\begin{pmatrix}1&1\\2&1\\5&1\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}-\begin{pmatrix}2\\3\\11\end{pmatrix}$ assumes its minimum.

which is the total squared error:
$(S_1-(at_1+b))^2+(S_2-(at_2+b))^2+(S_3-(at_3+b))^2$

Analysis: when $\binom{a}{b}$ varies, $\begin{pmatrix}1&1\\2&1\\5&1\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}$ spans a vector space of dimension 2 (i.e. a plane) in $\mathbb{R}^3$, whose base is $\begin{pmatrix}1\\2\\5\end{pmatrix}$, $\begin{pmatrix}1\\1\\1\end{pmatrix}$ (i.e. the image of $\binom{1}{0}$, $\binom{0}{1}$ under the transformation $\begin{pmatrix}1&1\\2&1\\5&1\end{pmatrix}$, resp.). So $\underset{A}{\begin{pmatrix}1&1\\2&1\\5&1\end{pmatrix}}\underset{u}{\begin{pmatrix}a\\b\end{pmatrix}}$ is the closest to $\underset{v}{\begin{pmatrix}2\\3\\11\end{pmatrix}}$

$\overset{(*)}{\Longleftrightarrow}$ $Au-v \perp \begin{pmatrix}1\\2\\5\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}$ (See Fig.3)

$\Longleftrightarrow$ $(Au-v)\cdot\begin{pmatrix}1\\2\\5\end{pmatrix}=0$, $(Au-v)\cdot\begin{pmatrix}1\\1\\1\end{pmatrix}=0$

matrix multiplication

i.e. $(1\ 2\ 5)(Au-v)=0$, $(1\ 1\ 1)(Au-v)=0$

i.e. $\begin{pmatrix}1&2&5\\1&1&1\end{pmatrix}(Au-v)=\begin{pmatrix}0\\0\end{pmatrix}$

i.e. $\underset{A^T}{\begin{pmatrix}1&2&5\\1&1&1\end{pmatrix}}\left[\underset{A}{\begin{pmatrix}1&1\\2&1\\5&1\end{pmatrix}}\begin{pmatrix}a\\b\end{pmatrix}-\begin{pmatrix}2\\3\\11\end{pmatrix}\right]=\begin{pmatrix}0\\0\end{pmatrix}$ i.e. $\begin{pmatrix}30&8\\8&3\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}-\begin{pmatrix}63\\16\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$

A^TA is a symmetric matrix

⋮

Notice that the equivalence (*), and thus the above method, is of generality, i.e. also valid in $\mathbb{R}^4, \mathbb{R}^5, \mathbb{R}^6, \cdots$ (See <u>Appendix I</u>). For example if we measure 5 pairs of data $\binom{t}{s}$, then we would get:

$\begin{pmatrix}\vdots&\vdots\\\vdots&\vdots\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}-\begin{pmatrix}\vdots\\\vdots\end{pmatrix}\perp\begin{pmatrix}\vdots\\\vdots\end{pmatrix},\begin{pmatrix}\vdots\\\vdots\end{pmatrix}$ which is in $\mathbb{R}^5$ and implies $\begin{pmatrix}\cdots\cdots\\\cdots\cdots\end{pmatrix}\left[\begin{pmatrix}\vdots&\vdots\\\vdots&\vdots\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}-\begin{pmatrix}\vdots\\\vdots\end{pmatrix}\right]=\begin{pmatrix}0\\0\end{pmatrix}$

i.e. $\begin{pmatrix}\cdot&\cdot\\\cdot&\cdot\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}-\begin{pmatrix}\cdot\\\cdot\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$.

Another problem: find a best-fitting circle for the 4 given points: $(1,2), (1,5), (3,1), (4,2)$.

We expect a circle $f(x,y)=x^2+y^2+ax+by+c=0$, which makes:

$$(\dagger)\begin{cases} f(1,2)=1^2+2^2+1a+2b+c=0\\ f(1,5)=1^2+5^2+1a+5b+c=0\\ f(3,1)= \qquad\qquad\qquad =0\\ f(4,2)= \qquad\qquad\qquad =0\end{cases} \xrightarrow[\text{separate constants}]{\text{i.e.}} \begin{pmatrix}1&2&1\\1&5&1\\3&1&1\\4&2&1\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}=\begin{pmatrix}-5\\-26\\-10\\-20\end{pmatrix}$$

find $\begin{pmatrix}a\\b\\c\end{pmatrix}$ such that $\begin{pmatrix}1&2&1\\1&5&1\\3&1&1\\4&2&1\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}$ is closest to $\begin{pmatrix}-5\\-26\\-10\\-20\end{pmatrix}$ in $\mathbb{R}^4$.

$$\begin{pmatrix}1&2&1\\1&5&1\\3&1&1\\4&2&1\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}-\begin{pmatrix}-5\\-26\\-10\\-20\end{pmatrix}\perp\begin{pmatrix}1\\1\\3\\4\end{pmatrix},\begin{pmatrix}2\\5\\1\\2\end{pmatrix},\begin{pmatrix}1\\1\\1\\1\end{pmatrix}.$$

$$\begin{pmatrix}1&1&3&4\\2&5&1&2\\1&1&1&1\end{pmatrix}\left[\begin{pmatrix}1&2&1\\1&5&1\\3&1&1\\4&2&1\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}-\begin{pmatrix}-5\\-26\\-10\\-20\end{pmatrix}\right]=\begin{pmatrix}0\\0\\0\end{pmatrix}\cdots\cdots(\ddagger)$$

$\vdots$

Final notes: the quantity which is made minimum by the solution of $(\ddagger)$ is the modulus of $\begin{pmatrix}1&2&1\\1&5&1\\3&1&1\\4&2&1\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}-\begin{pmatrix}-5\\-26\\-10\\-20\end{pmatrix}$, i.e. the modulus of $\begin{pmatrix}1^2+2^2+1a+2b+c\\1^2+5^2+1a+5b+c\\\vdots\end{pmatrix}$ in $(\dagger)$, i.e. the sum of squares of the values of $f(x,y)$ at the given points $(1,2)$, $(1,5),(3,1),(4,2)$, all of which are expected to be zero.

The above method is called <u>the method of least squares</u>, and is valid for finding the <u>best-fitting</u> curve of any given shape.

In 1809, Gauss published this method of least squares to find the best-fitting ellipse orbit $x^2+ay^2+bxy+cx+dy+e=0$ from observation data of celestial objects.

Appendix I. (*) is valid for $\mathbb{R}^4, \mathbb{R}^5, \mathbb{R}^6, \cdots$

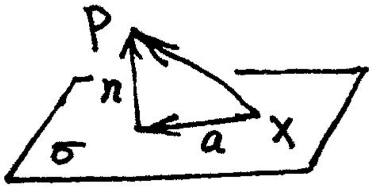

Fig. 4

In $\mathbb{R}^3$ (See Fig. 4), given a plane σ and a point P, when a point X moves on the plane σ, the distance $|\overrightarrow{PX}|$ assumes its minimum when X is such that $\overrightarrow{PX} \perp \sigma$.

This is clear by geometry: in a right triangle, the hypotenuse is always longer than a leg.

Rewrite the geometric argument in the language of vectors, we see that the conclusion is also true for any given point P and sub-space σ in the higher dimensional spaces $\mathbb{R}^4, \mathbb{R}^5, \mathbb{R}^6, \cdots$ (See Fig. 4):

$$
\begin{aligned}
|a+n|^2 &= (a+n)\cdot(a+n) \\
&= a\cdot a + 2a\cdot n + n\cdot n \\
&= |a|^2 + |n|^2 > |n|^2.
\end{aligned}
$$