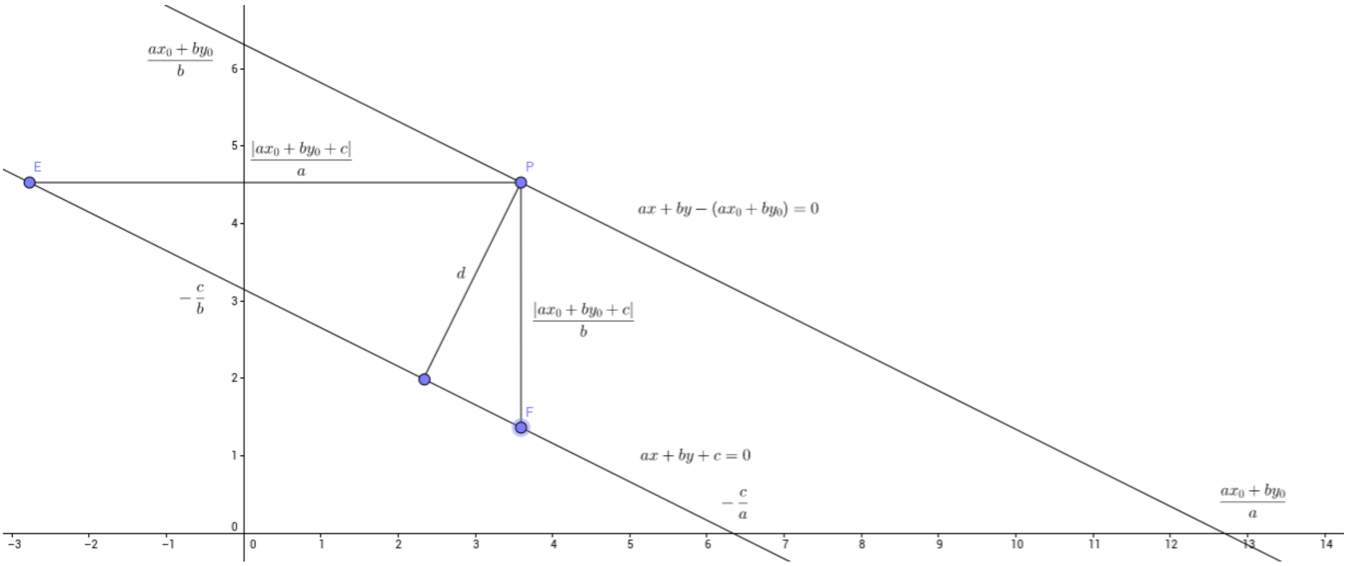
Exercise 1. Total least squares line fitting. (Pen and paper problem)
An overview of least squares line fitting is presented on the slide 13 of Lecture 4. Study it in detail and present the derivation with the following stages:

1) Given a line $ax+by-d=0$, where the coefficients are normalized so that $a^2+b^2=1$, show that the distance between a point (x_i, y_i) and the line is $|ax_i+by_i-d|$.

The distance from a point to a line is the perpendicular distance of the point to the line.



In this image, let's draw an arbitrary line $ax+by+c=0$ and a point (x_0, y_0). The equation of the line that passes through the point and parallel to the line has the equation $ax+by-(ax_0+by_0)=0$. The legs of the right triangle EPF can be obtained by the differences between the interception of x and y-axes of the two lines. In the figure,

$$EP=\frac{ax_0+by_0}{b}-(-\frac{c}{b})=\frac{ax_0+by_0+c}{b} \text{ and } PF=\frac{ax_0+by_0}{a}-(-\frac{c}{a})=\frac{ax_0+by_0+c}{a}$$

Since we are only interested in the magnitude, we have

$$EP=\frac{|ax_0+by_0+c|}{b} \text{ and } PF=\frac{|ax_0+by_0+c|}{a}$$

The hypothenuse EF is calculated using Pythagorean theorem:

$$EF=\sqrt{EP^2+PF^2}=\sqrt{\left(\frac{|ax_0+by_0+c|}{a}\right)^2+\left(\frac{|ax_0+by_0+c|}{b}\right)^2}=|ax_0+by_0+c|\frac{\sqrt{a^2+b^2}}{ab}$$

Denote d the distance of P to the line. Due to the area equality of the right triangle, we have:

$$d\cdot EF=EP\cdot PF \Rightarrow d\,|ax_0+by_0+c|\frac{\sqrt{a^2+b^2}}{ab}=\frac{|ax_0+by_0+c|}{b}\frac{|ax_0+by_0+c|}{a}$$

$$=> d\left|ax_0+by_0+c\right|\frac{\sqrt{a^2+b^2}}{ab}=\frac{\left|ax_0+by_0+c\right|^2}{ab}$$

$$=> d\left|ax_0+by_0+c\right|\sqrt{a^2+b^2}=\left|ax_0+by_0+c\right|^2$$

$$=> d=\frac{\left|ax_0+by_0+c\right|^2}{\left|ax_0+by_0+c\right|\sqrt{a^2+b^2}}=\frac{\left|ax_0+by_0+c\right|}{\sqrt{a^2+b^2}}\text{(proven)}$$

2) Thus, given n points (x_i, y_i), $i = 1, \ldots, n$, the sum of squared distances between the points and the line is $E = \Sigma_{i=1}^{n}(ax_i + by_i - d)^2$. In order to find the minimum of E, compute the partial derivative $\partial E/\partial d$, set it to zero, and solve d in terms of a and b.

The partial derivative of the sum of squared differences with respect to d is:

$$\frac{\partial E}{\partial d}=\sum_{i=1}^{n}\left(ax_i+by_i-d\right)^2\frac{\partial}{\partial d}=\sum_{i=1}^{n}\left(a^2x_i^2+b^2y_i^2+2abx_iy_i-2ax_id-2by_id+d^2\right)\frac{\partial}{\partial d}$$

$$=\sum_{i=1}^{n}\left(-2ax_i-2by_i-2d\right)=\sum_{i=1}^{n}-2\left(ax_i+by_i-d\right)$$

The minimum of E when the partial derivative is equal to 0:

$$\frac{\partial E}{\partial d}=\sum_{i=1}^{n}-2\left(ax_i+by_i-d\right)=0 => \sum_{i=1}^{n}d-\sum_{i=1}^{n}ax_i-\sum_{i=1}^{n}by_i=0$$

$$=> nd-a\sum_{i=1}^{n}x_i-b\sum_{i=1}^{n}y_i=0 => nd=a\sum_{i=1}^{n}x_i+b\sum_{i=1}^{n}y_i$$

$$=> d=a\frac{1}{n}\sum_{i=1}^{n}x_i+b\frac{1}{n}\sum_{i=1}^{n}y_i => d=a\overline{x}+b\overline{y}\ \text{(answer)}$$

3) Substitute the expression obtained for d to the formula of E, and show that then $E = (a\ b)U^\top U(a\ b)^\top$, where matrix U depends on the point coordinates (x_i, y_i).

Plugging in d from part (2) into the error formula:

$$E=\sum_{i=1}^{n}\left(ax_i+by_i-\left(a\overline{x}+b\overline{y}\right)\right)^2=\sum_{i=1}^{n}\left(a\left(x_i-\overline{x}\right)+b\left(y_i-\overline{y}\right)\right)^2$$

Let $U=\begin{bmatrix}x_1-\overline{x} & y_1-\overline{y}\\ \vdots & \vdots\\ x_n-\overline{x} & y_n-\overline{y}\end{bmatrix}$ => $E=\left\|\begin{bmatrix}x_1-\overline{x} & y_1-\overline{y}\\ \vdots & \vdots\\ x_n-\overline{x} & y_n-\overline{y}\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}\right\|^2=\left(U\begin{bmatrix}a\\ b\end{bmatrix}\right)^T\left(U\begin{bmatrix}a\\ b\end{bmatrix}\right)$

⇨ $E=\left(U\begin{bmatrix}a\\ b\end{bmatrix}\right)^T\left(U\begin{bmatrix}a\\ b\end{bmatrix}\right)=\begin{bmatrix}a & b\end{bmatrix}U^TU\begin{bmatrix}a\\ b\end{bmatrix}$ (proven)

4) Thus, the task is to minimize $||U(a\ b)^\top||$ under the constraint $a^2 + b^2 = 1$. The solution for $(a\ b)^\top$ is the eigenvector of $U^\top U$ corresponding to the smallest eigenvalue, and d can be solved thereafter using the expression obtained above in the stage two.

We can use the Lagrange multiplier technique to find the answer