On Finding the Shortest Distance of a Point From a Line: Which Method Do You Prefer?

Bhalchandra W Gore
Centre for Modeling and Simulation
Savitribai Phule Pune University
Ganesh Khind
Pune 411 007, India.
Email:
bwgore@cms.unipune.ac.in

The formula for the shortest distance of a point from a line can be derived in several different ways. Some typical methods are taught at the elementary (i.e., high-school and junior college) level. However, solving such 'school-book' problems using advanced mathematical methods is often overlooked and neglected. This article illustrates how this formula can be derived in various ways. Such a comparison will not only encourage the reader to explore and understand how and why mathematical techniques work, but it will also help understand the common thread between different branches of mathematics. This exercise also shows that 'the best way' to solve a mathematical problem is a misnomer.

1. Prologue

The shortest distance of a point from a line (or in general, from a hyperplane in the $\mathbb{R}^n$ Euclidean space) can be of use in many practical applications. In statistical analysis, a regression line is often fitted through the data points such that the variance of deviations is minimum. In the case of orthogonal regression, these deviations are nothing but the shortest distance of each data point from the fitted line. The second application is the design of a network in which the supply lines are tapped to feed the subscribers. The shortest distance of the subscriber from the line provides an estimate of the total length of the pipe or cable required to build such a network. The examples are water and CNG gas supply lines, the electric-supply distribution network, fluid supply lines in the factories, etc. One more application is in disaster management in the vicinity of high voltage electrical cables. A proper safety distance from such a cable must be maintained to avoid calamities. Hence, the shortest distance of any point in the surrounding terrain from the cable becomes a useful design parameter for laying

In statistical analysis, a regression line is often fitted through the data points such that the variance of deviations is minimum.

Keywords
Shortest distance, mathematical methods, optimization, Lagrange multiplier.

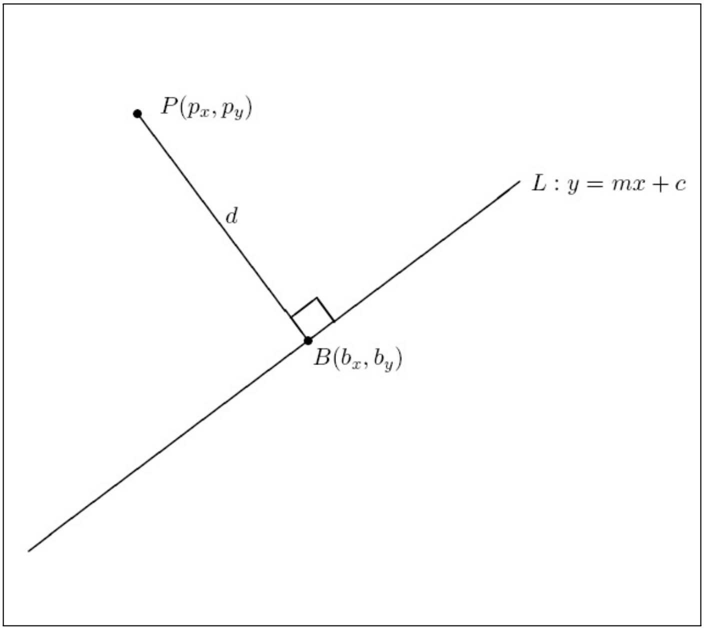

Figure 1. The diagram showing the details of the problem at hand.

the cable, as well as for constructing any structure in the vicinity of the high voltage cable.

Figure 1 depicts the different entities related to this problem: Point $P\,(p_x, p_y)$ is the given point. We wish to find its shortest distance from the line $L : y = mx + c$. Let $B(b_x, b_y)$ be the point on line L such that $PB \perp L$. It can be shown, using the Pythagoras theorem, that the perpendicular distance $d = l(PB)$ (see the *Figure*) is the shortest distance between point P and line L. Let us first derive the formula for the shortest distance using four elementary and familiar methods (Section 2). Later on in Section 3, let us derive the same formula using two mathematical optimization methods.

2. Elementary Methods

2.1 *Algebraic Method*

The given line has the slope m. Hence, the perpendicular line PB has the slope $-\frac{1}{m}$ and it passes through the point $P(p_x, p_y)$. Using the slope-point form, the equation of the line PB is :

$$y - p_y = \frac{-1}{m}(x - p_x). \tag{1}$$

The point B being the point of intersection of line L and line PB, we solve the equations of these two lines simultaneously to obtain the coordinates of point B :

$$b_x = \frac{mp_y + p_x - mc}{1 + m^2}; \qquad b_y = \frac{m^2 p_y + mp_x + c}{1 + m^2}. \tag{2}$$

Having done this, the required distance $d = l(PB)$ can be readily calculated as:

$$\begin{aligned}
d &= \sqrt{(b_x - p_x)^2 + \left(b_y - p_y\right)^2} \\
&= \sqrt{\left\{\frac{mp_y + p_x - mc}{1 + m^2} - p_x\right\}^2 + \left\{\frac{m^2 p_y + mp_x + c}{1 + m^2} - p_y\right\}^2} \\
&= \frac{1}{1 + m^2}\left\{[mp_y + p_x - mc - (1 + m^2)p_x]^2\right\} \\
&+ \left\{[m^2 p_y + mp_x + c - (1 + m^2)p_y]^2\right\}^{1/2} \\
&= \frac{1}{1 + m^2}\sqrt{m^2\left\{p_y - c - mp_x\right\}^2 + \left\{mp_x + c - p_y\right\}^2} \\
&= \left|\frac{mp_x - p_y + c}{\sqrt{1 + m^2}}\right|. \qquad (3)
\end{aligned}$$

It is worthwhile to note that in the step before the final one, the two squared terms under the square-root sign are of the form $(\alpha - \beta)^2$ and $(\beta - \alpha)^2$ and can therefore be taken out as a common factor. This leaves $\sqrt{1 + m^2}$ in the denominator. Also, to ensure non-negativity of the distance, we take the absolute value of the square-root obtained.

2.2 *Using Vectors*

The line L can be expressed in the two-dimensional vector form by noting that the point $C(0, c)$ lies on it, and the line is parallel to the vector $\vec{l} = \begin{bmatrix} 1 \\ m \end{bmatrix}$. Hence, coordinates of any point on line L

are given by the vector equation $\vec{r} = \vec{C} + t\,\vec{l}$, i.e.,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ c \end{bmatrix} + t \begin{bmatrix} 1 \\ m \end{bmatrix} = \begin{bmatrix} t \\ mt + c \end{bmatrix},$$

where $t \in \mathbb{R}$ is the parameter indicating the distance of the point (x, y) from C, and we have denoted the position vector of any point A by $\vec{A}$. Since the point B lies on this line, its coordinates must be in the same form as above. However, the value of parameter t corresponding to point B is not known yet. To obtain its value, we make use of the fact that $\overrightarrow{PB} \perp \vec{l}$, whereby, their scalar product should be zero (i.e., $\overrightarrow{PB} \cdot \vec{l} = 0$). Thus,

$$\overrightarrow{PB} = \vec{B} - \vec{P} = \begin{bmatrix} t - p_x \\ mt + c - p_y \end{bmatrix}$$

$$\therefore \quad \begin{bmatrix} t - p_x \\ mt + c - p_y \end{bmatrix} \cdot \begin{bmatrix} 1 \\ m \end{bmatrix} = 0$$

$$\therefore \quad (t - p_x) + m\left(mt + c - p_y\right) = 0$$

$$\therefore \quad t = \frac{mp_y + p_x - mc}{1 + m^2}.$$

Substituting this value of t and evaluating the components of $\overrightarrow{PB}$ we get:

$$\begin{bmatrix} t - p_x \\ mt + c - p_y \end{bmatrix} = \begin{bmatrix} \frac{mp_y + p_x - mc}{1+m^2} - p_x \\ m\left(\frac{mp_y + p_x - mc}{1+m^2}\right) + c - p_y \end{bmatrix}.$$

The required distance d is the magnitude of $\overrightarrow{PB}$. Hence,

$$d = |\overrightarrow{PB}| = \left\{ \left[\frac{mp_y + p_x - mc}{1 + m^2} - p_x \right]^2 + \left[m\left(\frac{mp_y + p_x - mc}{1 + m^2} \right) + c - p_y \right]^2 \right\}^{1/2}$$

One can easily verify that this expression simplifies to the required formula, (3).

This method is useful, in general, for finding the shortest distance of a point in the n-dimensional Euclidean space $\mathbb{R}^n$ from a hyperplane. Let the equation of the hyperplane be $(\vec{r} - \vec{q}) \cdot \vec{p} = 0$, where

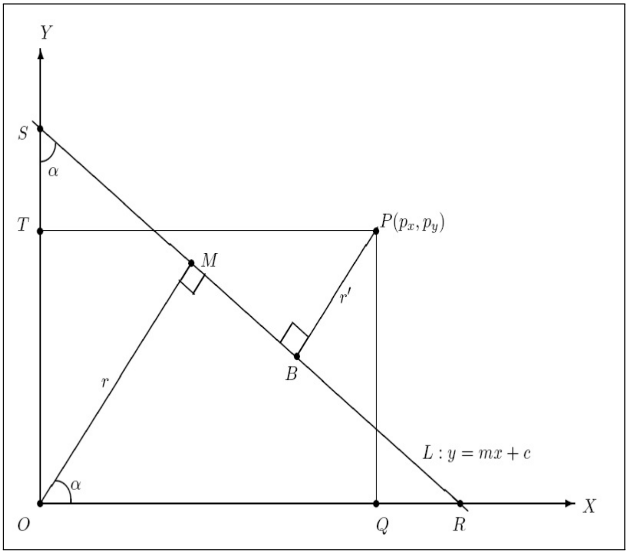

Figure 2. Geometric entities used in Sections 2.3 and 2.4. Here, $\angle OSR = \alpha$ because it is the angle between lines OS and SR which are perpendicular to lines OR and OM respectively.

$\vec{r}$ represents any point on the hyperplane, $\vec{q}$ a particular known point on the hyperplane, and $\vec{p}$ is the normal to the hyperplane. The shortest distance of any point $U(u_1, u_2, \cdots, u_n)$ in $\mathbb{R}^n$ from this hyperplane is given by:

$$d = \frac{\left|(\vec{u} - \vec{q}) \cdot \vec{p}\right|}{\left\|\vec{p}\right\|},$$

where $\|\cdots\|$ has been used to represent the Euclidean norm (i.e. magnitude) of vector.

2.3 *A Method Based on Coordinate Geometry*

This method is the most popular method in the elementary level textbooks. It involves the basic concepts such as the intercepts that the line makes with the coordinate axes, the basic trigonometric functions – sin and cos, and the change of coordinates due to shift of origin to a different point.

See *F*igure 2: Let OM be the normal from the origin to the line L and $l(OM) = r$. Let $\angle XOM = \alpha$. Let the line L intercept X-axis at point R and Y-axis at point S. We can find the coordinates of the two intercepts, using equation of line L, as $R(-\frac{c}{m}, 0)$ and $S(0, c)$.

From the right-angled $\triangle OMR$ we have,

$$\cos\alpha = \frac{OM}{OR} = \frac{-mr}{c},$$

and from the right-angled $\triangle OMS$ we have,

$$\sin\alpha = \frac{OM}{OS} = \frac{r}{c}.$$

From the trigonometric identity $\sin^2\alpha + \cos^2\alpha = 1$ we get:

$$\frac{m^2 r^2}{c^2} + \frac{r^2}{c^2} = 1.$$

This gives the distance of line L from origin,

$$r = \left|\frac{c}{\sqrt{1+m^2}}\right|. \tag{4}$$

Now, to find the distance d of point P from line L, we shift the origin to point P without changing the directions of the coordinate axes. Let the coordinates in the new system be represented by (x', y'). Then,

$$x' = x - p_x; \qquad y' = y - p_y.$$

The equation of the line L gets transformed to new coordinate system as:

$$\begin{aligned} y &= mx + c \\ \therefore y' + p_y &= m(x' + p_x) + c \\ \therefore y' &= mx' + \left\{mp_x - p_y + c\right\} \\ \therefore y' &= mx' + c', \end{aligned}$$

where $c' = mp_x - p_y + c$. According to (4), the distance of this transformed line from the transformed origin is,

$$\begin{aligned} r' &= \left|\frac{c'}{\sqrt{1+m^2}}\right| \\ &= \left|\frac{mp_x - p_y + c}{\sqrt{1+m^2}}\right|. \end{aligned}$$

As the new origin is point P, the line is still L, and $r' = l(PB)$ is the desired distance, we obtain the same result as in (3).

2.4 *Using Areas of Polygons*

Let $PQ \perp X$ and $PT \perp Y$. Then, we have $l(PQ) = p_y$ and $l(PT) = p_x$. Consider the quadrilateral $\Box PSOR$. It can be thought of as a union of two pairs of triangles: (1) $\triangle PSR \bigcup \triangle SOR$ and (2) $\triangle ORP \bigcup \triangle OPS$. These pairs are based on the partitions created by the two diagonals RS and OP. Note that these polygons are not explicitly shown in the figure provided, they need to be imagined.

We can write:

$$\mathcal{A}\{\Box PSOR\} = \mathcal{A}\{\triangle PSR\} + \mathcal{A}\{\triangle SOR\},$$

and,

$$\begin{aligned}
&\mathcal{A}\{\Box PSOR\} = \mathcal{A}\{\triangle ORP\} + \mathcal{A}\{\triangle OPS\} \\
&\therefore \mathcal{A}\{\triangle PSR\} + \mathcal{A}\{\triangle SOR\} = \mathcal{A}\{\triangle ORP\} + \mathcal{A}\{\triangle OPS\} \\
&\therefore \frac{1}{2} \times l(SR) \times l(PB) + \frac{1}{2} \times l(OS) \times l(OR) \\
&\quad = \frac{1}{2} \times l(OR) \times l(PQ) + \frac{1}{2} \times l(OS) \times l(PT) \\
&\therefore \left\{ \sqrt{\left(\frac{-c}{m}\right)^2 + c^2} \right\} \times r' + c \times \frac{-c}{m} = \frac{-c}{m} \times p_y + c \times p_x \\
&\therefore r' = \left| \frac{mp_x - p_y + c}{\sqrt{1 + m^2}} \right| .
\end{aligned}$$

The expression in (3) is thus established again.

3. Optimization-Based Methods

3.1 *A Simple Calculus-Based Derivation*

We now approach the problem using the advanced techniques based on the concept of mathematical optimization. We look at the problem of minimizing the distance ρ of point P from any point $Q(q_x, q_y)$ on line L, whereby $q_y = mq_x + c$. The distance ρ

can thus be formulated as a function of a single variable q_x:

$$\begin{aligned}\rho^2 &= (q_x - p_x)^2 + \left(q_y - p_y\right)^2 \\ &= (q_x - p_x)^2 + \left(mq_x + c - p_y\right)^2 \\ \therefore \rho &= \sqrt{(q_x - p_x)^2 + \left(mq_x + c - p_y\right)^2}\end{aligned}$$

Clearly, d is the minimum value of ρ; i.e., $d = \min \rho$. To find $\min \rho$, we employ the necessary first-derivative condition $\frac{d\rho}{dq_x} = 0$ which yields,

$$\frac{1}{2\rho}\left[2\left(q_x - p_x\right) + 2m\left(mq_x + c - p_y\right)\right] = 0$$
$$\therefore q_x = \frac{p_x + mp_y - mc}{1 + m^2},$$

which is exactly the expression we obtained in (2) for b_x. Needless to say, an equivalent expression can be obtained using the condition $q_y = mq_x + c$ and therefore, the point Q coincides with point B. Obviously, $\rho = d$ at this point, and we get the required formula (3). To make sure that the point thus obtained is indeed the location of a minimum, the second-derivative test (i.e. $\frac{d^2\rho}{dq_x^2} > 0$), should be applied. We leave this exercise to the reader.

3.2 *A Constrained Optimization Method*

Let us now see how the same formula can be derived using a method of optimizing a function $f(\vec{u})$ under an equality constraint of the form $h(\vec{u}) = 0$ in the n dimensional space $\mathbb{R}^n$. This formulation is often referred to as optimization using Lagrange coefficients. According to the theory of constrained mathematical optimization, there exists a Lagrange multiplier λ such that $\nabla f(\vec{u}) + \lambda \nabla h(\vec{u}) = 0$. Here, we assume that both the functions f and h have continuous first order partial derivatives so that the gradient can be defined, and we do so using the operator $\nabla \equiv \left[\frac{\partial}{\partial u_1}\ \frac{\partial}{\partial u_2}\ \cdots\ \frac{\partial}{\partial u_n}\right]^T$, where $[]^T$ denotes the transpose. The optimum value of function f under the constraint $h(\vec{u}) = 0$ can be found, provided the value of Lagrange multiplier λ obtained in

The optimum value of function f under the constraint $h(\vec{u}) = 0$ can be found, provided the value of Lagrange multiplier λ obtained in this manner is consistent with the given problem.

this manner is consistent with the given problem. If the number of constraints is more than one, there is a set of constraint functions $h_1, h_2, \cdots$, and the Lagrange multipliers are components of a vector and the system of equations $\nabla f(\vec{u}) + \lambda^T \nabla h(\vec{u}) = 0$, where $\lambda^T = [\lambda_1\ \lambda_2\ \cdots]$, must be solved along with the constraints $h_1(\vec{u}) = 0; h_2(\vec{u}) = 0; \cdots$ to obtain the optimum value of f. This method can be used directly to find the shortest distance of a point in the n-dimensional Euclidean space $\mathbb{R}^n$ from the hyperplane.

In order to find the shortest distance, we let $f(\vec{u}) \equiv (u_x - p_x)^2 + (u_y - p_y)^2$ as the objective function and $h(\vec{u}) \equiv u_y - mu_x - c = 0$ as the constraint function. Thus, our problem belongs to the $\mathbb{R}^2$ space. The gradients of these functions are:

$$\begin{aligned} \nabla f(\vec{u}) &= \begin{bmatrix} 2(u_x - p_x) \\ 2(u_y - p_y) \end{bmatrix} \\ \nabla h(\vec{u}) &= \begin{bmatrix} -m \\ 1 \end{bmatrix}. \end{aligned}$$

Then, using the condition $\nabla f(\vec{u}) + \lambda \nabla h(\vec{u}) = 0$, where λ is the Lagrange multiplier, we get the following two equations:

$$\begin{aligned} 2(u_x - p_x) + \lambda(-m) &= 0 \qquad \text{and} \\ 2(u_y - p_y) + \lambda &= 0. \end{aligned}$$

By eliminating the unknown parameter λ and using the constraint $h(\vec{u}) = 0$, we obtain:

$$u_x = \frac{p_x + mp_y - mc}{1 + m^2},$$

which is again exactly the expression as in (2) for b_x. The point (u_x, u_y) can hence, be shown to coincide with point B, and we again obtain the same formula of (3).

4. Epilogue

How different mathematical techniques can be employed to solve the same problem has been demonstrated in this article. This

This workout highlights the usefulness of different methods, with appropriate modeling of the entities involved, to solve trivial looking problems.

workout highlights the usefulness of different methods, with appropriate modeling of the entities involved, to solve trivial looking problems. In addition, it also teaches us how to view a given problem from different perspectives. If we encourage ourselves to try such exercises, it will aid the development of a habit of applying what we learn to solve different problems through independent thinking. The advantage of choosing the seemingly trivial problems is that the solutions are already known to us by simpler means. This gives us a chance to self-assess our efforts.

Suggested Reading

[1] https://en.wikipedia.org/wiki/Distance_from_a_point_to_a_line.

[2] S L Loney, *The Elements of Coordinate Geometry–Cartesian Coordinates*, G K Publisher, 2016.

[3] Edwin K P Chong and Stanislaw H Zak, *An Introduction to Optimization*, Wiley India Private Limited, 2010.