VOL. 88, NO. 1, FEBRUARY 2015 1

The shortest distance problem: an elementary solution

Katiuscia B. Teixeira
University of Central Florida,
Orlando, FL, USA
katiuscia.teixeira@ucf.edu

Introduction and Set-up

That the shortest path between two points in the (Euclidean) plane is a straight line is rather intuitive and even treated as common sense. However, this ancient problem is far from being trivial and the endeavor to understand it rigorously has prompted revolutionary ideas and tools throughout the history of mathematics. Euclid had proven that a straight line is always shorter than two consecutive lines; this is what we nowadays call the triangle inequality. It was Archimedes, however, the first one to tackle the shortest distance problem in the generality we refer nowadays.

The great mathematician Grothendieck [1928–2014] once said: *"...one should never try to prove anything that is not almost obvious"*, see [**3**], and it is hard to find a better fit to such a principle than the shortest distance problem.

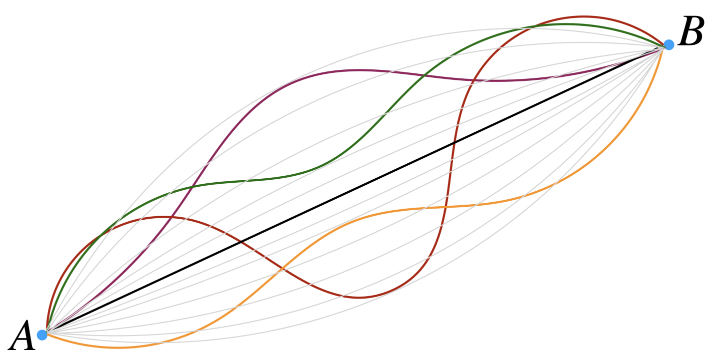

Figure 1 The shortest distance problem. While intuitive, there are infinitely many possibilities to join two points.

The mathematical formulation of the shortest distance problem can be easily done solely based on elements covered in any Calculus I course. However, due to the fact that "the set of candidates", that is the collection of all possible paths joining two points form an infinite dimensional space, traditional solutions of the problem depend on much heavier machineries pertaining to the field of Calculus of Variations.

Some of the classical textbooks on the theme are [**1, 2, 5, 6, 7**]. There are many other references that discuss traditional solutions by means of the Euler-Lagrange Equation associated to the minimization problem in question. As a result, the vast majority of STEM students conclude their degrees without being able to provide a satisfactory justification as to why a straight line is indeed the shortest path between two points.

The goal of this article is to remediate this by offering a self-contained solution to the shortest distance problem. But before, we would like to briefly discuss how to model the shortest distance path problem in the language of Calculus.

Set-up With convenient scaling and no loss of generality (see last section), let us image we are seeking for the shortest path between $A = (0,0)$ and $B = (1,b)$, for some $b \geq 0$. Also, to the benefit of the presentation, we will restrict to graph paths, i.e. we are interested in paths of the form: $(t, f(t))$, for some differentiable function $f\colon [0,1] \to \mathbb{R}$, verifying $f(0) = 0$ and $f(1) = b$. It might be convenient to give a name for the collection of all such functions:

$$\mathcal{P} := \left\{ f\colon [0,1] \to \mathbb{R} \;\middle|\; f(t) \text{ is differentiable}, \, f(0) = 0, \text{ and } f(1) = b \right\}. \tag{1}$$

The arc length of the path $(t, f(t))$ is given by the formula:

$$\ell(f) := \int_0^1 \sqrt{1 + [f'(t)]^2}\, dt. \tag{2}$$

The shortest distance path problem can now be properly formulated as:

$$\min \left\{ \ell(f) \;\middle|\; f \in \mathcal{P} \right\}, \tag{3}$$

and Archimedean's motto "the shortest path between two points is a straight line" framed as a precise mathematical theorem:

Theorem 1. *The only minimizer to Problem* (3) *is the function* $f(t) = bt$.

An elementary proof of Theorem 1

We start off by investigating the real function,

$$F(x) := \sqrt{1 + x^2},$$

aiming at estimating it from below. Direct application of the Chain Rule yields:

$$F'(x) = \frac{x}{\sqrt{1+x^2}} \quad \text{and} \quad F''(x) = \frac{1}{(1+x^2)^{3/2}}.$$

In particular, since $F''(x) > 0$, F is a strictly convex function. Let $T_b(x)$ denote the tangent line to $F(x)$ at $x = b$. Easily we compute

$$\begin{aligned} T_b(x) &= F(b) + F'(b) \cdot (x - b) \\ &= \sqrt{1+b^2} + \frac{b}{\sqrt{1+b^2}} (x - b). \end{aligned} \tag{4}$$

Being a convex function, F is above its tangent lines, so in particular $F(x) \geq T_b(x)$, for all $x \in \mathbb{R}$. In fact, this inequality is strict, unless $x = b$. To put it differently, the function

$$e(x) := F(x) - T_b(x) \tag{5}$$

is convex, non-negative, and $e(x) > 0$ for all $x \neq b$.

We are ready to prove the shortest path between $(0,0)$ and $(1,b)$ is $f(t) = bt$. For that, let $f\colon [0,1] \to \mathbb{R}$ be any element of $\mathcal{P}$. In view of (5), $F(x) = e(x) + T_b(x)$, for all x, which implies:

$$F(f'(t)) = e(f'(t)) + T_b(f'(t)), \tag{6}$$

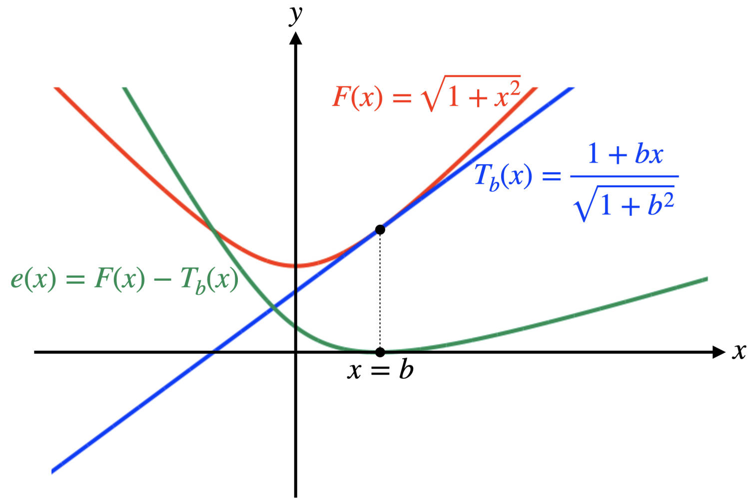

Figure 2 Graphs of the three functions involved in the proof.

for all $t \in (0,1)$. Hence,

$$\begin{aligned} \ell(f) &:= \int_0^1 \sqrt{1+[f'(t)]^2}dt = \int_0^1 F(f'(t))dt \\ &= \int_0^1 \{e(f'(t)) + T_b(f'(t))\}\, dt \qquad (7) \\ &= \int_0^1 e(f'(t))dt + \int_0^1 T_b(f'(t))dt. \end{aligned}$$

The first term, $\int_0^1 e(f'(t))dt$, is greater than or equal to zero, and in fact, it is strictly positive, unless $f'(t) \equiv b$ for all $t \in (0,1)$. As for the second term, in view of (4), it can be computed as,

$$\begin{aligned} \int_0^1 T_b(f'(t))dt &= \int_0^1 \{F(b) + F'(b)\,(f'(t)-b)\}\, dt \\ &= \sqrt{1+b^2} + \frac{b}{\sqrt{1+b^2}} \int_0^1 (f'(t)-b)\, dt. \end{aligned} \qquad (8)$$

We recognize the first term, $\sqrt{1+b^2}$, as our target minimizing quantity. As for the second term, using the Fundamental Theorem of Calculus, we compute:

$$\frac{b}{\sqrt{1+b^2}} \int_0^1 (f'(t)-b)\, dt = \frac{b}{\sqrt{1+b^2}}\,(f(1)-f(0)-b) = 0. \qquad (9)$$

Combining (6), (7), (8), and (9) we finally obtain:

$$\ell(f) = \sqrt{1+b^2} + \int_0^1 e(f'(t))dt \geq \sqrt{1+b^2}, \qquad (10)$$

and the above inequality is strict unless $f'(t) \equiv b$. Theorem 1 is confirmed. □

Final remarks

Bonus information. Initially we comment that one can actually use the expression $\int_0^1 e(f'(t))dt$ as to quantify the "length excess" in terms of how much $f'(t)$ deviates from being b. For $\delta > 0$, consider the sets

$$A_\delta := \left\{t \in (0,1) \mid f'(t) < b - \delta\right\} \quad \text{and} \quad B_\delta := \left\{t \in (0,1) \mid f'(t) \geq b - \delta\right\},$$

and assume both have positive measures, $|A_\delta| > 0$ and $|B_\delta| = 1 - |A_\delta| > 0$. We break the excess term as:

$$E := \int_0^1 e(f'(t))dt = \int_{A_\delta} e(f'(t))dt + \int_{B_\delta} e(f'(t))dt, \tag{11}$$

where according to (5), $e(x) = \sqrt{1+x^2} - \dfrac{1+bx}{\sqrt{1+b^2}}$; a non-negative, strictly convex function vanishing only at $x = b$. We readily estimate:

$$\int_{A_\delta} e(f'(t))dt \geq |A_\delta| e(b-\delta). \tag{12}$$

Now, since $\int_0^1 f'(t)dt = b$ we have:

$$\int_{B_\delta} f'(t)dt = b - \int_{A_\delta} f'(t)dt \geq (1 - |A_\delta|)\, b + \delta |A_\delta| = |B_\delta| b + \delta |A_\delta|. \tag{13}$$

Applying Jensen's inequality, see for instance [**4**] for an elegant proof, we estimate

$$e\left(\frac{1}{|B_\delta|}\int_{B_\delta} f'(t)dt\right) \leq \frac{1}{|B_\delta|}\int_{B_\delta} e(f'(t))dt, \tag{14}$$

which combined with (13) yields:

$$\int_{B_\delta} e(f'(t))dt \geq |B_\delta| e\left(b + \delta\frac{|A_\delta|}{|B_\delta|}\right).$$

Finally, we can estimate the length excess by

$$E \geq |A_\delta| e(b-\delta) + |B_\delta| e\left(b + \delta\frac{|A_\delta|}{|B_\delta|}\right) > 0. \tag{15}$$

Arbitrary intervals. Now let's us briefly comment on the fact that the problem has been modeled over the unit interval $[0,1]$ is not restrictive. Indeed, if a differentiable function f is defined over an generic interval $[c,d]$ and take arbitrary values $f(c)$ and $f(d)$, we simply apply the result proven to the new function, $\tilde{f}\colon [0,1] \to \mathbb{R}$ given as

$$\tilde{f}(t) = \frac{f(c + t(d-c)) - f(c)}{d-c}.$$

To see that, one can easily check that $\tilde{f}(0) = 0$ and $\tilde{f}(1) = \frac{f(d)-f(c)}{d-c} =: b$. Thus,

$$\sqrt{1 + \left(\frac{f(d)-f(c)}{d-c}\right)^2} \leq \ell(\tilde{f}). \tag{16}$$

On the other hand, making a change of variable, $y = c + t(d - c)$, we compute:

$$\begin{aligned} \ell(\tilde{f}) &= \int_0^1 \sqrt{1 + \left(f'(c + t(d-c))\right)^2} dt \\ &= \frac{1}{d-c} \int_c^d \sqrt{1 + \left(f'(y)\right)^2} dy. \end{aligned} \tag{17}$$

Comparing (16) and (17) we finally reach:

$$\ell(f) \geq \sqrt{(d-c)^2 + (f(d) - f(c))^2},$$

with equality if, and only if,

$$f(t) = f(c) + \frac{f(d) - f(c)}{d - c}(t - c).$$

Summary We offer an elementary solution to the shortest distance problem. In addition to being accessible to freshman Calculus students, our approach also provides a bit of extra information, namely a way quantify the “length excess” in terms of how much the path deviates from being a straight line.

Notes on contributor

Katiuscia Teixeira (MR Author ID: 1490277) is an Assistant Professor in the Department of Mathematics at the University of Central Florida, FL USA. Her research focuses on Mathematics Education. She is also interested in new insightful proofs of important results from undergraduate and graduate mathematical courses.

REFERENCES

1. Courant, R., and Hilbert, D., *Methods of Mathematical Physics. Vol I*, Interscience Publishers, Inc., New York, N.Y., 1953. xv+561 pp.
2. Dacorogna, B., *Introduction to the Calculus of Variations*, Imperial College Press, London, 2015. x+311 pp. ISBN: 978-1-78326-551-0
3. Jackson, Allyn *Comme appelé du néant—as if summoned from the void: the life of Alexandre Grothendieck. II.* Notices Amer. Math. Soc. 51 (2004), no. 10, 1196–1212.
4. Sun, Dennis L. *Proof without words: convex hulls and Jensen’s inequality.* College Math. J. 52 (2021), no. 4, 298. 26D15
5. Kot, M., *A First Course in the Calculus of Variations*, Student Mathematical Library, 72. American Mathematical Society, Providence, RI, 2014. x+298 pp. ISBN: 978-1-4704-1495-5
6. Giaquinta, M., and Hildebrandt, S., *Calculus of Variations I. The Lagrangian Formalism,* Springer-Verlag, Berlin, 1996. xxx+474 pp. ISBN: 3-540-50625-X
7. Giusti, Enrico *Direct methods in the calculus of variations.* World Scientific Publishing Co., Inc., River Edge, NJ, 2003. viii+403 pp. ISBN: 981-238-043-4