

On the Maximum Arc Length of Monotonic Functions

Marc Deléglise and Andrew Markoe

May 29, 2013

Abstract

We revisit a problem solved in 1963 by Zaanen & Luxemburg in this monthly: what is the largest possible length of the graph of a monotonic function on an interval? And is there such a function that attains this length?

This is an interesting and intriguing problem with a somewhat surprising answer, that should be of interest to a broad spectrum of mathematicians starting with upper level undergraduates.

The proof given by Zaanen & Luxemburg is very short and elegant but not accessible to an undergraduate. We give here a longer, but elementary, proof.

MSC Primary 26A30, 26A48, 28A10

1 Introduction.

This article concerns a simply stated but fascinating problem in analysis. The problem concerns *monotonic functions*, which are functions that are either *non-increasing* or *nondecreasing*.

Problem (Arc Length Problem).

Let $[a, b]$ be a closed interval. What is the maximum arc length that can be attained by a monotonic function $f : [a, b] \rightarrow [f(a), f(b)]$? Is there a function that attains this maximum?

This problem was solved by A. C. Zaanen and W. A. J. Luxemburg [3], [4] in 1963. The problem was also mentioned by P. R. Halmos in an interesting article about teaching mathematics [1, p. 467].

Although Zaanen and Luxemburg stated their theorem only in the continuous case, we believe that their proof is valid, without modification, in the case of a discontinuous monotonic function. The proof given here applies to this general case.

2 Notation and Prerequisites.

The minimum prerequisite for reading this paper is to know the definition of *Lebesgue measure* on the real line, which we denote by ℓ , and the definition and elementary properties of *sets of measure zero*, which we also call *null sets*. Recall that a statement is said to be *true almost everywhere* if it is true on the complement of a set of measure zero. The elementary properties of null sets that we require are:

1. The union of countably many sets of measure zero is of measure zero.
2. A subset of a set of measure zero is of measure zero.

The *length of a measurable subset* of the line is its Lebesgue measure, and in particular the length of an interval $I = [a, b]$ is $\ell(I) = b - a$.

Definition 2.1. A function $f : [a, b] \rightarrow \mathbb{R}$ is said to be *nondecreasing* if $x < y \implies f(x) \leq f(y)$ for all $x, y \in [a, b]$. Reversing the second inequality defines a *nonincreasing function*.

We say that a function is *monotonic* if it is either nondecreasing or nonincreasing.

If the inequality $f(x) \leq f(y)$ is strict, $f(x) < f(y)$ for all $x, y \in [a, b]$, then we say that f is *strictly increasing*

Definition 2.2. A *partition* of an interval $[a, b]$ of \mathbb{R} is a finite increasing sequence $(x_0, x_1, x_2, \dots, x_n)$ where n is an arbitrary positive integer and $x_0 = a$ and $x_n = b$.

A *subinterval* of a partition is of the form $[x_{j-1}, x_j]$.

Definition 2.3. Let f be a function defined on the interval $[a, b]$. If $X = (x_0, x_1, x_2, \dots, x_n)$ is a partition of $[a, b]$, then we let

$$\lambda_f(X) = \sum_{j=1}^n \sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2}.$$

The *arc length of the function* $f : [a, b] \rightarrow \mathbb{R}$ is defined as

$$\Lambda_f(a, b) = \sup \lambda_f(X)$$

where the supremum is taken over all the partitions X of $[a, b]$.

If $I = [a, b]$ we also write $\Lambda_f(I)$ for $\Lambda_f(a, b)$. If there is no confusion we drop the subscript f from the notation.

Definition 2.4. A function $f : [a, b] \rightarrow \mathbb{R}$ is said to be of *bounded variation* if it is the difference of two nondecreasing functions.

It can be shown that $\Lambda_f(a, b)$ is finite, if and only if f is of bounded variation. In particular $\Lambda_f(a, b)$ is finite for each nondecreasing function f . Proposition 3.1 gives an effective prove of this result.

It can be shown that arc length is additive: If $c \in [a, b]$, then

$$\Lambda_f(a, b) = \Lambda_f(a, c) + \Lambda_f(c, b). \quad (2.1)$$

We use Lebesgue's theorem on the derivative of a monotonic function, Fubini's theorem on term by term differentiation of an infinite series, and the elementary theory of interval functions. Although we state these results, theorems 1, 2, and 3, without proof, the interested reader may find a short and beautiful treatment of these topics in only 19 pages in the book of F. Riesz and B. Sz.-Nagy [2, pages 3 - 21].

Theorem 1 (Lebesgue). *A monotonic function has a finite derivative almost everywhere.*

Theorem 2 (Fubini). *Let f_n be a sequence of nondecreasing functions such that the series*

$$s(x) = \sum_{n=1}^{\infty} f_n(x)$$

is pointwise convergent on the interval $[a, b]$. Then has a finite derivative almost everywhere. Moreover the series may be differentiated term by term in the sense that the equation

$$s'(x) = \sum_{n=1}^{\infty} f'_n(x)$$

is true almost everywhere.

The theory of interval functions is a useful tool that leads easily to important results in other areas such as Riemann integration, total variation, and arc length. This theory was in vogue in the early twentieth century, but it has lately fallen into obscurity.

Definition 2.5 (Interval Functions). *An interval function is a real valued function defined on the family of closed subintervals of a fixed closed interval $[a, b]$.*

A simple example is $f[c, d] = d - c$ for every closed interval $[c, d]$ contained in $[a, b]$.

By convention, if $a \geq b$, then we let $f[a, b] = -f[b, a]$.

Definition 2.6 (Integral of an Interval Function). *A real number I is the integral of an interval function f on $[a, b]$, denoted by $\int f[a, b]$, if for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every partition x_0, x_1, \dots, x_n of $[a, b]$ with maximum subinterval length $< \delta$ it is true that*

$$\left| \sum_{j=1}^n f[x_{j-1}, x_j] - I \right| < \varepsilon. \quad (2.2)$$

Example 2.1 (An interval function with all nonnegative terms and with integral equal to zero). Let f be a function of bounded variation on $[a, b]$. Define $g [c, d] = \Lambda_f (c, d) - \sqrt{(d - c)^2 + (f(d) - f(c))^2}$. The integral of g is zero, because

$$\begin{aligned} & \left| \sum_{j=1}^n g [x_{j-1}, x_j] - 0 \right| \\ &= \left| \sum_{j=1}^n \Lambda_f (x_{j-1}, x_j) - \sum_{j=1}^n \sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2} \right|. \end{aligned}$$

By additivity of arc length (2.1), $\sum_{j=1}^n \Lambda_f (x_{j-1}, x_j) = \Lambda_f (a, b)$ and thus the previous expression is

$$\left| \Lambda_f (a, b) - \sum_{j=1}^n \sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2} \right|.$$

By the definition of arc length, for every $\varepsilon > 0$ we can find $\delta > 0$ such that for every partition of $[a, b]$ with maximum interval length $< \delta$,

$$\left| \Lambda_f (a, b) - \sum_{j=1}^n \sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2} \right| < \varepsilon.$$

This shows that $\int g [a, b] = 0$.

Definition 2.7 (Derivative of an Interval Function). We say that the interval function f is *differentiable* at c with derivative $f'(c)$ if the limit

$$f'(c) = \lim_{x \rightarrow c} \frac{f [c, x]}{x - c}$$

exists.

Theorem 3 (First Fundamental Interval Function Theorem). *Let f be an interval function on $[a, b]$ that satisfies the two conditions:*

1. f has nonnegative values, and
2. $\int f [a, b] = 0$.

Then, the derivative of f is zero almost everywhere.

3 Characterization of Monotonic Functions with Maximum arc length.

In this section we prove Theorem 4, stated and proved by Zaannen & Luxemburg's in [3] and [4].

Theorem 4. [Zaanen & Luxemburg] Let f be a continuous, monotonic function defined on the closed interval $[a, b]$.

Then f has the maximum arc length $(b - a) + |f(b) - f(a)|$ if and only if it has a zero derivative almost everywhere.

Possibly replacing f by $-f$, it is sufficient to prove this theorem in the case of nondecreasing functions. Our proof is divided in a proof of sufficiency, Theorem 5, and a proof of necessity, Theorem 6.

Proposition 3.1. Let $f : [a, b] \rightarrow \mathbb{R}$ be a nondecreasing function. Then $\Lambda_f(a, b)$ is bounded above by $b - a + f(b) - f(a)$ and is bounded below by both $b - a$ and $f(b) - f(a)$.

Proof. Let $x_0, x_1, x_2, \dots, x_n$ be a partition of $[a, b]$. By the triangle inequality and the hypothesis that f is nondecreasing, for each j , $1 \leq j \leq n$,

$$x_j - x_{j-1} \leq \sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2} \leq x_j - x_{j-1} + f(x_j) - f(x_{j-1}).$$

Adding these inequalities we get

$$b - a \leq \sum_{j=1}^n \sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2} \leq b - a + f(b) - f(a). \quad (3.1)$$

Since $\Lambda_f(a, b)$ is the least upper bound of

$$\sum_{j=1}^n \sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2}$$

taken on all partitions of $[a, b]$, by inequality (3.1) we obtain

$$b - a \leq \Lambda_f(a, b) \leq b - a + f(b) - f(a).$$

In the same way, starting from

$$f(x_j) - f(x_{j-1}) \leq \sqrt{(x_j - x_{j-1})^2 + (f(x_j) - f(x_{j-1}))^2}$$

we get $f(b) - f(a) \leq \Lambda_f(a, b)$. □

Proposition 3.2. Let f be a nondecreasing, continuous function defined on an interval $[a, b]$ of \mathbb{R} . If U is a set on which f has a zero derivative, then $f(U)$ is a null set.

Proof. In the case where f is strictly increasing, the proof of theorem 3.2 is particularly simple. We define $g = f^{-1}$, the reciprocal function of f . If $y = f(x)$ with $x \in U$, from $f'(x) = 0$ we deduce $g'(y) = \infty$. Thus, for each $y \in f(U)$, g does not admit a finite derivative at y , and $f(U)$ is a subset of the set of points where g has no derivative. By Lebesgue's theorem, this is a null set, and this ends the proof in this case.

In the case where f is not strictly increasing, we have to work a little more. Because f is nondecreasing and continuous, $f^{-1}(y)$ is a closed interval for each $y \in f([a, b])$, and we define a function $g : [f(a), f(b)] \mapsto [a, b]$ by

$$g(y) = \text{the right end point of the interval } f^{-1}(y).$$

We begin by proving the following properties of g .

Property 1 : $\mathbf{f(g(y)) = y}$ for each \mathbf{y} in $\mathbf{[f(a), f(b)]}$, by definition of g .

Property 2 : \mathbf{g} is strictly increasing.

Let $y_1 < y_2$. Suppose, by contradiction, that $g(y_1) \geq g(y_2)$. Then, since f is nondecreasing we may write $y_1 = f(g(y_1)) \geq f(g(y_2)) = y_2$, which contradicts $y_1 < y_2$.

Property 3 : If \mathbf{g} is continuous at \mathbf{y} , and $\mathbf{y} > \mathbf{f(a)}$, then $\mathbf{f^{-1}(y) = \{g(y)\}}$.

By property 1, $g(y) \in f^{-1}(y)$. By contradiction, let us suppose that $f^{-1}(y)$ is not reduced to $g(y)$. Then, by definition of $g(y)$, there is some $c \in [a, b]$ such that $f^{-1}(y) = [c, g(y)]$, with $c < g(y)$. Since $f(a) < y$ we may choose an increasing sequence y_n of points in $[f(a), y)$ such that $y_n \rightarrow y$. Then $g(y_n) \rightarrow g(y)$, by continuity of g at y , and, for n sufficiently large, $g(y_n) \in [c, g(y)] = f^{-1}(y)$. Thus $y_n = f(g(y_n)) = y$, which is a contradiction.

Property 4 : If $\mathbf{y} \in \mathbf{f(U) \setminus f(a)}$ then \mathbf{g} is not differentiable at \mathbf{y} .

By contradiction, let us suppose g is differentiable at y . Then f is also differentiable at $g(y)$. This is because g is also continuous at y , and by property 3, $\mathbf{f^{-1}(y)}$ consists of a single point $x = g(y)$. But $y \in f(U)$ and thus $g(y)$ must be in U . By the hypothesis on U it follows that f must be differentiable at $g(y)$. With the differentiability of g at y and f at $g(y)$, we are justified in using the chain rule:

$$1 = (f \circ g)'(y) = f'(g(y))g'(y). \quad (3.2)$$

The left-hand side equals 1 because by property 1, $f \circ g$ is the identity function on $[f(a), f(b)]$.

As we have already observed, $g(y) = x \in U$ and hence $f'(g(y)) = f'(x) = 0$ by the hypothesis on U . This together with (3.2), gives the contradiction $0 = 1$.

The proof of theorem 3.2 is completed as follows: since g is nondecreasing, by Lebesgue's theorem, the set of points where g is not differentiable is a null set. By property 4, $f(U) \setminus f(a)$ is a subset of this set, thus, it is a null set. The singleton $f(a)$ being also a null set, $f(U)$ is a null set. \square

Theorem 5. *If f is a nondecreasing, continuous, function defined on the closed interval $[a, b]$ that has a zero derivative almost everywhere, then $\Lambda_f(a, b)$ is equal to $(b - a) + (f(b) - f(a))$.*

Proof. By proposition 3.1, $\Lambda_f(a, b) \leq b - a + f(b) - f(a)$. It remains to prove the converse inequality. Before doing this we recall from theorem 3.2 that $F_y = f^{-1}(y)$ is a closed interval. The lengths of these intervals form an uncountable positive family whose sum is convergent because the partial sums are bounded by $b - a$. Therefore, at most countably many of the intervals F_y have positive length.

Choose any $\varepsilon > 0$. By the hypothesis that $f' = 0$ almost everywhere, there is a sequence (I_n) of disjoint subintervals of $[a, b]$ such that f has a zero derivative on $U = [a, b] \setminus \bigcup_{n=1}^{\infty} I_n$, and

$$\sum_{n=1}^{\infty} \ell(I_n) \leq \varepsilon. \quad (3.3)$$

We show that if $n \neq m$, then $f(I_n) \cap f(I_m)$ and $f(U) \cap f(I_n)$ have measure zero. Suppose that $y \in f(I_n) \cap f(I_m)$. Then $\exists s \in I_n, t \in I_m$ such that $y = f(s) = f(t)$. Also, the distance between s and t is positive because I_n and I_m are disjoint open intervals. This means that $y \in f(F_y)$ where the interval F_y has positive length. Because there are only countably many such intervals F_y , then there are only countably many y in $f(I_n) \cap f(I_m)$. Thus the intersection has measure zero. The set $f(U) \cap f(I_n)$ has measure zero because it is a subset of $f(U)$, which has measure zero by proposition 3.2.

Now,

$$[f(a), f(b)] = f([a, b]) = f(U) \cup \bigcup_{n=1}^{\infty} f(I_n).$$

Therefore the length of $[f(a), f(b)]$ is the linear Lebesgue measure of the set $f(U) \cup f(I_n)$. If this union were disjoint, then we could use the additivity of Lebesgue measure to find the length. However, in the weakly increasing case the union is not disjoint. But fortunately, there are only countably many pairwise intersections of the terms of the union, and we have shown that the Lebesgue measure of each intersection is zero. Therefore, the measure of the union will indeed be the sum of the measures of the terms.

Therefore,

$$f(b) - f(a) = \ell(f([a, b])) = \ell(f(U)) + \sum_{n=1}^{\infty} \ell(f(I_n)).$$

By theorem 3.2, $\ell(f(U)) = 0$ and hence $f(b) - f(a) = \sum_{n=1}^{\infty} \ell(f(I_n))$. Because of the convergence of this series, there is some integer p such that

$$\sum_{n=1}^p \ell(f(I_n)) \geq f(b) - f(a) - \varepsilon. \quad (3.4)$$

After reindexing, if necessary, we can assume that the p intervals $I_n = (a_n, b_n)$ in this sum are in increasing order. We then define $V_0 = [a, a_1], V_1 = [b_1, a_2], V_2 = [b_2, a_3], \dots, V_{p-1} = [b_{p-1}, a_p], V_p = [b_p, b]$. We see that

$$[a, b] = \bigcup_{n=1}^p I_n \cup V_j \quad (3.5)$$

and this union is disjoint. Thus

$$\sum_{j=0}^p \ell(V_j) = b - a - \sum_{n=1}^p \ell(I_n) \geq b - a - \varepsilon, \quad (3.6)$$

since, from (3.3), $\sum_{n=1}^p \ell(I_n) \leq \varepsilon$. From the disjoint union (3.5), using the additivity of arc length on disjoint intervals, theorem 3.1, inequality (3.4) and inequality (3.6) we get

$$\begin{aligned}\Lambda_f(a, b) &= \sum_{n=1}^p \Lambda_f(I_n) + \sum_{j=0}^p \Lambda_f(V_j) \\ &\geq \sum_{n=1}^p \ell(f(I_n)) + \sum_{j=0}^p \ell(V_j) \geq f(b) - f(a) - \varepsilon + b - a - \varepsilon.\end{aligned}$$

Since ε is an arbitrary positive number this proves that $\Lambda_f(a, b) \geq f(b) - f(a) + b - a$ and ends the proof of theorem 5. \square

Remark 3.1. The hypothesis of continuity was used in two places. Once to apply Proposition 3.2, and once to guarantee that $f([a, b])$ is an interval so that $[f(a), f(b)] = f([a, b])$.

Lemma 3.1. *Let f be nondecreasing and suppose that the arc length of f is equal to $(b - a) + (f(b) - f(a))$ on an interval $[a, b]$. Let $c < x$ in $[a, b]$. Then the length of f restricted to the subinterval $[c, x]$ is equal to $(x - c) + (f(x) - f(c))$.*

Proof. Case 1: $c = a$. Suppose for contradiction that there were a point x in $[a, b]$ such that $\Lambda(a, x)$ were not equal to $(x - a) + (f(x) - f(a))$. By theorem 3.1, $\Lambda(a, x)$ must be less than $(x - a) + (f(x) - f(a))$. Also by this theorem $\Lambda(x, b) \leq (b - x) + (f(b) - f(x))$. Because arc length is additive, then the total arc length would satisfy

$$\begin{aligned}\Lambda(a, b) &= \underbrace{(b - a) + (f(b) - f(a))}_{\text{by hypothesis}} \\ &= \Lambda(a, x) + \Lambda(x, b) \\ &< (x - a) + (f(x) - f(a)) + (b - x) + (f(b) - f(x)) \\ &= (b - a) + (f(b) - f(a)) = \Lambda(a, b).\end{aligned}$$

This gives the contradiction $\Lambda(a, b) < \Lambda(a, b)$. Thus $\Lambda(a, x) = (x - a) + (f(x) - f(a))$.

Case 2: $a < c \leq b$. By the preceding equation with $x = c$, we obtain $\Lambda(a, c) = (c - a) + (f(c) - f(a))$. Thus

$$\begin{aligned}\Lambda(c, x) &= \Lambda(a, x) - \Lambda(a, c) \\ &= ((x - a) + (f(x) - f(a))) - ((c - a) + (f(c) - f(a))) \\ &= (x - c) + (f(x) - f(c)).\end{aligned}$$

\square

The next result shows how to calculate the limit of the ratio of arc length to abscissa length. This result may not be well known, although it is obvious if you draw a picture. The proof requires the theory of interval functions, their integrals, and derivatives as described in section 2.

Lemma 3.2. *Let f be a function of bounded variation on an interval I . Then for almost every $c \in I$,*

$$\lim_{x \rightarrow c} \frac{\Lambda_f(c, x)}{|x - c|} = \sqrt{1 + (f'(c))^2}.$$

Proof. Consider the interval function $g(a, b) = \Lambda_f(a, b) - \sqrt{(b - a)^2 + (f(b) - f(a))^2}$. Because we defined $g(c, x) = -g(x, c)$ it is easy to show that

$$\frac{g(c, x)}{x - c} = \frac{\Lambda_f(c, x) - \sqrt{(x - c)^2 + (f(x) - f(c))^2}}{|x - c|} \quad (3.7)$$

for $x \neq c$.

In example 2.1 we showed that $g(a, b)$ has all nonnegative terms and its integral is equal to zero. By the first fundamental theorem of interval functions, the derivative of g exists almost everywhere with the value zero. By this and equation (3.7), for almost every c in I , we have

$$0 = \lim_{x \rightarrow c} \frac{g(c, x)}{x - c} = \lim_{x \rightarrow c} \frac{\Lambda_f(c, x) - \sqrt{(x - c)^2 + (f(x) - f(c))^2}}{|x - c|}. \quad (3.8)$$

Let c be a point where f is differentiable. Then

$$\begin{aligned} \lim_{x \rightarrow c} \frac{\sqrt{(x - c)^2 + (f(x) - f(c))^2}}{|x - c|} &= \sqrt{1 + \left(\lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c} \right)^2} \\ &= \sqrt{1 + (f'(c))^2}. \end{aligned}$$

Therefore if f is differentiable at c , and if equation (3.8) holds, then $\lim_{x \rightarrow c} \frac{\Lambda_f(c, x)}{|x - c|} = \sqrt{1 + (f'(c))^2}$.

Because a function of bounded variation is the difference of monotonic functions, then the Lebesgue differentiation theorem implies that f is differentiable almost everywhere. Because the union of two measure zero sets has measure zero, then at almost every c in I , it is true that both (3.8) is valid and $f'(c)$ exists. This proves that for almost every c

$$\lim_{x \rightarrow c} \frac{\Lambda_f(c, x)}{|x - c|} = \sqrt{1 + (f'(c))^2}.$$

□

Theorem 6. *If a nondecreasing function f defined on $[a, b]$ has graph length $(b - a) + (f(b) - f(a))$, then f must have a zero derivative almost everywhere.*

Proof. By Lemma 3.2, $\lim_{x \rightarrow c} \frac{\Lambda(c, x)}{|x - c|} = \sqrt{1 + (f'(c))^2}$ for almost every c . By lemma 3.1, $\Lambda(c, x) = (x - c) + (f(x) - f(c))$. Hence, for almost every c ,

$$\begin{aligned} \sqrt{1 + (f'(c))^2} &= \lim_{x \rightarrow c} \frac{\Lambda_f(c, x)}{|x - c|} = \lim_{x \rightarrow c} \frac{(x - c) + (f(x) - f(c))}{|x - c|} \\ &= \pm \left(1 + \lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c} \right). \end{aligned}$$

Because the derivative of a monotonic function exists almost everywhere, then the preceding equation shows that $\sqrt{1 + (f'(c))^2} = \pm(1 + f'(c))$ for almost every c . Squaring this equation immediately shows that $f'(c) = 0$ for almost every c . \square

It is not completely obvious that theorem 4 is not vacuous, ie that there exists continuous nondecreasing functions f defined on $[a, b]$ satisfying $f(b) > f(a)$ and whose derivative is almost everywhere zero. But it is not too hard to construct them using theorem 2. The function l in figure 1 is such a function. It is an example due to Riesz & Sz-Nagy. Thus, in spite of initial appearances, this function has maximal length. We refer the reader to [2, pp. 48, 49] for the precise construction and proof.

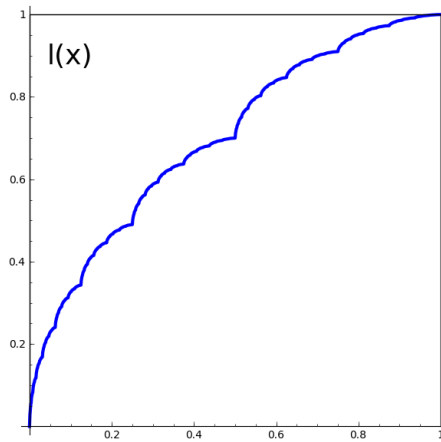


Figure 1: An increasing continuous function applying $[0, 1]$ on itself, with a derivative which is almost everywhere 0. Its arc length has the maximum value $1 + 1 = 2$

4 The discontinuous case

We state in this section the more general form of Zaanen & Luxemburg's Theorem, applying to every monotonic functions. Therefore we are interested in monotonic functions that may be discontinuous. The two curves at left of Figure 2 are the graphs of non decreasing functions g and f mapping 0 on 0, and 1 on 1. The function f is discontinuous at $0, x_1 = 0.2, x_2 = 0.4, x_3 = 0.7$ and 1.

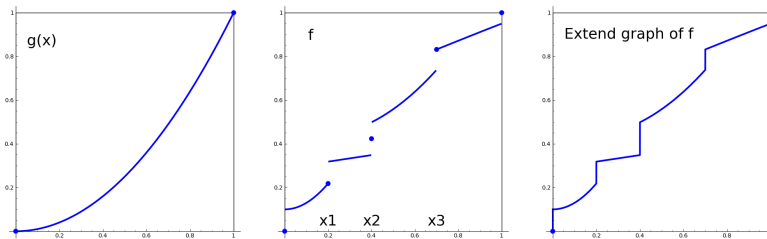


Figure 2: The graphs of g , f and the *extended graph* of f .

The arc length of every monotonic function given by Definition 2.3, is a well defined real number. This could seem unnatural for a discontinuous function like f . But it will appear that this number has a very simple interpretation : it is the length of the continuous curve obtained from the graph of f by adding vertical line segments at the jump discontinuities. We call this curve the *extended graph* of f reading the two words as one, because this curve is no longer the graph of a function. The third curve on Figure 2 is the extended graph of f .

We will use the classical decomposition of every monotonic function as the sum of a step-function, whose discontinuity points are the discontinuity points of f , which is called the *jump function* of f , and a continuous monotonic function, *the continuous part* of f . Figure 3 shows the graphs of f , of its jump-function, and of its continuous part. The following definitions and results up to items 1 and 2 of Proposition 4.1 follow the text of the Riesz & Nagy's book [2].

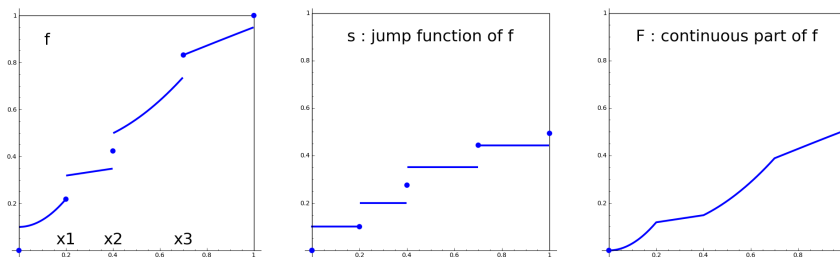


Figure 3: The graphs of f , its jump function s and $F = f - s$.

Definition 4.1. Let f be a nondecreasing function defined on $[a, b]$.

If $x > a$ the *left jump* $J_f^-(x)$ of f at x is $J_f^-(x) = f(x) - \lim_{t \rightarrow x^-} f(t)$

If $x < b$ the *right jump* $J_f^+(x)$ of f at x is $J_f^+(x) = \lim_{t \rightarrow x^+} f(t) - f(x)$
 We define $J_f^-(a) = J_f^+(b) = 0$, and $J_f(x) = J_f^-(x) + J_f^+(x)$ for $x \in [a, b]$.

We write $J^+(x)$, $J^-(x)$, $J(x)$, omitting the f symbol when there is no ambiguity. Then f is continuous at x if and only if $J(x) = 0$, i.e., if and only if $J^-(x) = J^+(x) = 0$.

The discontinuities of the function f of Figure 2 are : A positive right jump at 0, x_1 , x_2 and 1. A positive left jump at x_2 , x_3 and 1.

Lemma 4.1. *Let f be nondecreasing on $[x, y]$, and $x < t < y$. The three numbers*

$$J^+(x), J(t), J^-(y) \text{ are bounded above by } f(y) - f(x). \quad (4.1)$$

Proof. We prove the first statement, the others are similar. Because $y > x$ and f is nondecreasing, then $\lim_{t \rightarrow x^+} f(t) \leq f(y)$. By definition $J^+(x) = \lim_{t \rightarrow x^+} f(t) - f(x)$ and hence $J^+(x) \leq f(y) - f(x)$. \square

Lemma 4.2. *Let f be nondecreasing on $[a, b]$ and $a \leq x < y \leq b$. The sum $\sum_{x < t < y} J(t)$ is convergent, and*

$$J^+(x) + \left(\sum_{x < t < y} J(t) \right) + J^-(y) \leq f(y) - f(x). \quad (4.2)$$

Proof. It is sufficient to prove that for each finite subset $\{t_1 < t_2 < \dots < t_n\}$ of (x, y) , we have

$$J^+(x) + \sum_{i=1}^n J(t_i) + J^-(y) \leq f(y) - f(x). \quad (4.3)$$

For $i = 0, 1, \dots, n$, choose a_i in (x, y) such that $a_{i-1} < t_i < a_i$ for $1 \leq i \leq n$. Inequalities (4.1) give $J^+(x) \leq f(a_0) - f(x)$, $J(t_i) \leq f(a_i) - f(a_{i-1})$ for $1 \leq i \leq n$, and $J^-(y) \leq f(y) - f(a_n)$. Adding all these inequalities we get (4.3). \square

Definition 4.2. Let f be a nondecreasing function defined on $[a, b]$.

1. The *jump-function* of f is the function s defined on $[a, b]$ by

$$s(x) = J_f^+(a) + \sum_{a < t < x} J_f(t) + J_f^-(x) = \sum_{a \leq t < x} J_f(t) + J_f^-(x). \quad (4.4)$$

It is nondecreasing because all the jumps of f are nonnegative.

2. The *continuous part* of f is the function F defined on $[a, b]$ by

$$F(x) = f(x) - s(x). \quad (4.5)$$

Remark 4.1. $s(a) = 0$ and $s(b)$ is the sum of the lengths of the vertical segments of the extended graph of f .

Remark 4.2. The graph of F is the continuous curve obtained by joining the connected components of the graph of f .

Lemma 4.3. *Let f be nondecreasing on $[a, b]$, and s its jump-function.*

1. *There are at most countably many points where f is discontinuous.*
2. *For each $x \in [a, b]$, $J_s^-(x) = J_f^-(x)$ and $J_s^+(x) = J_f^+(x)$. In particular, the functions f and s have the same discontinuity points.*
3. *The function s has almost everywhere a zero derivative.*

Proof. *Item 1.* By (4.2) the sum $\sum_{a < t < b} J_f(t)$ is convergent. Since the $J_f(t)$'s are nonnegative, by theorem 5, there are at most countably many points t such that $J_f(t) > 0$.

Item 2. We prove that $J_s^-(x) = J_f^-(x)$. If $x = a$ there is nothing to prove because, by definition 4.1, in this case, the two numbers are zero. Thus we suppose $x > a$. Let ε be an arbitrary positive number. By convergence of $\sum_{a \leq t \leq b} J_f(t)$, there exists a finite subset $R = \{x_1, x_2, \dots, x_M\}$ of $[a, b]$ such that $\sum_{t \in [a, b] \setminus R} J_f(t) < \varepsilon$. Since R is finite, we can choose $\alpha < x$ such that (α, x) has a void intersection with R . For u in (α, x) , Equation (4.4) defining s gives

$$s(x) - s(u) = J_f^+(u) + \sum_{u < t < x} J_f(t) + J_f^-(x). \quad (4.6)$$

Since u and t are such that $u < t < x$, they do not belong to R , and we have $J_f^+(u) + \sum_{u < t < x} J_f(t) < \varepsilon$ that, with (4.6), gives $s(x) - s(u) \leq J_f^-(x) + \varepsilon$. From (4.6) and the fact that jumps are nonnegative, $J_f^-(x) \leq s(x) - s(u)$. Thus, for u in (α, x) , we have

$$J_f^-(x) \leq s(x) - s(u) \leq J_f^-(x) + \varepsilon.$$

Since ε is arbitrary, this proves that $J_s^-(x) = \lim_{u \rightarrow x^-} s(x) - s(u) = J_f^-(x)$.

Similarly when $x < b$, choosing $\beta > x$ such that (β, x) does not meet T , then for $v \in (\beta, x)$, $J_f^+(x) \leq s(v) - s(x) \leq J_f^+(x) + \varepsilon$, and that gives

$$J_s^+(x) = \lim_{v \rightarrow x^+} s(v) - s(x) = J_f^+(x).$$

Item 3. For each $t \in [a, b]$ define the function χ_t by

$$\chi_t(x) = \begin{cases} 0 & \text{if } x < t \\ J_f^-(t) & \text{if } x = t \\ J_f(t) & \text{if } x > t \end{cases}$$

Then, χ_t is nondecreasing, with a zero derivative at each $x \neq t$, and χ_t is a nonzero function if and only if f is discontinuous at t . Let D be the set of discontinuity points of f . Then (4.4) may be rewritten

$$s(x) = \sum_{t \in D} \chi_t(x),$$

and, by the Fubini theorem (theorem 2), we have almost everywhere $s'(x) = \sum \chi'_t(x) = 0$, because each time $\chi'_t(x)$ exists, its value is zero. \square

Proposition 4.1. *Let f be a nondecreasing function on $[a, b]$, s its jump-function and $F = f - s$ its continuous part.*

1. F is continuous and nondecreasing.
2. F and f have almost everywhere the same derivative.
3. The arc-lengths of f and F satisfy

$$\Lambda_F(a, b) = \Lambda_f(a, b) - s(b). \quad (4.7)$$

Remark 4.3. The fact that the arc length of f is the same as the length of its extended graph results from Equation (4.7), with remarks 4.1 and 4.2

Proof.

Item 1. Let $x \geq a$. By lemma 4.3, item 2, f and s have the same left-jump at x and thus $F = f - s$ is left continuous at x . In the same way, F is right continuous at each $x \leq b$. Thus F is continuous.

If $a \leq x < y \leq b$, by equation (4.4) defining s , equation (4.2) of lemma 4.2 may be written $s(y) - s(x) \leq f(y) - f(x)$, which is equivalent to $F(y) - F(x) \geq 0$, and thus F is nondecreasing.

Item 2. For almost every x , $f'(x)$ and $s'(x)$ exist. Since, by item 3 of lemma 4.3, $s'(x) = 0$ almost everywhere, we get almost everywhere $F'(x) = f'(x) - s'(x) = f'(x) - 0 = f'(x)$.

Item 3. Let $\mathbf{X} = (x_0 = a, x_1, \dots, x_n) = b$ be a partition of $[a, b]$. For $i = 1, 2, \dots, n$, using complex notation and the triangle inequality

$$\begin{aligned} \lambda_F(x_{i-1}, x_i) &= |x_i - x_{i-1} + \imath(F(x_i) - F(x_{i-1}))| \\ &= |x_i - x_{i-1} + \imath(f(x_i) - f(x_{i-1})) - \imath(s(x_i) - s(x_{i-1}))| \\ &\geq |x_i - x_{i-1} + \imath(f(x_i) - f(x_{i-1}))| - |s(x_i) - s(x_{i-1})|. \end{aligned}$$

Since s is nondecreasing, we have $|s(x_i) - s(x_{i-1})| = s(x_i) - s(x_{i-1})$. Adding these inequalities we get $\lambda_F(\mathbf{X}) \geq \lambda_f(\mathbf{X}) - s(b) + s(a) = \lambda_f(\mathbf{X}) - s(b)$, since, $s(a) = 0$. Taking the supremum over all the partitions of $[a, b]$ we get $\Lambda_F(a, b) \geq \Lambda_f(a, b) - s(b)$.

It remains to prove the opposite inequality. Let ε be an arbitrary positive number. By the definition of $\Lambda_f(a, b)$ and the definition of s (see (4.4)), there exists a partition $\mathbf{X} = (x_0, x_1, \dots, x_n)$ of $[a, b]$ and a subset $T = \{t_1, \dots, t_k\}$ of $[a, b]$ such that

$$\lambda_F(\mathbf{X}) \geq \Lambda_F(a, b) - \varepsilon \quad \text{and} \quad \sum_{i=1}^k J_f(t_i) \geq s(b) - \varepsilon. \quad (4.8)$$

The second inequality follows from the convergence of $s(b) = \sum_{t \in [a, b]} J_f(t)$.

We know that if we add a point to \mathbf{X} we do not decrease $\lambda_F(\mathbf{X})$, thus we may suppose $T \subset \{x_0, x_1, \dots, x_n\}$.

For each x_i that is equal to some t in T we do the following : If $x_i < b$ and $J_f^+(x_i) > 0$, we add at the right of $x_i = t$, close to x_i , a new $x = t^r$. If $t = x_i > a$ and if $J_f^-(x_i) > 0$, we add at the left of t , close to t , a t^ℓ . We thus obtain a partition \mathbf{X}' of $[a, b]$, which is a refinement of X . We take care to choose these t^ℓ and t^r sufficiently close to t to ensure that in the new partition \mathbf{X}' , each t^r is the right neighbor of t , and each t^ℓ is the left neighbor of t . Because T is finite, we may also assume that there are no points of T between t and the new points.

Using the continuity of F we may suppose that

$$|t^r - t| \leq \frac{\varepsilon}{k}, \quad |t^\ell - t| \leq \frac{\varepsilon}{k}, \quad |F(t^r) - F(t)| \leq \frac{\varepsilon}{k}, \quad |F(t^\ell) - F(t)| \leq \frac{\varepsilon}{k}. \quad (4.9)$$

We partition the pairs of consecutive points of \mathbf{X}' in three classes : C_r contains the (t, t^r) for $t \in T$ with $J_f^+(t) > 0$; C_ℓ contains the (t^ℓ, t) for $t \in T$ with $J_f^-(t) > 0$, and C_0 all the other pairs of consecutive points of \mathbf{X}' . We then write

$$\lambda_f(\mathbf{X}') = \sum_{(x', x'') \in C_0} \lambda_f(x', x'') + \sum_{(t, t^r) \in C_r} \lambda_f(t, t^r) + \sum_{(t^\ell, t) \in C_\ell} \lambda_f(t^\ell, t). \quad (4.10)$$

Subtracting from (4.10) the same equation with F instead of f , we get

$$\begin{aligned} \lambda_f(\mathbf{X}') - \lambda_F(\mathbf{X}') &= \sum_{(x', x'') \in C_0} \lambda_f(x', x'') - \lambda_F(x', x'') \\ &+ \sum_{(t, t^r) \in C_r} \lambda_f(t, t^r) - \lambda_F(t, t^r) + \sum_{(t^\ell, t) \in C_\ell} \lambda_f(t^\ell, t) - \lambda_F(t^\ell, t). \end{aligned} \quad (4.11)$$

- For (x', x'') in C_0 , using that F and s are nondecreasing we write

$$\begin{aligned} \lambda_f(x', x'') &= |x'' - x' + \iota(F(x'') - F(x') + s(x'') - s(x'))| \\ &\geq |x'' - x' + \iota(F(x'') - F(x'))| = \lambda_F(x', x''). \end{aligned} \quad (4.12)$$

- If $(t, t^r) \in C_r$, (4.9) gives $\lambda_F(t, t^r) = |t^r - t + \iota F(t^r) - F(t)| \leq \frac{2\varepsilon}{k}$ while, using once more that F and s are nondecreasing,

$$\lambda_f(t, t^r) = |t^r - t + \iota F(t^r) - F(t) + s(t^r) - s(t)| \geq s(t^r) - s(t).$$

By (4.1) of lemma 4.1, $s(t^r) - s(t) \geq J_s^+(t)$. But from lemma 4.3, f and s have the same jumps and therefore $s(t^r) - s(t) \geq J_f^+(t)$. Thus we get

$$\lambda_f(t, t^r) - \lambda_F(t, t^r) \geq J_f^+(t) - \frac{2\varepsilon}{k}. \quad (4.13)$$

- If $(t^\ell, t) \in C_\ell$, (4.9), as in the previous case we get

$$\lambda_f(t^\ell, t) - \lambda_F(t^\ell, t) \geq J_f^-(t) - \frac{2\varepsilon}{k}. \quad (4.14)$$

From (4.11), (4.12), (4.13), and (4.14) we deduce

$$\begin{aligned}\lambda_f(\mathbf{X}') &\geq \lambda_F(\mathbf{X}') + \sum_{t \in T, J^+(t) > 0} \left(J^+(t) - \frac{2\varepsilon}{k} \right) + \sum_{t \in T, J^-(t) > 0} \left(J^-(t) - \frac{2\varepsilon}{k} \right) \\ &\geq \lambda_F(\mathbf{X}') + \left(\sum_{t \in T} J(t) \right) - 4\varepsilon.\end{aligned}$$

Since \mathbf{X}' is a refinement of \mathbf{X} we have $\lambda_F(\mathbf{X}') \geq \lambda_F(\mathbf{X})$, and using (4.8), we get

$$\lambda_f(\mathbf{X}') \geq \Lambda_F(a, b) - \varepsilon + s(b) - \varepsilon - 4\varepsilon \geq \Lambda_F(a, b) + s(b) - 6\varepsilon.$$

This ends the proof because ε is arbitrary small. \square

Theorem 7. [*General form of Zaanen & Luxemburg's Theorem*] Let f be a monotonic function defined on the closed interval $[a, b]$.

Then f has the maximum arc length $(b - a) + |f(b) - f(a)|$ if and only if it has a zero derivative almost everywhere.

Proof. Without loss of generality we may suppose that f is non decreasing. Let s the jump function of f and $F = f - s$ its continuous part. From equations (4.7), (4.5) and $s(a) = 0$ we get

$$\Lambda_f(a, b) - [b - a + f(b) - f(a)] = \Lambda_F(a, b) - [b - a + F(b) - F(a)]. \quad (4.15)$$

Now, by point 2 of Proposition 4.1, f has an almost everywhere zero derivative if and only if F has an almost everywhere zero derivative. By Theorem 4, this is equivalent to $\Lambda_F(a, b) - [b - a + F(b) - F(a)] = 0$. With (4.15), this ends the proof. \square

5 Historical Notes.

The arc length problem has been in the folklore for a long time. Riesz and Sz.-Nagy [2] certainly had the tools to solve the problem, but there is no mention of it there. The first published references that the authors were able to find are Zaanen and Luxemburg [3] and [4]. There is also a reference to the problem in Halmos, [1, p. 467].

Zaanen and Luxemburg gave a short proof of Theorem 4 using deep results in measure theory. Our aim was to present the proof using simpler methods. For example, the proof of theorem 5, due to the first author in the continuous monotonic case, only uses Lebesgue's theorem and the definition of Lebesgue measure. The proof of the converse, theorem 6 is due to the second author.

Zaanen and Luxemburg included the hypothesis that f be continuous. However, we believe that the proof in [4] is valid for the discontinuous case, but the authors were either unaware of this or choose for some reason not to state the result in full generality.

6 Appendix : the theorem of Interval Functions

Proof. This is the proof given in [2, pp. 11-12] with a few added details. Let f satisfying the hypotheses of Theorem 3. From (2.2) with $I = 0$, for each n there exists δ_n such that

$$\left| \sum_{j=1}^p f[x_{j-1}, x_j] \right| < 2^{-n} \quad (6.1)$$

for every partition of $[a, b]$ with maximal interval length $< \delta_n$.

Because f has nonnegative values, we can drop the absolute value sign, and (6.1) is also true if x_1, x_2, \dots, x_p is a partition of an arbitrary subinterval $[c, d]$ of $[a, b]$.

Now define $F_n(x)$ by

$$F_n(x) = \sup \sum_{j=1}^p f[x_{j-1}, x_j] \quad (6.2)$$

where $x_0 = a, x_1, \dots, x_p = x$ is an arbitrary partition of $[a, x]$ whose maximal interval length is less than δ_n . We prove the following items:

- For each n and for all $x \in [a, b]$, $F_n(x) \leq 2^{-n}$.
- For $a \leq x < y \leq b$, with $y - x < \delta_n$, we have

$$F_n(x) + f[x, y] \leq F_n(y). \quad (6.3)$$

- F_n is nondecreasing.

1. Item 1 is obvious by the definitions of δ_n and F_n .
2. To prove item 2 we observe that since $|x - y| < \delta_n$, for every partition x_0, x_1, \dots, x_p of $[a, x]$ with maximal interval length $< \delta_n$, the sequence x_0, x_1, \dots, x_p, y is a partition of $[a, y]$ with maximal interval length $< \delta_n$. Therefore, by definition of $F_n(y)$

$$\sum_{j=1}^p f[x_{j-1}, x_j] + f[x, y] \leq F_n(y).$$

Taking the supremum of the first member of this inequality over all the partitions x_0, x_1, \dots, x_p of $[a, x]$ (with $\max |x_j - x_{j-1}| < \delta_n$) we get

$$F_n(x) + f[x, y] \leq F_n(y).$$

3. Item 3 is then an immediate consequence of item 2 and the hypothesis that f is a nonnegative interval function.

Now we prove that, for every $c, x \in [a, b]$,

$$\text{if } |x - c| < \delta_n \text{ then } \left| \frac{f[c, x]}{x - c} \right| \leq \left| \frac{F_n(x) - F_n(c)}{x - c} \right|. \quad (6.4)$$

If $x > c$, (6.3) gives $F_n(c) + f[c, x] \leq F_n(x)$ and, since $f[c, x] \geq 0$, this can be written as $\left| \frac{f[c, x]}{x - c} \right| \leq \left| \frac{F_n(x) - F_n(c)}{x - c} \right|$. If $x < c$, then reversing the roles of x and c in (6.3) gives the same equation.

From item 1, the $F_n(x)$ form a convergent series $s(x) = \sum_{n=1}^{\infty} F_n(x)$ of nondecreasing functions. By Fubini's theorem $s'(x) = \sum_{n=1}^{\infty} F_n'(x)$ for almost every x . Let C be the set of x for which this true. It is sufficiency to prove that $f'(c) = 0$ for $c \in C$.

Let $c \in C$ and $\varepsilon > 0$ be given. Because of the convergence of $\sum_{n=1}^{\infty} F_n'(c)$, there exists an n such that $|F_n'(c)| < \varepsilon/2$. By the definition of derivative, there is a γ such that

$$\text{if } |x - c| < \gamma \text{ then } \left| \frac{F_n(x) - F_n(c)}{x - c} - F_n'(c) \right| < \frac{\varepsilon}{2}. \quad (6.5)$$

Let $x \in [a, b]$ be such that $|x - c| < \min(\gamma, \delta_n)$. Then (6.4) and (6.5) are satisfied and, with $|F_n'(c)| < \varepsilon/2$ we have

$$\left| \frac{f[c, x]}{x - c} \right| \leq \left| \frac{F_n(x) - F_n(c)}{x - c} \right| \leq \left| \frac{F_n(x) - F_n(c)}{x - c} - F_n'(c) \right| + |F_n'(c)| < \varepsilon.$$

Since ε is arbitrary small, this proves that $\lim_{x \rightarrow c} \frac{f[c, x]}{x - c} = 0$. Thus $f'(c) = 0$ for every $c \in C$ and this ends the proof. \square

References

- [1] P. R. Halmos, E. E. Moise, and G. Piranian, *The Problem of Learning to Teach*, Amer. Math. Monthly **82** (1975), no. 5, 466–476. MR 1537724
- [2] F. Riesz and B. Sz.-Nagy, *Functional analysis*, Dover Books on Advanced Mathematics, Dover Publications, New York, 1990, Translated from the second French edition by Leo F. Boron, Reprint of the 1955 original. MR 1068530 (91g:00002)
- [3] A. C. Zaanen and W. A. J. Luxemburg, *Advanced problem 5029*, Amer. Math. Monthly **69** (1962), no. 5, 438–439.
- [4] ———, *Solution to advanced problem 5029*, Amer. Math. Monthly **70** (1963), no. 6, 674–675.

Marc Deléglise Ph.D in mathematics, Université Lyon 1, 1991, maître de conférences in this university, now retired. His research interests are number theory and computational number theory.

Institut Camille Jordan, Mathématiques Université Claude Bernard (Lyon 1), 21 Avenue Claude Bernard, F-69622 Villeurbanne cédex, France.

Andrew Markoe B.S. in mathematics, the City College of New York, 1964, Ph.D. in mathematics, Brown University, 1969, Member, Institute for Advanced Study, 1970-1971. He is a former chairman of the Department of Mathematics, Rider University and he is currently Professor of Mathematics there. His research interests are in several complex variables, real analysis and Radon transforms.

Department of Mathematics, Rider University, Lawrenceville, NJ 08648, USA