

Functions of Bounded Variation: Part 2

Continuity and Bounded Variation

In the last post we saw that continuity is not essential to the property of being a function of bounded variation. However monotonicity is absolutely essential in the sense that every function of bounded variation can be expressed as a sum or difference of monotone functions. But does that mean that continuity is not at all required? Can we have a function which is discontinuous everywhere and still be of bounded variation? The answer is NO! As an example the function $f(x) = 0$ when x is irrational and $f(x) = 1$ when x is rational is not of bounded variation. We can choose a partition to consists of equal number of rational and irrational points lying alternately and then the variation can be seen as a linear function of the number of points of subdivision so that the variation is not bounded.

So everywhere discontinuous functions are ruled out. So how many discontinuities we can have for a function and the function still be of bounded variation? In order to answer this question we need to study the discontinuities of a monotone function.

Discontinuities of a Monotone Function

A monotone function can only have discontinuities of the first kind i.e. jump discontinuities.

More formally:

If f is monotone on $[a, b]$ then for each point $c \in (a, b)$ the left and right hand limits $f(c-)$ and $f(c+)$ exist and for the end points $f(a+)$ and $f(b-)$ exist.

To establish this let's assume that f is increasing in $[a, b]$ and take a point $c \in (a, b)$ and consider the set $A = \{f(x) \mid x \in (c, b]\}$. Clearly the set A is bounded below by $f(c)$ and is obviously non-empty. Hence $l = \inf A$ exists. We shall prove that $f(c+) = l$. Since $l = \inf A$ given any $\epsilon > 0$ there exists a member $f(d) \in A$ such that $0 \leq f(d) - l < \epsilon$. Since f is increasing it follows that for any $x \in (c, d]$ we have $l \leq f(x) \leq f(d)$ so that $0 \leq f(x) - l \leq f(d) - l < \epsilon$. Thus if $\delta = d - c > 0$ then $0 \leq f(x) - l < \epsilon$ whenever $0 < x - c < \delta$. Thus $f(c+) = l$. Similarly we can prove that $f(c-), f(a+), f(b-)$ exist.

Also if f is increasing then it is clear that $f(a) \leq f(a+), f(b-) \leq f(b)$ and $f(c-) \leq f(c) \leq f(c+)$ for any $c \in (a, b)$. The inequalities get reversed in case the function f is decreasing in $[a, b]$.

The difference $|f(c+) - f(c-)|$ is called the jump of f at c . If case of the monotone functions the jump of the function is bounded in a sense precisely described below:

If f is monotone in $[a, b]$ and $P = \{x_0, x_1, x_2, \dots, x_n\}$ is a partition of $[a, b]$ then

$$\sum_{i=1}^{n-1} |f(x_i+) - f(x_i-)| \leq |f(b) - f(a)|$$

Clearly if f is increasing in $[a, b]$ and $c_i \in (x_i, x_{i+1})$ for $i = 1, 2, \dots, n - 1$ then

$f(x_i+) \leq f(c_i)$ and $f(c_{i-1}) \leq f(x_i-)$ so that

$$|f(x_i+) - f(x_i-)| = f(x_i+) - f(x_i-) \leq f(c_i) - f(c_{i-1})$$

so that sum mentioned above is no more than $f(c_{n-1}) - f(c_0)$. Therefore the sum is obviously no more than $f(b) - f(a)$.

The above result helps us to precisely describe the amount of discontinuities a monotone function may have. What we are now going to establish is that:

If f is monotone in $[a, b]$ then the discontinuities of f in $[a, b]$ form a countable set.

Let f be increasing in $[a, b]$. If f is discontinuous at a point then its jump at that point is positive. Hence we can classify the discontinuities of f into sets $D_m, m = 1, 2, 3, \dots$ such that D_m contains points of discontinuities where the jump is greater than $1/m$. Clearly then the set D of discontinuities of f is given by the union of all such D_m . If points x_1, x_2, \dots, x_n are in D_m then by the last result $n/m \leq f(b) - f(a)$ and therefore n must be finite. It follows that each D_m contains only a finite number of points (this number might be zero in some cases). Hence D which is a union of a countable number of finite sets is countable.

Since a function of bounded variation can be expressed as the difference of two increasing functions, it follows that:

The discontinuities of a function of bounded variation in a closed interval form a countable set.

Continuous Functions of Bounded Variation

Let f be a function of bounded variation in $[a, b]$ and for $x \in [a, b]$ let $V_f(a, x)$ be the total variation of f in interval $[a, x]$. We shall see below that the continuity of f implies the continuity of $V_f(a, x)$ and vice-versa. More formally:

If f is of bounded variation in $[a, b]$ then both the functions $f(x)$ and $V_f(a, x)$ have the same discontinuities in $[a, b]$.

Let $x \in (a, b)$ be an interior point of the interval $[a, b]$ and let $a < x < y \leq b$. We will prove that continuity of f at x implies the continuity of V_f at x and vice-versa. The argument can be modified accordingly to handle the end points of the interval $[a, b]$. Let's also write $V(x)$ in place of $V_f(a, x)$ to simplify notation.

Clearly we have $0 \leq |f(y) - f(x)| \leq V_f(x, y) = V(y) - V(x)$. Taking limits when $y \rightarrow x+$ we see that $0 \leq |f(x+) - f(x)| \leq V(x+) - V(x)$. Similarly we can prove that $0 \leq |f(x) - f(x-)| \leq V(x) - V(x-)$. Together these inequalities show that if V is continuous at x then f is also continuous at that point.

Next we assume that the function f is continuous at $c \in (a, b)$. Then for any $\epsilon > 0$ there is a $\delta' > 0$ such that $|x - c| < \delta'$ implies $|f(x) - f(c)| < \epsilon/2$. Let's take a partition $P = \{x_0, x_1, x_2, \dots, x_n\}$ of $[c, b]$ such that

$$V_f(c, b) - \frac{\epsilon}{2} < \sum_{i=1}^n |f(x_i) - f(x_{i-1})| = \sum_{i=1}^n |\Delta f_i|$$

Let $\delta = \min(\delta', x_1 - x_0)$ and consider any point x such that $0 < x - c < \delta$. Let

$$P' = P \cup \{x\} = \{x'_0 = c, x'_1 = x, x'_2, \dots, x'_m = b\}$$

be the partition of $[c, b]$ obtained by adding point x to P . Since P' is finer than P we have

$$V_f(c, b) - \frac{\epsilon}{2} < \sum_{i=1}^n |f(x_i) - f(x_{i-1})| \leq \sum_{i=1}^m |f(x'_i) - f(x'_{i-1})|$$

We also have

$$|f(x'_1) - f(x'_0)| = |f(x) - f(c)| < \frac{\epsilon}{2}$$

and hence

$$V_f(c, b) - \frac{\epsilon}{2} < \frac{\epsilon}{2} + \sum_{i=2}^m |f(x'_i) - f(x'_{i-1})| \leq \frac{\epsilon}{2} + V_f(x, b)$$

Thus $V_f(c, b) - V_f(x, b) < \epsilon$. Now it is easy to observe that

$$\begin{aligned} 0 &\leq V(x) - V(c) = V_f(a, x) - V_f(a, c) \\ &= V_f(c, x) = V_f(c, b) - V_f(x, b) < \epsilon \end{aligned}$$

Thus we observe that $0 < x - c < \delta$ implies $0 \leq V(x) - V(c) < \epsilon$ so that V is continuous from right at c . Similarly we can prove that V is continuous from left at c . This establishes that V is continuous at the points at which f is continuous.

Now that we have established various properties of functions of bounded variation, it is time to focus on one of their applications namely the formalization of an arc-length.

Length of a Curve

Let's restrict ourselves to a curve in a plane (i.e. a 2-dimensional curve) and recall that such a curve is given in parametrized form by the equations $x = f(t), y = g(t)$ where f and g are continuous functions in some closed interval $[a, b]$. If we have $f(a) = f(b), g(a) = g(b)$ then the curve is said to be closed otherwise it is normally called an arc. If the curve does not intersect itself except possibly at the end points a, b then the curve is called simple.

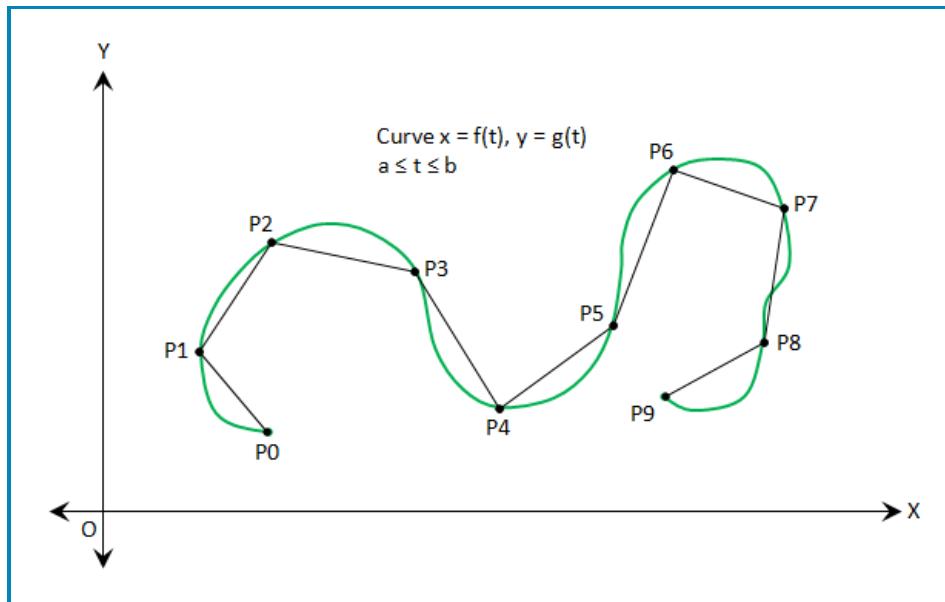
We intuitively associate a number called a length with a curve (remember the high school definition of π as the ratio of circumference to diameter of a circle), but are we certain that every curve must have a length? What is the true conception of length of a curve? Could there be curves for which a reasonable concept of length can not be provided?

To answer these questions let's try to understand how we can measure the length of a curve practically. A simple solution is to divide the curve into many small parts and then approximate the length of each part by the length of the chord joining end-points of this part. The expectation is that as we make finer and finer subdivisions of the curve our approximation

would get better. Since the length of a line segment between the two points is the shortest distance between those two points we should expect that our approximation is from below and that the actual length of curve is greater than any approximation we could make by this method.

We now formalize this approach. Let $x = f(t), y = g(t)$ where f, g are continuous in $[a, b]$ be a given curve. Let $P = \{t_0, t_1, t_2, \dots, t_n\}$ be a partition of the interval $[a, b]$. We mark the points $(f(t_i), g(t_i))$ on the curve as P_i . Let $P_{i-1}P_i$ denote the length of line segment joining P_{i-1} and P_i . And we form the sum

$$L(P) = \sum_{i=1}^n P_{i-1}P_i = \sum_{i=1}^n \sqrt{\{f(t_i) - f(t_{i-1})\}^2 + \{g(t_i) - g(t_{i-1})\}^2}$$



Length of a Curve

As we make partition P finer the value $L(P)$ increases. If the set of values of $L(P)$ for all partitions P of $[a, b]$ is bounded above, then this set has a supremum, say Λ , which is defined as the *length of the curve under consideration* and the curve itself is said to be a *rectifiable curve*. If these values of $L(P)$ are not bounded then the curve is not rectifiable and does not possess a length.

From the definition of $L(P)$ we see that

$$\sum_{i=1}^n |f(t_i) - f(t_{i-1})| \leq L(P) \leq \sum_{i=1}^n |f(t_i) - f(t_{i-1})| + \sum_{i=1}^n |g(t_i) - g(t_{i-1})|$$

$$\sum_{i=1}^n |g(t_i) - g(t_{i-1})| \leq L(P) \leq \sum_{i=1}^n |f(t_i) - f(t_{i-1})| + \sum_{i=1}^n |g(t_i) - g(t_{i-1})|$$

From the above inequalities it is clear that the values $L(P)$ are bounded if and only both the sums seen above for f, g are bounded. These are the sums used to find variation of f, g for a partition P of $[a, b]$. Thus we have the following characterization of rectifiable curves:

A curve given by equations $x = f(t), y = g(t)$ where f, g are continuous in $[a, b]$ is rectifiable if and only if both the functions f and g are of bounded variation in $[a, b]$.

If the curve is defined as $y = f(x)$ where f is continuous in $[a, b]$ then the curve is rectifiable if and only if the function f is of bounded variation in $[a, b]$.

If we denote the length of curve C given by $x = f(t), y = g(t), t \in [a, b]$ by $\Lambda_C(a, b)$ then it is easy to see that $\Lambda_C(a, b) = \Lambda_C(a, c) + \Lambda_C(c, b)$ for $c \in (a, b)$ and this allows us to define $\Lambda_C(a, a) = 0$. Thus the length of a curve is an additive function. Also from the definition it can be easily seen that the function $\Lambda_C(a, x)$ is continuous function of x in $[a, b]$ (the reader should be able to note this continuity follows because of the continuity of the total variation of f and g). Therefore the length of a curve is a continuous function.

P. S.: I owe my own understanding of the concept of functions of bounded variation to the remarkable book "*Mathematical Analysis*" by Tom M. Apostol. Most of the content in this and the previous post is based on material presented in this book. Readers should grab a copy of this book and delve further into these topics.

By Paramanand Singh
Monday, July 16, 2012

Labels: Mathematical Analysis