Functions with zero derivative in an interval are constant

In this note we will prove a Theorem we can use to show that $\log(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}x^n}{n}$ if $-1 < x < 1$.

Theorem: *If $f(x)$ and $g(x)$ are differentiable in an open interval I (I may be unbounded) and $f'(x) = g'(x)$ for all $x \in I$, then $f(x) - g(x) \equiv C$ in I, where C is a constant.*

As usual is is easier to prove this if we work with $h(x) := f(x) - g(x)$ in I and show that

$$\text{(1)} \qquad \textit{if } h'(x) = 0 \textit{ for all } x \in I, \textit{ then for every } x_o \in I, \quad h(x) = h(x_o) \textit{ for all } x \in I.$$

We start with an idea: show that, given $\epsilon > 0$, for any point x in I such that $x \neq x_o$, $|h(x) - h(x_o)| \leq \epsilon$. If we can do this, then $h(x) \neq h(x_o)$ is impossible. Otherwise we could choose $\epsilon = |h(x) - h(x_o)|/2 > 0$ and obtain the contradiction $0 < |h(x) - h(x_o)| \leq \epsilon = |h(x) - h(x_o)|/2$. Our proof that $|h(x) - h(x_o)| \leq \epsilon$ will involve getting from x to x_o in small steps that can be described in terms of small intervals centered at the points of a closed interval.

To choose the small intervals we let $a := \min\{x, x_o\}$, $b := \max\{x, x_o\}$. For each $x \in [a, b]$ we know that $h'(x) = 0$. Therefore there exists $\delta = \delta_x > 0$ such that $\left|\frac{h(y) - h(x)}{y - x}\right| = \left|\frac{h(y) - h(x)}{y - x} - 0\right| < \frac{\epsilon}{b - a}$ if $|y - x| < \delta$. We constrain δ by demanding that $0 < \delta < b - a$. Because $h(y)$ is continuous, we can let y converge to $x \pm \delta_x$ (if those points are in $[a, b]$) so we have to allow

$$\text{(2)} \qquad \left|\frac{h(y) - h(x)}{y - x}\right| \leq \frac{\epsilon}{b - a} \quad \text{if } |y - x| \leq \delta_x \text{ and } y \in [a, b].$$

We then put $I_x := (x - \delta_x, x + \delta_x)$. These are our small intervals I_x, one for each $x \in [a, b]$.

We will use a very important Covering Lemma to "find," among our infinite collection of intervals I_x, a finite number, N, of intervals I_k, $1 \leq k \leq N$, that "cover" $[a, b]$ in the sense that their union contains $[a, b]$. We'll write $I_k = (x_k - \delta_k, x_k + \delta_k)$ instead of $I_{x_k} = (x_k - \delta_{x_k}, x_k + \delta_{x_k})$, a much too cumbersome notation!

We will use a "selection procedure" to find (for some positive integer M), $M + 1$ "special" points y_k, $0 \leq k \leq M$, such that $a = y_0$, $y_M = b$ and $y_k < y_{k+1}$ for $1 \leq k < M$. What makes these points special is that we will show, using (2), that $|h(y_k) - h(y_{k-1})| \leq \epsilon\frac{|y_k - y_{k-1}|}{b-a}$ for each k under consideration. To make sure (2) can be applied we have to build the points y_k so that, of any two consecutive ones, one is the center of one of our intervals I_k, and the other is in I_k. Then

$$(3) \qquad |h(x) - h(x_o)| = |h(b) - h(a)| = \left|\sum_{k=1}^{M}[h(y_k) - h(y_{k-1})]\right| \leq \sum_{k=1}^{M}|h(y_k) - h(y_{k-1})| \leq \frac{\epsilon}{(b-a)}\sum_{k=1}^{M}|y_k - y_{k-1}|.$$

Since $|y_k - y_{k-1}| = y_k - y_{k-1}$, we can summarize the last chain of equalities and inequalities as

$$|h(x) - h(x_o)| = |h(b) - h(a)| \leq \frac{\epsilon}{(b-a)}\sum_{k=1}^{M}(y_k - y_{k-1}) = \frac{\epsilon}{(b-a)}(b-a) = \epsilon.$$

The following Covering Lemma takes an infinite collection of open intervals that "covers" $[a,\, b]$ (meaning the union of the intervals contains $[a,\, b]$) and throws away all but a finite number of them. The finite number not thrown away still cover $[a,\, b]$. We will then have to carefully build the special points mentioned earlier, based on the selected intervals. The Covering Lemma and variations on it are extremely applicable in many area of mathematics.

Covering Lemma: *If* $[a,\, b]$ *is a closed and bounded interval and for each* $x \in [a,\, b]$ *there exists an open interval* $I_x = (x - \delta_x,\, x + \delta_x)$, *where* $\delta_x > 0$ *for each* $x \in [a,\, b]$, *then there exists a finite set* $\{x_1,\, \ldots,\, x_N\}$ *of points in* $[a,\, b]$ *such that*

$$[a,\, b] \subseteq \bigcup_{i=1}^{N}(x_i - \delta_{x_i},\, x_i + \delta_{x_i}) = \bigcup_{i=1}^{N} I_{x_i}; \quad \textit{the finite union of open intervals still covers (contains)}\ [a,\, b].$$

Note 1: The notation here for the finite set of intervals that *covers* $[a,\, b]$ is very ugly. When we apply the Lemma we will replace the δ_{x_i} by δ_i, and write, instead,

$$[a,\, b] \subseteq \bigcup_{i=1}^{N}(x_i - \delta_i,\, x_i + \delta_i).$$

Note 2: We will assume that there is only one interval I_x for each $x \in [a,\, b]$. This is for our convenience.

Proof of the Covering Lemma: This proof resembles that of the Intermediate Value Theorem. We "construct" a carefully chosen set S, show that it has a supremum, and then prove that the supremum gives us what we want, in this case a finite covering.

The choice of the set S comes from a common mathematical "idea:" *proceed in small steps.* We define

$$S := \{y \in [a,\, b] : \text{ the interval } [a,\, y] \text{ is covered by finitely many of the sets } I_x\}.$$

Since we want $[a,\, b]$ covered by finitely many of the sets I_x, the thing we want to prove is that $b \in S$.

First we notice that S is bounded above by b. To prove that S is not empty, we notice that the interval $[a, a]$ is contained in I_a because $[a, a]$ contains a alone. Hence S is bounded above and non-empty. Thus $c := \sup S$ exists. But since $a \in I_a$, the interval $[a, a + \delta_a/2]$ is covered by the interval I_a so $a + \delta_a/2 \in S$ as well. Actually, $a{+}t\delta_a{\in}S$ for all t, $0{<}t{<}1$; we use $t{=}1/2$ just for convenience. Thus every upper bound for S is strictly larger than a. In particular, $\sup S > a$.

We proceed in two steps: show that $c \in S$, then show that this implies $c = b$, so that $b \in S$.

Oddly enough, we don't use contradiction in the first step. We notice that $c \in I_c = (c - \delta_c, c + \delta_c)$. Thus since $c = \sup S$ there exists $x \in S$ such that $x > c - \delta_c$. It is therefore true that $x \in (c - \delta_c, c + \delta_c)$. There are finitely many intervals I_y that cover $[a, x]$ and by adding to this collection the interval I_c we obtain a finite covering of $[a, c]$. Thus $c \in S$. This completes the first step.

If $c = b$ we are done. If $c < b$, we can find a contradiction. For the proof that $[a, c]$ can be covered by finitely many intervals I_y actually shows that the interval $[a, c + \delta_c/2]$ is covered by the same intervals, and so, if $c < b$ there are numbers c' in $(c, c + \delta_c/2]$ that are less than b. Each such $c' \in S$ because c' is in $[a, b]$ and $c' \le b$, and $[a, c']$ can be covered by finitely many intervals I_y. This would give an element of S larger than $\sup S$. Thus $c = b$. Since $c \in S$, we have $b \in S$. This completes the proof of the Covering Lemma.

Remark: It was very important that the intervals I_x in the Covering Lemma were *open intervals*. We needed to be sure that c being in I_c meant that points on both sides of c belonged to I_c.

A technical observation about finite coverings, based on a selection procedure

The purpose of the next part of this note is to build the "special" points y_k, $0 \le k \le M$, mentioned earlier. The reason for the word "technical" is that some of the points y_k have to be centers of covering intervals, and some of the y_k have to belong to *two* covering intervals so they can provide a "link" between the points that are centers of covering intervals. I can imagine this observation as useful for studying the behavior of volatile stock prices.

We will invent a term to use in the following construction by Finite Induction. We will say that the interval I_K *has higher rank* than a different interval I_J if $x_K + \delta_K > x_J + \delta_J$ or if $x_K + \delta_K = x_J + \delta_J$ but $x_K - \delta_K < x_J - \delta_J$. You should prove on your own that, in the second case (namely $x_K + \delta_K = x_J + \delta_J$ but $x_K - \delta_K < x_J - \delta_J$), it is true that $I_J \subseteq I_K$ and $I_J \neq I_K$. By the way we chose them, our intervals, if different, have different centers.

Remarks: If A and B are among our intervals, and $A \neq B$ and $B \subseteq A$ then A outranks B. If $A \neq B$ and neither contains the other, then each has a point not in the other. If A contains an upper bound for B (by the way, all of our intervals are bounded) then A outranks B. If A does not contain an upper bound for B then B contains an upper bound for A since neither interval contains the other. This is not true of sets in general, but is true of intervals. In particular, if a number p does not belong to an interval A then p is either an upper bound for A or is a lower bound for A. You should prove all the mathematical statements in this paragraph.

An important point, left to you to prove, is that, of any two different ones of our intervals, one outranks the other.

To begin our construction we select the interval I_{i_o} of highest rank that contains a. We continue the construction as follows. If the interval just selected contains b, no more intervals are selected. If the interval just selected is, say, $I_i = (x_i - \delta_i,\, x_i + \delta_i)$ and it does not contain b, we next select the interval I_j of highest rank that contains $x_i + \delta_i$, the right-hand endpoint of the interval just selected. This new interval I_j becomes the "interval just selected." We then return to the second sentence of this papagraph to continue the selection process.

Since no open interval contains either of its endpoints, the number of intervals from which we can make a selection decreases by at least one each time we make a selection. The selection thus stops after a finite number of steps. Since b is contained in at least one interval of the collection, any interval that contains b outranks all those intervals that do not contain b. Thus each interval that contains b remains a candidate for selection. Eventually an interval that contains b must be selected. Otherwise, infinitely many intervals would have to be selected, and this is impossible.

Now that intervals have been selected, we will use their centers and possibly some other points to build our "special points" y_k.

The centers of the selected intervals increase strictly. Suppose that I_i and then I_j are selected. Then $x_i + \delta_i < x_j + \delta_j$, so if on the contrary we had $x_j \leq x_i$, we would have $x_i + \delta_i < x_j + \delta_j \leq x_i + \delta_j$, which implies that $\delta_j > \delta_i$. But then $x_j - \delta_j \leq x_i - \delta_j < x_i - \delta_i$. This means that $I_i \subseteq I_j$, and I_j outranks I_i, so I_i is not the interval of highest rank containing *any* of its points. Since I_i *is* a selected interval, this is a contradiction, so the centers do indeed increase strictly. We notice that the inequality $\delta_j > \delta_i$ was only proven true under the erroneous assumption that $x_j \le x_i$. We actually know nothing at all about the relative sizes of δ_i and δ_j.

If $x_i < x_j$ are the centers of consecutively chosen intervals I_i and I_j, there must exist a point y that lies in both of $\bar{I}_i$ and $\bar{I}_j$ such that $x_i \le y \le x_j$. We will prove this. If $x_i \ge x_j - \delta_j$, we set $y := x_i$. If $x_i < x_j - \delta_j$ we set $y := x_j - \delta_j$. In the first case, $x_i \ge x_j - \delta_j$, we have $x_j - \delta_j \le x_i < x_j$, so $y = x_i \in I_i \cap \bar{I}_j \subseteq \bar{I}_i \cap \bar{I}_j$. In the second case, $x_i < x_j - \delta_j$, we have $x_i < x_j - \delta_j < x_i + \delta_i$ since $x_i + \delta_i \in I_j$. But then $y = x_j - \delta_j \in I_i \cap \bar{I}_j \subseteq \bar{I}_i \cap \bar{I}_j$.

We chose the points y because they belonged to both of two consecutively selected intervals. We can now say that $h(x_i) - h(x_j) = h(x_i) - h(y) + h(y) - h(x_j)$, and hence, by the inequality (2) and the triangle inequality,

$$|h(x_i) - h(x_j)| \le |h(x_i) - h(y)| + |h(y) - h(x_j)| \le \frac{\epsilon}{b-a}\left(|x_i - y| + |y - x_j|\right) = \frac{\epsilon}{b-a}|x_i - x_j| = \frac{\epsilon}{b-a}(x_j - x_i).$$

We also have $|h(x_{i_o}) - h(a)| \le \frac{\epsilon}{b-a}(x_{i_o} - a)$ and $|h(b) - h((x_{j_o})| \le \frac{\epsilon}{b-a}(b - x_{j_o})$. This makes the argument leading up to (3) work, and in essence completes the proof! The “special points” were implicit here.

We can be a little more formal about the “special points.” First we notice that $y = \max\{x_i,\, x_j - \delta_j\}$. Let us denote by I_{i_o} the first selected interval, and denote by I_{i_o} the last selected interval. They might be the same. If they are not, the intervals $[x_i,\, \max\{x_i,\, x_j - \delta_j\}]$ and $[\max\{x_i,\, x_j - \delta_j\},\, x_i]$, constructed for all consecutive pairs of selected intervals, together with $[a,\, x_{i_o}]$ and $[x_{j_o},\, b]$, cover $[a,\, b]$ and meet only in single points. We next drop any of these intervals that contains only one point. For example, it is possible that $x_{i_o} = a$, and then $[a,\, x_{i_o}] = [a,\, a]$ contains only the one point, a. The remaining intervals all have positive length, and their union is $[a,\, b]$. At last we can define our “special points,” by putting the endpoints of our intervals in increasing order, dropping repetitions. We name these endpoints y_k. Our Theorem is proved.

An application of the Theorem

We know that $\log(1+x)$ is differentiable for $x > -1$, and that its derivative is $\frac{1}{1+x}$ there. We also know that the power series $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}x^n}{n}$ converges absolutely for $|x| < 1$, is thus differentiable term-by-term, having derivative $\sum_{n=1}^{\infty} (-1)^{n-1}x^{n-1} = \frac{1}{1+x}$. Therefore, in the open interval $(-1,\, 1)$ the series and the function have the same derivative. Thus they differ by a constant. The constant is zero, as we can see by choosing $x = 0$ as the point at which to evaluate the series and the function $\log(1+x)$. Therefore,

$$\log(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}x^n}{n}, \quad \text{if} \quad -1 < x < 1.$$