ROLLE'S THEOREM AND THE MEAN VALUE THEOREM

WILLIAM A. LAMPE

Recall the

Theorem on Local Extrema. *If $f(c)$ is a local extremum, then either f is not differentiable at c or $f'(c) = 0$. That is, at a local max or min f either has no tangent, or f has a horizontal tangent there.*

We will use this to prove

Rolle's Theorem. *Let $a < b$. If f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b) and $f(a) = f(b)$, then there is a c in (a, b) with $f'(c) = 0$. That is, under these hypotheses, f has a horizontal tangent somewhere between a and b.*

Rolle's Theorem, like the Theorem on Local Extrema, ends with $f'(c) = 0$. The proof of Rolle's Theorem is a matter of examining cases and applying the Theorem on Local Extrema.

Proof. We seek a c in (a, b) with $f'(c) = 0$. That is, we wish to show that f has a horizontal tangent somewhere between a and b.

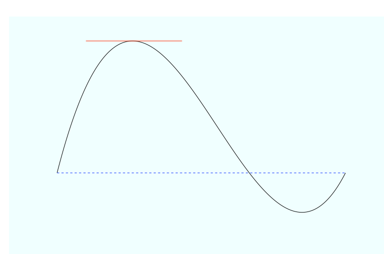
Since f is continuous on the closed interval $[a, b]$, the Extreme Value Theorem says that f has a maximum value $f(M)$ and a minimum value $f(m)$ on the closed interval $[a, b]$. Either $f(M) = f(m)$ or $f(M) \neq f(m)$.

Case 1. We suppose the maximum value $f(M) = f(m)$, the minimum value. So all values of f on $[a, b]$ are equal, and f is constant on $[a, b]$. Then $f'(x) = 0$ for all x in (a, b). So one may take c to be anything in (a, b); for example, $c = \frac{a+b}{2}$ would suffice.

Case 2. Now we suppose $f(M) \neq f(m)$. So at least one of $f(M)$ and $f(m)$ is not equal to the value $f(a) = f(b)$.

Case 2.a

We first consider the case where the maximum value $f(M) \neq f(a) = f(b)$. (See the figure to the right.)

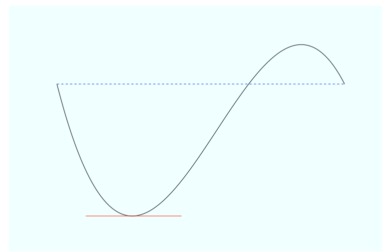


So M is neither a nor b. But M is in $[a, b]$ and not at the end points. So M must be in the open interval (a, b). We have the maximum value $f(M) \geq f(x)$ for all x in the closed interval$[a, b]$ which contains the open interval (a, b). So we also have $f(M) \geq f(x)$ for every x in the open interval (a, b), Since M is also in the open interval (a, b), this means by definition that $f(M)$ *is a local maximum.*

Since M is in the open interval (a, b), by hypothesis we have that f is differentiable at M. *Now* by the Theorem on Local Extrema, we have that f has a horizontal tangent at m; that is, we have that $f'(M) = 0$. So we take $c = M$, and we are done with this case.

Case 2.b

We now consider the case where the minimum value $f(m) \neq f(a) = f(b)$. (This case is very similar to the previous case. Also, see the figure to the right.)



So m is neither a nor b. But m is in $[a, b]$ and not at the endpoints. So m must be in the open interval (a, b). We have the minimum value $f(m) \leq f(x)$ for all x in the closed interval $[a, b]$ which contains the open interval (a, b). Thus $f(m) \leq f(x)$ for every x in the open interval (a, b), Since m is also in the open interval (a, b), this means by definition that $f(m)$ *is a local minimum.*

Since m is in the open interval (a, b), by hypothesis we have that f is differentiable at m. *Now* by the Theorem on Local Extrema, we have that f has a horizontal tangent at m; that is, we have that $f'(m) = 0$. So we take $c = m$, and we are done with this case.

Our list of cases covers all possibilities which ends the proof.

Next we give an application of Rolle's Theorem and the Intermediate Value Theorem.

Example. We show that $x^5 + 4x = 1$ has exactly one solution. Let $f(x) = x^5 + 4x$. Since f is a polynomial, f is continuous everywhere. $f'(x) = 5x^4 + 4 \geq 0 + 4 = 4 > 0$ for all x. So $f'(x)$ is never 0. So by Rolle's Theorem, no equation of the form $f(x) = C$ can have 2 or more solutions. In particular $x^5 + 4x = 1$ has at most one solution.

$f(0) = 0^5 + 4 \cdot 0 = 0 < 1 < 5 = 1 + 4 = f(1)$. Since f is continuous everywhere, by the Intermediate Value Theorem, $f(x) = 1$ has a solution in the interval $[0, 1]$.

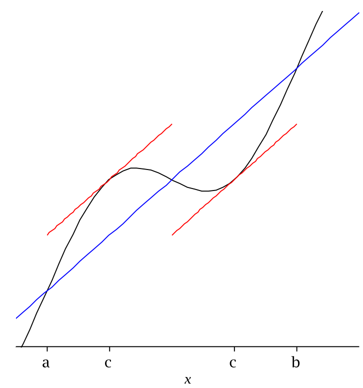
Together these reults say $x^5 + 4x = 1$ has exactly one solution, and it lies in $[0, 1]$.

The traditional name of the next theorem is the Mean Value Theorem. A more descriptive name would be Average Slope Theorem.

Mean Value Theorem. *Let $a < b$. If f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then there is a c in (a, b) with*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

The Mean Value Theorem says that under appropriate smoothness conditions the slope of the curve at some point between a and b is the same as the slope of the line joining $\langle a, f(a)\rangle$ to $\langle b, f(b)\rangle$. The figure to the right shows two such points, each labeled c.



If f satisfies the hypotheses of the Rolle's Theorem, then the Mean Value theorem also applies and $f(b) - f(a) = 0$. For the c given by the Mean Value Theorem we have $f'(c) = \frac{f(b)-f(a)}{b-a} = 0$. So the Mean Value Theorem says nothing new in this case, but it does add information when $f(a) \neq f(b)$.

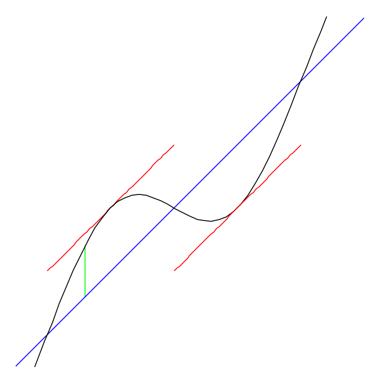
The proof of the Mean Value Theorem is accomplished by finding a way to apply Rolle's Theorem. One considers the line joining the points $\langle a, f(a)\rangle$ and $\langle b, f(b)\rangle$. The difference between f and that line is a function that turns out to satisfy the hypotheses of Rolle's Theorem, which then yields the desired result.

Proof. Suppose f satisfies the hypotheses of the Mean Value Theorem. The line joining the points $\langle a, f(a)\rangle$ and $\langle b, f(b)\rangle$ has equation

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

We let g be the difference between f and this line.

That is, $$g(x) = f(x) - \left[f(a) + \frac{f(b) - f(a)}{b - a}(x - a)\right].$$



$g(x)$ is the height of the vertical green line in the figure to the right.

g is the difference of two continuous functions. So g is continuous on $[a, b]$.

g is the difference of two differentiable functions. So g is differentiable on (a, b). Moreover, the derivative of g is the difference between the derivative of f and the derivative (slope) of the line. That is,

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

Both f and the line go through the points $\langle a, f(a)\rangle$ and $\langle b, f(b)\rangle$. So the difference between them is 0 at a and at b. Indeed,

$$g(a) = f(a) - \left[f(a) + \frac{f(b)-f(a)}{b-a}(a-a)\right] = f(a) - [f(a)+0] = 0, \qquad \text{and}$$
$$g(b) = f(b) - \left[f(a) + \frac{f(b)-f(a)}{b-a}(b-a)\right] = f(b) - [f(a)+f(b)-f(a)] = 0.$$

So Rolle's Theorem applies to g. So there is a c in the open interval (a, b) with $g'(c) = 0$. Above we calculated $g'(x)$. Using that we have

$$0 = g'(c) = f'(c) - \frac{f(b)-f(a)}{b-a}$$

which is what we needed to prove.

Example. We illustrate The Mean Value Theorem by considering $f(x) = x^3$ on the interval $[1, 3]$.

f is a polynomial and so continuous everywhere. For any x we see that $f'(x) = 3x^2$. So f is continuous on $[1, 3]$ and differentiable on $(1, 3)$. So the Mean Value theorem applies to f and $[1, 3]$.

$$\frac{f(b)-f(a)}{b-a} = \frac{f(3)-f(1)}{3-1} = \frac{27-1}{2} = 13.$$

$f'(c) = 3c^2$. So we seek a c in $[1, 3]$ with $3c^2 = 13$.

$3c^2 = 13$ iff $c^2 = \frac{13}{3}$ iff $c = \pm\sqrt{\frac{13}{3}}$.

$-\sqrt{\frac{13}{3}}$ is not in the interval $(1, 3)$, but $\sqrt{\frac{13}{3}}$ is a little bigger than $\sqrt{\frac{12}{3}} = \sqrt{4} = 2$. So $\sqrt{\frac{13}{3}}$ is in the interval $(1, 3)$.

So $c = \sqrt{\frac{13}{3}}$ is in the interval $(1, 3)$, and

$$f'(c) = f'\left(\sqrt{\frac{13}{3}}\right) = 13 = \frac{f(3)-f(1)}{3-1} = \frac{f(b)-f(a)}{b-a}.$$

PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 81, Number 1, January 1981

AN ELEMENTARY PROOF OF THE STONE-WEIERSTRASS THEOREM

BRUNO BROSOWSKI AND FRANK DEUTSCH[1]

ABSTRACT. An elementary proof of the Stone-Weierstrass theorem is given.

In this note we give an elementary proof of the Stone-Weierstrass theorem. The proof depends only on the *definitions* of compactness ("each open cover has a finite subcover") and continuity ("the inverse images of open sets are open"), two simple consequences of these definitions (i.e. "a closed subset of a compact space is compact," and "a positive continuous function on a compact set has a positive infimum"), and the elementary Bernoulli inequality:

$$(1 + h)^n \geqslant 1 + nh \qquad (n = 1, 2, \ldots)$$

if $h \geqslant -1$.

In the beautiful and elementary proof of the classical Weierstrass theorem given by Kuhn [**1**], it is observed that it suffices to be able to approximate, by polynomials, the step function which is 1 on the interval $[0, \frac{1}{2})$ and 0 on the interval $[\frac{1}{2}, 1]$ uniformly *outside* of each neighborhood of $\frac{1}{2}$. The main step in our proof (Lemma 2) is the general analogue of this. It shows that it suffices to be able to approximate, by elements of the subalgebra, a given "generalized step function" (i.e. a function which is 0 on a closed set and 1 off the set) uniformly on the closed set and off a neighborhood of this set.

It should be remarked that when our proof is specialized to the classical case of polynomials in $C[a, b]$, it is even "simpler" than Kuhn's proof in the sense that no "change of variables" argument is necessary, nor is it necessary to appeal to the fact that continuous functions on $[a, b]$ are uniformly continuous. (Kuhn's proof also used the Bernoulli inequality.)

In particular, it is perhaps worth emphasizing that, in contrast to many proofs of the Stone-Weierstrass theorem, we do *not* appeal to any of the following facts:

(a) the classical Weierstrass theorem (nor even the special case of uniformly approximating $f(t) = |t|$ on $[-1, 1]$ by polynomials);

(b) that the closure of a subalgebra is a subalgebra;

(c) that the closure of a subalgebra is a sublattice.

Let T be a compact topological space and $C(T)$ the set of all real-valued continuous functions on T. A neighborhood of a point in T is an open set which

---

Received by the editors July 10, 1979 and, in revised form, April 24, 1980.

1980 *Mathematics Subject Classification*. Primary 46J30, 46E15; Secondary 41A30.

Key words and phrases. Stone-Weierstrass theorem, Weierstrass theorem, approximation by subalgebras, polynomial approximation.

[1]On leave from the Pennsylvania State University, University Park, Pennsylvania 16802.

0002-9939/81/0000-0019/$02.00

contains the point. Let $\mathfrak{A}$ be a subset of $C(T)$ with the properties:

(i) x, y in $\mathfrak{A}$, α, β in $\mathbf{R}$ implies $\alpha x + \beta y \in \mathfrak{A}$;

(ii) x, y in $\mathfrak{A}$ implies $x \cdot y \in \mathfrak{A}$;

(iii) $1 \in \mathfrak{A}$;

(iv) if t_1 and t_2 are distinct points in T, then there exists $x \in \mathfrak{A}$ such that $x(t_1) \neq x(t_2)$.

In other words, $\mathfrak{A}$ is a "subalgebra of $C(T)$ which contains constants and separates points."

The Stone-Weierstrass theorem may be stated as follows. *If $\mathfrak{A}$ is a subalgebra of $C(T)$ which contains constants and separates points, then the elements of $C(T)$ can be uniformly approximated by the elements of $\mathfrak{A}$.* That is, given $f \in C(T)$ and $\varepsilon > 0$ there exists $g \in \mathfrak{A}$ such that $\sup_{t \in T}|f(t) - g(t)| < \varepsilon$.

It is convenient to divide the proof into three steps. The essential step is Lemma 1. For brevity, the norm notation $\|x\| = \sup\{|x(t)| \,|\, t \in T\}$ will sometimes be used.

LEMMA 1. *Let $t_0 \in T$ and let U be a neighborhood of t_0. Then there is a neighborhood $V = V(t_0)$ of t_0, $V \subset U$, with the following property. For each $\varepsilon > 0$, there exists $x \in \mathfrak{A}$ such that*

(1) $0 \leq x(t) \leq 1, t \in T$;

(2) $x(t) < \varepsilon, t \in V$;

(3) $x(t) > 1 - \varepsilon, t \in T \setminus U$.

PROOF. For each $t \in T \setminus U$, the point separating property (iv) implies that there is a function $g_t \in \mathfrak{A}$ with $g_t(t) \neq g_t(t_0)$. Then the function $h_t = g_t - g_t(t_0) \cdot 1$ is in $\mathfrak{A}$ and $h_t(t) \neq h_t(t_0) = 0$. The function $p_t = (1/\|h_t^2\|)h_t^2$ is in $\mathfrak{A}$, $p_t(t_0) = 0$, $p_t(t) > 0$, and $0 \leq p_t \leq 1$.

Let $U(t) = \{s \in T | p_t(s) > 0\}$. Then $U(t)$ is a neighborhood of t. By compactness of $T \setminus U$, there exist a finite number of points $\{t_1, t_2, \ldots, t_m\}$ in $T \setminus U$ such that $T \setminus U \subset \bigcup_1^m U(t_i)$. Let $p = (1/m)\sum_1^m p_{t_i}$. Then $p \in \mathfrak{A}$, $0 \leq p \leq 1$, $p(t_0) = 0$, and $p > 0$ on $T \setminus U$.

Again using the compactness of $T \setminus U$, there exists $0 < \delta < 1$ such that $p \geq \delta$ on $T \setminus U$. Let $V = \{t \in T | p(t) < \delta/2\}$. Then V is a neighborhood of t_0 and $V \subset U$.

Let k be the smallest integer which is greater than $1/\delta$. Then $k - 1 \leq 1/\delta$ which implies that $k \leq (1 + \delta)/\delta < 2/\delta$. Thus $1 < k\delta < 2$. Consider the functions q_n defined by

$$q_n(t) = \left[1 - p^n(t)\right]^{k^n} \qquad (n = 1, 2, \ldots).$$

Clearly, $q_n \in \mathfrak{A}$, $0 \leq q_n \leq 1$, and $q_n(t_0) = 1$. For each $t \in V$, $kp(t) \leq k\delta/2 < 1$ so that, by Bernoulli's inequality,

$$q_n(t) \geq 1 - \left[kp(t)\right]^n \geq 1 - (k\delta/2)^n \to 1$$

uniformly on V. For $t \in T \setminus U$, $kp(t) \geq k\delta > 1$ and, using Bernoulli's inequality,

$$q_n(t) = \frac{1}{k^n p^n(t)}\left[1 - p^n(t)\right]^{k^n} k^n p^n(t) < \frac{1}{\left[kp(t)\right]^n}\left[1 - p^n(t)\right]^{k^n}\left[1 + k^n p^n(t)\right]$$

$$< \frac{1}{\left[kp(t)\right]^n}\left[1 - p^n(t)\right]^{k^n}\left[1 + p^n(t)\right]^{k^n} = \frac{1}{\left[kp(t)\right]^n}\left[1 - p^{2n}(t)\right]^{k^n}$$

$$< \frac{1}{(k\delta)^n} \to 0$$

uniformly on $T \setminus U$.

Thus for n sufficiently large, the function q_n has the property that $0 \leqslant q_n \leqslant 1$, $q_n < \varepsilon$ on $T \setminus U$, and $q_n > 1 - \varepsilon$ on V. The result follows by taking $x = 1 - q_n$. □

LEMMA 2. *Let A and B be disjoint closed subsets of T. Then for each* $0 < \varepsilon < 1$, *there exists $x \in \mathfrak{A}$ such that*

(1) $0 \leqslant x(t) \leqslant 1, t \in T$;

(2) $x(t) < \varepsilon, t \in A$;

(3) $x(t) > 1 - \varepsilon, t \in B$.

PROOF. Let $U = T \setminus B$. For each $t \in A$, choose the neighborhood $V(t)$ of t as in Lemma 1. Then there exists a finite set of points $\{t_1, t_2, \ldots, t_m\}$ in A such that $A \subset \bigcup_1^m V(t_i)$. By the choice of $V(t_i)$, there exist $x_i \in \mathfrak{A}$ $(i = 1, 2, \ldots, m)$ with $0 \leqslant x_i \leqslant 1$, $x_i < \varepsilon/m$ on $V(t_i)$, and $x_i > 1 - \varepsilon/m$ on $T \setminus U = B$. Then the function $x = x_1 \cdot x_2 \cdots x_m$ is in $\mathfrak{A}$, $0 \leqslant x \leqslant 1$, $x < \varepsilon/m \leqslant \varepsilon$ on $\bigcup_1^m V(t_i) \supset A$, and (using Bernoulli's inequality) $x > (1 - \varepsilon/m)^m \geqslant 1 - \varepsilon$ on B. □

Finally, we turn to the proof of the Stone-Weierstrass theorem. Let $f \in C(T)$ and $\varepsilon > 0$. To complete the proof, it suffices to show the existence of a $g \in \mathfrak{A}$ such that

$$|f(t) - g(t)| < 2\varepsilon, \qquad t \in T. \tag{$*$}$$

By replacing f by $f + \|f\|$, we can assume that $f \geqslant 0$. We may also assume that $\varepsilon < \frac{1}{3}$. Choose an integer n so that $(n - 1)\varepsilon \geqslant \|f\|$. Define the sets A_j, B_j $(j = 0, 1, \ldots, n)$ by

$$A_j = \left\{t \in T | f(t) \leqslant \left(j - \tfrac{1}{3}\right)\varepsilon\right\}, \qquad B_j = \left\{t \in T | f(t) \geqslant \left(j + \tfrac{1}{3}\right)\varepsilon\right\}.$$

Note that A_j and B_j are disjoint closed sets in T, $\varnothing = A_0 \subset A_1 \subset \cdots \subset A_n = T$, and $B_0 \supset B_1 \supset \cdots \supset B_n = \varnothing$. For each $j = 0, 1, \ldots, n$, Lemma 2 implies that there is $x_j \in \mathfrak{A}$, with $0 \leqslant x_j \leqslant 1$, $x_j < \varepsilon/n$ on A_j, and $x_j > 1 - \varepsilon/n$ on B_j.

Then the function $g = \varepsilon \Sigma_0^n x_i$ is in $\mathfrak{A}$. For any $t \in T$, we have $t \in A_j \setminus A_{j-1}$ for some $j \geqslant 1$ which implies that

$$\left(j - \tfrac{4}{3}\right)\varepsilon < f(t) \leqslant \left(j - \tfrac{1}{3}\right)\varepsilon \tag{$**$}$$

and

$$x_i(t) < \varepsilon/n \quad \text{for every } i \geqslant j. \tag{$***$}$$

Also, $t \in B_i$ for every $i \leqslant j - 2$ which implies

$$x_i(t) > 1 - \varepsilon/n \quad \text{for every } i \leqslant j - 2. \tag{$****$}$$

Using (***), we obtain

$$g(t) = \varepsilon \sum_{0}^{j-1} x_i(t) + \varepsilon \sum_{j}^{n} x_i(t)$$
$$\leq j\varepsilon + \varepsilon(n - j + 1)\varepsilon/n \leq j\varepsilon + \varepsilon^2 < \left(j + \tfrac{1}{3}\right)\varepsilon.$$

Using (****), we obtain for $j \geq 2$

$$g(t) \geq \varepsilon \sum_{0}^{j-2} x_i(t) \geq (j-1)\varepsilon(1 - \varepsilon/n)$$
$$= (j-1)\varepsilon - ((j-1)/n)\varepsilon^2 > (j-1)\varepsilon - \varepsilon^2 > \left(j - \tfrac{4}{3}\right)\varepsilon.$$

The inequality $g(t) > (j - \frac{4}{3})\varepsilon$ is trivially true for $j = 1$. Thus

$$|f(t) - g(t)| \leq \left(j + \tfrac{1}{3}\right)\varepsilon - \left(j - \tfrac{4}{3}\right)\varepsilon < 2\varepsilon. \quad \square$$

References

1. H. Kuhn, *Ein elementarer Beweis des Weierstrasschen Approximationssatzes*, Arch. der Math. (Basel) **15** (1964), 316–317.

Fachbereich Mathematik, J. W. Goethe-Universität Frankfurt, Robert Mayer-Strasse 6-10, D-6000 Frankfurt, West Germany (Current address of Bruno Brosowski)

Current address (Frank Deutsch): Department of Mathematics, Pennsylvania State University, University Park, Pennsylvania 16802