Proof of the Extreme Value Theorem

Theorem: If f is a continuous function defined on a closed interval $[a, b]$, then the function attains its maximum value at some point c contained in the interval.

Proof:
There will be two parts to this proof. First we will show that there must be a finite maximum value for f (this was not done in class); second, we will show that f must attain this maximum value. Note that the methods in the two parts are very similar.

Part I

First we show that there must be a finite maximum value. We will accomplish this by assuming that there is no such value, and then deriving a contradiction.

Assume there is no maximum value; then certainly this means that the function attains larger and larger values. In particular, there must be some point c_1 with $f(c_1) > 1$, and a point c_2 with $f(c_2) > 2$, and a point c_3 with $f(c_3) > 3$,...

Now look at all of these points $c_1, c_2, ...$ It is certainly not clear that they converge to some single point in $[a, b]$; in fact they probably don't. We will however be able to find a "subsequence" that does converge to a single point; this will prove useful at the end of the proof.

We find the subsequence as follows. First, split the interval $[a, b]$ into two equally sized pieces.

We notice that there are infinitely many points c_n in $[a, b]$– so, therefore, there must be infinitely many c_n's in either the left half or the right half. Whichever half that is we will refer to as $[a_1, b_1]$. Since there are infinitely many of the c_n's in $[a_1, b_1]$, we can certainly pick one – any one, it doesn't matter which. We will call it d_1.

Now we recall that, because of our previous choice, there are infinitely many points c_n in $[a_1, b_1]$– so, therefore, there must be infinitely many c_n's in either the left half or the right half. Whichever half that is we will refer to as $[a_2, b_2]$. Since there are infinitely many of the c_n's in $[a_2, b_2]$, we can certainly pick one – any one, it doesn't matter which. We will call it d_2.

We continue in this way and end up with a sequence of points, $d_1, d_2, d_3, ...$ This sequence has two properties: 1) every point in the sequence is one of the original c_n's, and 2) the sequence converges to some point in $[a, b]$. The latter property we observe since the intervals from which we choose the d_n's are getting smaller and smaller as n gets larger. Let $\lim_{n\to\infty} d_n = d$.

We now observe that

$$f(d) = f(\lim_{n\to\infty} d_n) = \lim_{n\to\infty} f(d_n) = \lim_{n\to\infty} f(c_n)$$

(Note that we can pass the limit outside of the function since we are given that the function is continuous, and we recall that this is one of the properties of continuous functions we discussed several weeks ago; the last equality we can conclude since every d_n was chosen as one of the c_n in the first place.)

Since we chose the c_n's originally such that the values $f(c_n)$ increased to infinity, we see that this limit must not exist. But of course, the above equality would then suggest that $f(d)$ does not exist. This of course contradicts the assumption that f is defined on the entire closed interval.

Having derived a contradiction, we see that our original assumption must be false, so in fact there must be a finite maximum value.

(continued on next page...)

Part II

We have now shown that there must be a finite maximum value. Let M be the smallest such maximum value. It remains to show that there exists some point such that the value of the function there is M.

If M really is the smallest possible maximum value, then certainly this means that the function attains values closer and closer to M. In particular, for some ϵ there must be some point c_1 with $f(c_1) = M - \epsilon$, and a point c_2 with $f(c_2) = M - \frac{\epsilon}{2}$, and a point c_3 with $f(c_3) = M - \frac{\epsilon}{3}$,... (Note that these points have nothing to do with the c_n's that we chose in Part I! I use the same variable to emphasize the similarity in the ideas we are using in the two parts...)

Now look at all of these points $c_1, c_2, ...$ It is certainly not clear that they converge to some single point in $[a, b]$; in fact they probably don't. We will however be able to find a "subsequence" that does converge to a single point; this will prove useful at the end of the proof.

We find the subsequence as follows. First, split the interval $[a, b]$ into two equally sized pieces.

We notice that there are infinitely many points c_n in $[a, b]$– so, therefore, there must be infinitely many c_n's in either the left half or the right half. Whichever half that is we will refer to as $[a_1, b_1]$. Since there are infinitely many of the c_n's in $[a_1, b_1]$, we can certainly pick one – any one, it doesn't matter which. We will call it d_1.

Now we recall that, because of our previous choice, there are infinitely many points c_n in $[a_1, b_1]$– so, therefore, there must be infinitely many c_n's in either the left half or the right half. Whichever half that is we will refer to as $[a_2, b_2]$. Since there are infinitely many of the c_n's in $[a_2, b_2]$, we can certainly pick one – any one, it doesn't matter which. We will call it d_2.

We continue in this way and end up with a sequence of points, $d_1, d_2, d_3, ...$ This sequence has two properties: 1) every point in the sequence is one of the original c_n's, and 2) the sequence converges to some point in $[a, b]$. The latter property we observe since the intervals from which we choose the d_n's are getting smaller and smaller as n gets larger. Let $\lim_{n\to\infty} d_n = d$.

We now observe that

$$f(d) = f(\lim_{n\to\infty} d_n) = \lim_{n\to\infty} f(d_n) = \lim_{n\to\infty} f(c_n)$$

(Note that we can pass the limit outside of the function since we are given that the function is continuous, and we recall that this is one of the properties of continuous functions we discussed several weeks ago; the last equality we can conclude since every d_n was chosen as one of the c_n in the first place)

Since we chose the c_n's originally such that the values $f(c_n)$ were getting closer and closer to M, we see that this limit must equal M. But of course, the above equality then suggests that $f(d) = M$.

This of course shows what we were trying to prove in the first place – the function f attains its maximum value at the point d in the interval $[a, b]$.

This concludes the proof.

Part II – Alternate

We have now shown that there must be a finite maximum value. Let M be the smallest such maximum value. It remains to show that there exists some point such that the value of the function there is M.

We will prove this by contradiction. So we assume the contrary – in other words, suppose that there is no value in $[a, b]$ where the function f is equal to M – and from this we will derive a contradiction.

Consider the function

$$g(x) = \frac{1}{M - f(x)}$$

Because f does not equal M anywhere, the denominator is never zero. So, as the ratio of two continuous functions, this function g must be continuous.

And as we have already shown, any continuous function defined on a closed interval must be bounded. So, g must have some upper bound, which we call k. We then reason as follows:

$$\begin{aligned} g(x) &\leq k \\ \frac{1}{M - f(x)} &\leq k \\ M - f(x) &\geq \frac{1}{k} \\ M - \frac{1}{k} &\geq f(x) \end{aligned}$$

Note that g is always positive, and therefore so is k. So, $(M - \frac{1}{k})$ is less than M.

Thus, according to the above equation, $(M - \frac{1}{k})$ is an upper bound for f which is in fact smaller than M, contradicting our definition of M as the smallest upper bound for f.

Because we have derived a contradiction, we conclude that our original assumption must be false.

This of course shows what we were trying to prove in the first place – the function f attains its maximum value at some point in the interval $[a, b]$.

This concludes the proof.