Real Analysis

Karan Elangovan

May 16, 2021

Contents

1 Limits and Continuity

1.1 Limits

The value of a function f at a, in the abscence of any other information about f, gives absolutely no information about f for values close to a. A behaviour that is of significant interest is when f "approaches" a value (which is not necessarily fa) at a.

This is in the sense that by considering a sufficiently small neighborhood of a, all the images of f are arbitrarily close to a. We formalise this intuition in defining the limit of f at a.

Definition 1 (Limit)**.** Let f be defined on some punctured open neighborhood of a.

Then l is the limit of f at a, or $\lim_{x \to a} f(x) = l$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for all $0 < |x - a| < \delta$ we have that $|f(x) - l| < \epsilon$.

In our defintion we used the definite article, suggesting that the limit is unique. This is true however, so there is no issue.

Theorem 1 (Uniqueness of Limit)**.** The limit of a function f at a, when it exists, is unique.

Proof. Let l_1 amd l_2 be the limits of f at a. Assume for the sake of contradiction that $l_1 \neq l_2$, then we may assume by symmetry that $l_1 < l_2$.

Taking $\epsilon = \frac{l_2 - l_1}{2}$, we have two punctured open neighborhoods of a in which $|f(x) - l_1| < \epsilon$ and $|f(x) - l_2| < \epsilon$, respectively. So in their intersection we have that $f(x) < \frac{l_1 + l_2}{2}$ and $f(x) > \frac{l_1 + l_2}{2}$, which is a contradiction.

Hence we have $l_1 = l_2$, so the limit is unique. □

We now show that taking the limit behaves nicely with addition and multiplication. This is unsuprising given the intuitive idea behind a limit.

Theorem 2. Let $\lim_{x\to a} f(x) = l_1$ and $\lim_{x\to a} g(x) = l_2$. Then we have

1. $\lim_{x\to a}(f+g)(x) = l_1 + l_2$
2. $\lim_{x\to a}(fg)(x) = l_1 l_2$
3. If l_1 is non-zero, then $\lim_{x\to a}\left(\frac{1}{f}\right)(x) = \frac{1}{l_1}$

Proof. The sum rule follows trivially from the triangle inequality. The product rule requires a small trick however.

We consider an arbitrary small $\epsilon > 0$ and x sufficiently close to a such that

$$|f(x) - l_1| < \epsilon$$
$$|g(x) - l_2| < \epsilon.$$

We now have

$$\begin{aligned} &|f(x)g(x) - l_1 l_2| \\ &= |f(x)g(x) - l_1 g(x) + l_1 g(x) - l_1 l_2| \\ &< |g(x)||f(x) - l_1| + |l_1||g(x) - l_2| \\ &< \epsilon(|g(x)| + |l_1|) \\ &< \epsilon((|l_2| + \epsilon) + |l_1|) \\ &< \epsilon(|l_1| + |l_2| + 1). \end{aligned}$$

So we have the product rule.

Now we show the rule for reciprocals. First we note that as l_1 is non-zero, f is non-vanishing on some punctured open neighborhood of a, and so $\frac{1}{f}$ is defined on a punctured open neighborhood of a. We then have for sufficiently close x that

$$\begin{aligned}
&|\left(\frac{1}{f}\right)(x) - \frac{1}{l_1}| \\
&= |\frac{1}{f(x)} - \frac{1}{l_1}| \\
&= |\frac{l_1 - f(x)}{l_1 f(x)}| \\
&< \frac{1}{|l_1||f(x)|}\epsilon.
\end{aligned}$$

In a similar manner to the proof of the product rule, we may show that the ϵ coefficient is bounded above by some quantity constant with respect to x and ϵ, and so we have the reciprocal rule. □

1.2 Continuous Functions

Intuitively a continuous function is one for which the graph has no sudden jumps, infinite oscillations, etc. That is, one which we may "draw without lifting our pen from the paper". However this is only a vague heuristic from which we may infer results visually.

Definition 2 (Continuity at a Point)**.** The function f is continous at a if

$$\lim_{x \to a} f(x) = f(a).$$

Analytically this is extremely convienient as when we use continuous functions to reason about limits we no longer need to keep in mind the pesky condition that $f(a)$ itself must be ignored. For in the case of a continuous function, f actually takes on the limit value at a.

Using the results we have shown of limits of products, sums and reciprocals, we trivially have that continuous functions are closed under addition, multiplication and reciprocals (when the function does not vanish at a).

From this we have that every rational function is continuous everywhere apart from those points at which the denominator vanishes. So we may evaluate many limtis of rational functions by simply evaluating them at that point.

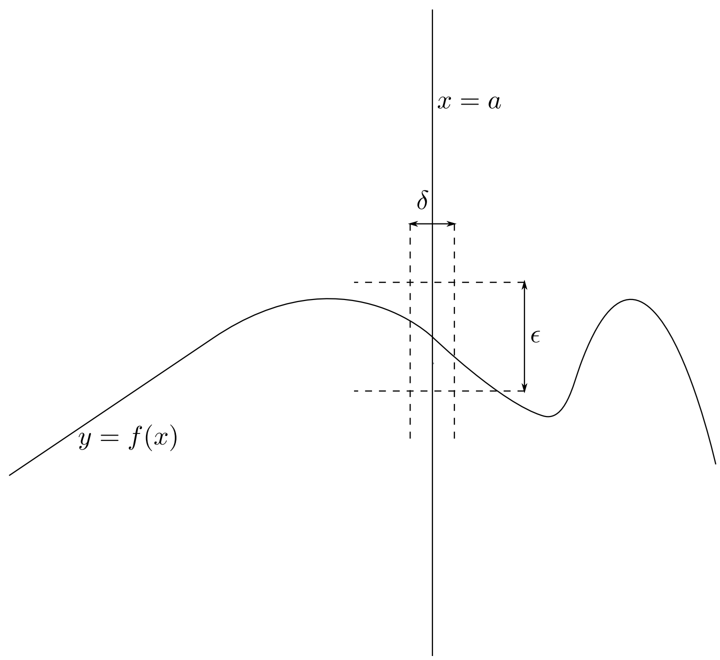

Figure 1: Illustration of continuity at a point

An extremely useful result is that the continuous functions are closed under composition.

Theorem 3 (Composition of Continuous Functions)**.** Let g be continuous at a and f continuous at $g(a)$. Then $f \circ g$ is continuous at a.

Proof. Consider an arbitrary $\epsilon > 0$. By continuity we have a $\delta_1 > 0$ such that

$$|x - g(a)| < \delta_1 \implies |f(x) - f \circ g(a)| < \epsilon.$$

Again by continuity we have a $\delta_2 > 0$ such that

$$|x - a| < \delta_2 \implies |g(x) - g(a)| < \delta_1.$$

So for $|x - a| < \delta_2$ we have

$$|f \circ g(x) - f \circ g(a)| < \epsilon.$$

□

For more non-trivial results, we need to define the notion of continuity on an interval. In the case of an open interval we simply require that the f is continuous at each point in the interval. For the case of a closed or half-closed interval, we make a small ammendment by using left and right hand limits at the boundary points.

An important point to note then is that if f is continuous on $[a, b]$ it may be discontinuous at either of the boundary points, as indeed f may not even be defined on a neighborhood of a or b.

Whilst we will never consider continuity on an arbitrary subset A, we may do so by using the topological definition of continuity with $\mathbb{R}$ under the usual metric topology and A under the subspace topology. This of course coincides with our definitions for the special cases of intervals.

Theorem 4 (Intermediate Value Theorem)**.** Let f be continuous on $[a, b]$ and k between $f(a)$ and $f(b)$. Then there exists a $c \in [a, b]$ such that $f(c) = k$.

Proof. Assume that $f(a) \leq f(b)$ as the other case uses similar logic. We also assume that k is strictly between $f(a)$ and $f(b)$, as if this is not the case then $c = a$ or $c = b$ suffices.

We define the set $A = \{x \in [a, b] : (\forall z \in [a, x] : f(z) < k)\}$. Clearly A is an interval. We see $a \in A$ and A is bounded above by b, so the supremum c of A exists.

Say $f(c) < k$, then have $c \in A$ (if this were not the case then c would be a limit point of A, and as A is an interval, this means that $A = [a, c)$, but then as $f(c) < k$ this would mean $c \in A$, which is a contradiction) However by continuity and $c < b$ we then have a $\delta > 0$ such that $c+\delta \in A$, which contradicts c being an upper bound.

Say $f(c) > k$, then we have $c \notin A$. So this means c is a limit point of A, and as A is an interval this means that $A = [a, c)$. However by continuity we have a $\delta > 0$ such that $f(c - \delta) > k$ which contradicts $A = [a, c)$.

So by trichotomy, we must have that $f(c) = k$. $\square$

Theorem 5. Let f be continuous on $[a, b]$. Then f is bounded on $[a, b]$.

Proof. We will show that f is bounded above, as showing it is bounded below uses similar logic.

We define the set $A = \{x \in [a, b] : f \text{ is bounded above on } [a, x]\}$. A is bounded above by b and $a \in A$, so the supremum α of A exists.

Say $\alpha < b$, then by continuity we would have a $\delta > 0$ such that $\alpha + \delta \in A$, contradicting α being an upper bound. If $\alpha > b$ then there would also be a contradiction as α would need to be a limit point.

Hence $\alpha = b$. If $\alpha \notin A$, then it would be a limit point of A, and so f would be bounded above on $[a, b - \epsilon]$ for every sufficiently small $\epsilon > 0$, and so by continuity at b, f would be bounded above on $[a, b]$. $\square$

Theorem 6. Let f be continuous on $[a, b]$. Then f takes on a maximum and minimum value on the interval.

Proof. We will just prove that it takes on a maximum, as the minimum uses similar logic.

We know that f is bounded above on $[a, b]$. So let α be the least of these upper bounds. Say that f never takes on the value of α. Then $f(x) < \alpha$ for all $x \in [a, b]$, and so the denominator of $\frac{1}{\alpha - f}$ is non-vanishing on $[a, b]$ and so the function is continuous.

However α is the supremum and does not belong to the set of images of f, so it is a limit point, so we may make f arbitrarily close to α and so $\frac{1}{f-\alpha}$ becomes arbitrarily large on $[a, b]$, contradicting that it must be bounded above by continuity.

So f takes on the value of α at some $c \in [a, b]$, and so $f(c) = \alpha$ is a maximum value. □

1.3 Uniform Continuity

In our formulation of continuity, we require the existence of an appropriate $\delta > 0$. However this δ may depend on ϵ. Of course the case of a function in which δ does not depend on ϵ is trivial, as this must just be a function that is constant on some neighborhood of a.

However now consider the case of a function that is continuous on an interval. In this case δ may also vary depending on which point we choose in the interval. This motivates our definition of uniform continuity.

Definition 3 (Uniform Continuity)**.** We say f is uniformly continuous on the interval I if for all $\epsilon > 0$ we have a $\delta > 0$ such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon.$$

Where x and y are arbitrary points in I.

Of course when the interval is of infinite length (e.g. the entirety of $\mathbb{R}$) even nicely behaved functions like polynomials fail to satisfy uniform continuity.

Even on an interval of finite length, we may have continuity without uniform continuity, for example with $f(x) = \frac{1}{x}$ on $(0, 1)$. Indeed as Figure 2 illustrates, we may even have a bounded continuous function on a finite interval fail to satisfy uniform continuity.

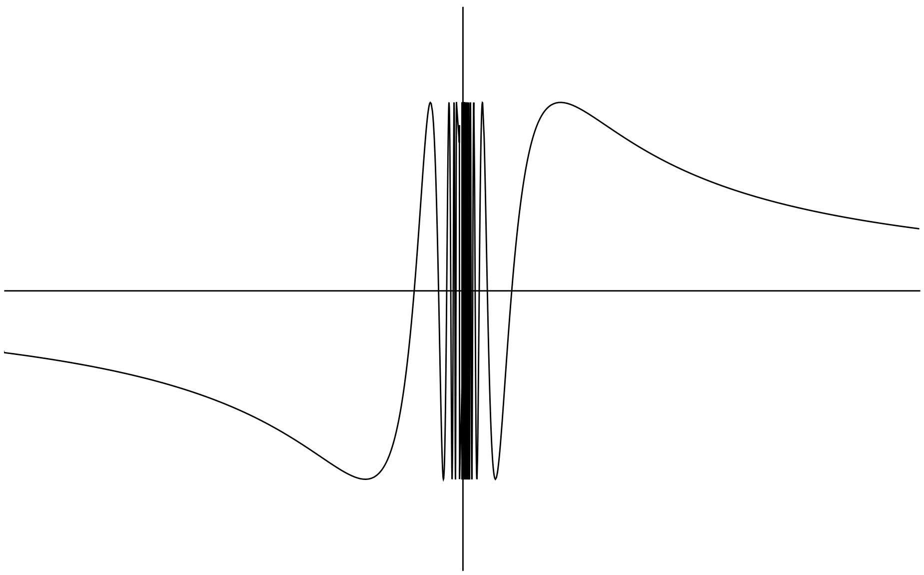

Figure 2: On the interval $(0, 1)$ the function $f(x) = \sin(\frac{1}{x})$ is continuous, however due to its infinite oscillations as we approach the origin, it fails to satisfy uniform continuity.

However it turns out that we may show a large class of functions of interest are uniformly continuous, namely those that are continuous on a closed finite interval.

Theorem 7. Let f be continuous on $[a, b]$. Then f is also uniformly continuous on $[a, b]$.

Proof. Let us assume the contrary. Then we have an $\epsilon > 0$ such that for all

$\delta > 0$ we have x, y such that

$$|x - y| < \delta$$
$$|f(x) - f(y)| \geq \epsilon.$$

Letting $\delta = \frac{1}{n}$ for $n \in \mathbb{Z}^+$ yields sequences (x_n) and (y_n) in $[a, b]$ such that

$$|x_n - y_n| < \frac{1}{n}$$
$$|f(x_n) - f(y_n)| \geq \epsilon.$$

By the Bolzano Weierstrass Theorem we have a subsequence (x_{n_k}) converging to a limit $\alpha \in [a, b]$. As $|x_{n_k} - y_{n_k}| < \frac{1}{n_k}$ we have that (y_{n_k}) also converges to α.

f is continuous at α, and (x_{n_k}) and (y_{n_k}) converge to α. So we have for sufficiently large k that

$$|f(x_{n_k}) - f(\alpha)| < \frac{\epsilon}{2}$$
$$|f(y_{n_k}) - f(\alpha)| < \frac{\epsilon}{2}.$$

Adding and applying the triangle inequality yields

$$|f(x_{n_k}) - f(y_{n_k})| < \epsilon.$$

Which provides the necessary contradiction. $\square$

2 Derivatives

2.1 The Derivative

Definition 4 (Derivative)**.** We define the derivative of f at a, $f'(a)$ as

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

When this limit exists. If the limit exists we say that f is differentiable at a.

This has an obvious physical interpretation of an instantaeneous rate of change of some quantity, in particular we can say that velocity is the derivative of displacement. Essentially we are taking better and better approximations of the rate of change by measuring over smaller and smaller time intervals.

Graphically the derivative may be seen as the gradient of the line that arises as the "limit" (in an intuitive sense as we have not formally defined the notion of a limit of points) of the secant lines through $(a, f(a))$ and $(a+h, f(a+h))$ as h tends to 0. Indeed we will define the tangent line to the graph of a function at a as the line through $(a, f(a))$ with gradient $f'(a)$.

Similarly to continuity, we will define the notion of differentiability on an open interval as being differentiable at every point. We will avoid considering the derivative over closed intervals (though this could be easily done by considering left and right hand limits of the quotient).

Theorem 8. If f is differentiable at a then f is continuous at a.

Proof. By differentiability we have

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

$$0 = \lim_{h \to 0} h.$$

Multiplying yields

$$\lim_{h\to 0}(f(a+h)-f(a))=0.$$

$\square$

Of course the converse fails to hold (for example the absolute value function is non-differentiable at 0)

When a function is differentiable on an interval, differentiation at every point gives rise to another function, f'. If this function turns out to be differentiable as well, we may differentiate once more to yield f'' and so on.

2.2 Diffrentiation

Theorem 9 (Sum Rule)**.** Let f and g be differentiable at a. Then $f+g$ is differentiable at a with derivative $f'(a)+g'(a)$

Proof. We have

$$\begin{aligned} f'(a) &= \lim_{h\to 0}\frac{f(a+h)-f(a)}{h} \\ g'(a) &= \lim_{h\to 0}\frac{g(a+h)-g(a)}{h}. \end{aligned}$$

Summing these equations yields

$$f'(a)+g'(a)=\lim_{h\to 0}\frac{(f+g)(a+h)-(f+g)(a)}{h}.$$

$\square$

Theorem 10 (Product Rule)**.** Let f and g be differentiable at a. Then fg is differentiable at a with derivative $f'(a)g(a)+f(a)g'(a)$.

Proof. We have

$$
\begin{aligned}
& f'(a)g(a) + f(a)g'(a) \\
&= g(a) \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} + f(a) \lim_{h\to 0} \frac{g(a+h)-g(a)}{h} + \lim_{h\to 0} (g(a+h)-g(a) \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} \\
&= \lim_{h\to 0} \frac{(f(a+h)-f(a))g(a) + (g(a+h)-g(a))f(a) + (g(a+h)-g(a))(f(a+h)-f(a))}{h} \\
&= \lim_{h\to 0} \frac{(fg)(a+h)-(fg)(a)}{h}.
\end{aligned}
$$

□

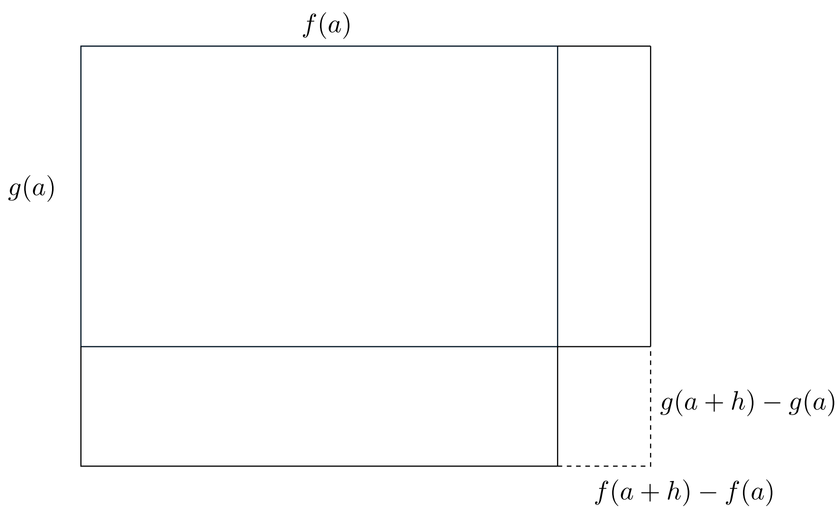

Figure 3: Geometric illustration of the manipulation in the derivation of the product rule.

Theorem 11. Let f be differentiable and non-vanishing at a. Then $\frac{1}{f}$ is differentiable at a with derivative

$$(\frac{1}{f})'(a) = -\frac{f'(a)}{f(a)^2}$$

Proof. We have the limit quotient as

$$\frac{(\frac{1}{f})(a+h) - (\frac{1}{f})(a)}{h}$$
$$= \frac{f(a) - f(a+h)}{hf(a)f(a+h)}.$$

As h tends to 0, this approaches

$$-\frac{f'(a)}{f(a)^2}.$$

□

Theorem 12 (Chain Rule)**.** Let f be differentiable at $g(a)$ and g be differentiable at a. Then $f \circ g$ is differentiable at a with derivative

$$(f \circ g)'(a) = f'(g(a))g'(a).$$

Proof. We define the function ϕ by

$$\phi(h) = \begin{cases} \frac{(f\circ g)(a+h)-(f\circ g)(a)}{g(a+h)-g(a)} & \text{if } g(a+h) - g(a) \neq 0 \\ f'(g(a)) & \text{if } g(a+h) - g(a) = 0 \end{cases}$$

So we have the limit quotient as

$$\frac{(f \circ g)(a+h) - (f \circ g)(a)}{h} = \phi(h)\frac{g(a+h) - g(a)}{h}.$$

So if we can show ϕ is continuous at 0, then we are done.

We have that

$$|\phi(h) - f'(g(a))| = |\frac{f(g(a)+k) - f(g(a))}{k} - f'(g(a))|$$

Where $k = g(a+h) - g(a)$. By the continuity of g at a, for sufficiently small h, k may be made arbitrarily small. And so the quantity above may be made arbitrarily small, hence ϕ is continuous at 0. $\square$

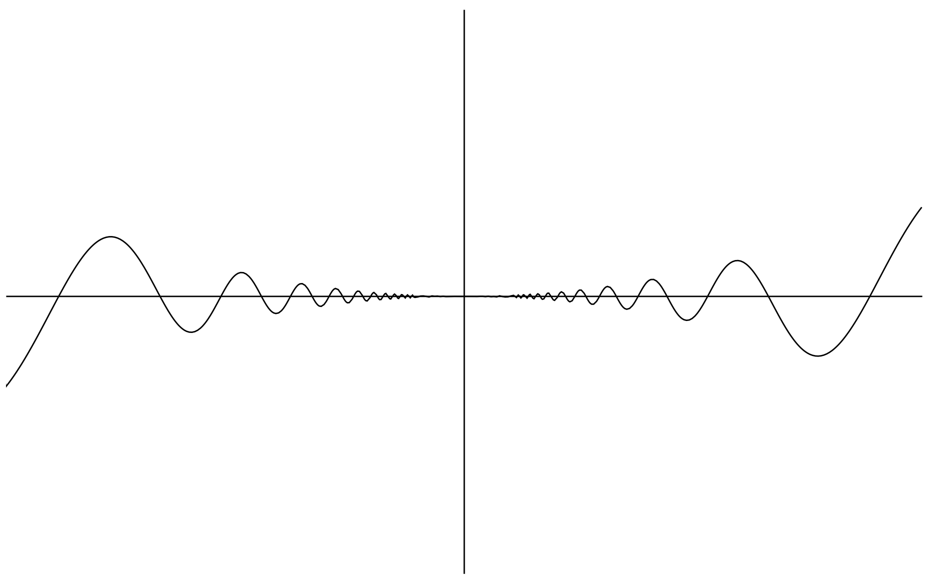

Figure 4: It is necessary for us to use the trick of defining ϕ in the derivation of the chain rule as it is possible for the quantity $g(a+h) - g(a)$ to vanish infinitely many times on any neighborhood of a. For example $g(x) = x^2 \sin(\frac{1}{x})$ when $x \neq 0$, $g(0) = 0$, is differentiable at 0.

2.3 Theorems on Derivatives

Definition 5 (Extremum Point)**.** Let f be defined on an interval I. Then we say $x \in I$ is a maximum point if for all $y \in I$ we have $f(x) \geq f(y)$. We define a minimum point similarly. We say a point is an extremum if it's a maximum or a minimum.

Definition 6 (Local Extremum)**.** Let f be defined on an open interval I. Then we say $x \in I$ is a local extremum point if we have some neighborhood of x in which x is an extremum point of f.

We see then that all extremum points are local extrema.

Definition 7 (Stationary Point)**.** We say a is a stationary point of f if f is differentiable at a and $f'(a) = 0$.

Theorem 13. Let f be differentiable on an open interval I and $a \in I$ a local extremum of f. Then we have

$$f'(a) = 0.$$

That is, a is a stationary point of f.

Proof. We will consider only the case when a is a local maximum, as then the case of a local minimum follows from considering $-f$.

We now proceed by contradiction, assuming the derivative does not vanish at a. So it is either positive or negative. Say it is positive, then on some sufficiently small open interval of a, we have

$$\frac{f(a+h) - f(a)}{h} > 0$$

So for sufficiently small $h > 0$ we have $f(a+h) > f(a)$, contradicting local maximality. We may arrive at a similar contradiction in the case the derivative is negative. $\square$

The converse however does not hold, for example define f by

$$f(x) = x^3.$$

Then 0 is a stationary point of f, however it is not a local extremum.

Theorem 14 (Rolle's Theorem)**.** Let f be continuous on $[a, b]$, differentiable on (a, b) and satisfy $f(a) = f(b)$. Then we have a stationary point in (a, b).

Proof. By continuity on the interval, f must take on a maximum and minimum value on $[a, b]$. If either of them is in (a, b) then we have the desired stationary point. Otherwise they both occur at the endpoints of the interval, and as they are equal this means the maximum and minimum are equal, so f is constant so every point in (a, b) is stationary. $\square$

We may generalise Rolle's Theorem to the much more useful Mean Value Theorem.

Theorem 15 (Mean Value Theorem)**.** Let f be continuous on $[a, b]$ and differentiable on (a, b). Then we have a point $a \in (a, b)$ such that

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

That is, f' takes on the value of the "average gradient" over the interval at some point.

Proof. We will consider an auxillary function g over $[a, b]$ which we define to be the distance between vertically aligned points on $y = f(x)$ and the line through $(a, f(a))$ and $(b, f(b))$, as in Figure 5. So we have

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

As we would expect from the geometric interpretation of g, its derivative is the difference between the derivative of f and the slope of the line, which is the average gradient. So if g has a stationary point on the interval (a, b) we are done.

This is provided by Rolle's Theorem, as the graph of f and the line intersect at the boundaries of the intervals, and so $g(a) = g(b) = 0$. $\square$

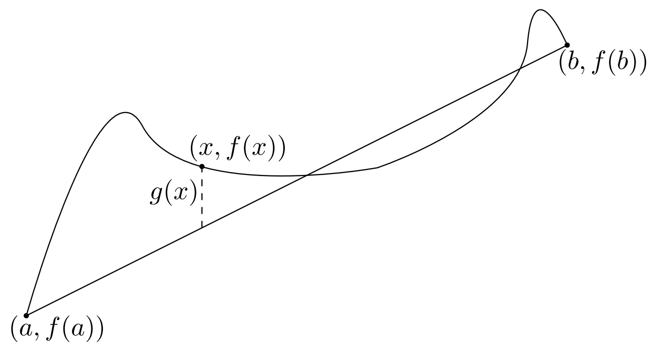

Figure 5: An illustration of the auxillary function g used in deriving the Mean Value Theorem.

We now have a particularly significant corollary of the Mean Value Theorem.

Theorem 16. Let f and g be differentiable on an open interval I. Then $f' = g'$ precisely when $f = g + C$ for some constant C.

Proof. First assume their derivatives cohere. By the sum and product rule we have

$$(f - g)' = 0.$$

Now say that $f - g$ is non-constant on I. Then by the Mean Value Theorem we would have a point in I with non vanishing derivative, which cannot be. Hence $f - g = C$.

The converse is trivial as constants vanish under differentiation. $\square$

We may also use derivativese to characterise monotonicity.

Theorem 17. Let f be differentiable on an open interval I. If $f'(x) > 0$ for all $x \in I$, then f is monotone increasing on I.

Proof. Consider arbitrary $a < b$ in I. Then by the Mean Value Theorem we have $x \in (a, b)$ such that

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

As f' is positive on I, we have

$$\begin{aligned} \frac{f(b) - f(a)}{b - a} &> 0 \\ f(b) &> f(a). \end{aligned}$$

$\square$

Of course we may show a similar result for when f is monotone decreasing. It is also important to note that the converse does not hold, as we may have montone increasing function whose derivative vanishes at isolated points, such as $f(x) = x^3$.

However if we assume that f' is continuous we do have that if f is monotone increasing its derivative cannot be negative anywhere, as otherwise we would have an interval in which f' is negative, meaning f would be monotone decreasing on that interval.

Theorem 18. Let f be continuous at a, differentiable on some punctured neighborhood of a and such that the limit $\lim_{x \to a} f'(x)$ exists.

Then we have f is differentiable at a and f' is continuous at a.

Proof. Let $\lim_{x\to a} f'(x) = \alpha$ and consider an arbitrary $\epsilon > 0$.

For sufficiently small $h > 0$ we have by the Mean Value Theorem that

$$\frac{f(a+h)-f(a)}{h} = f'(x).$$

For some $x \in (a, a+h)$. So if h is sufficiently small we have

$$|f'(x) - \alpha| < \epsilon$$
$$|\frac{f(a+h)-f(a)}{h} - \alpha| < \epsilon$$

Using similar logic for $h < 0$ we have that f is differentiable at a with derivative α. $\square$

Theorem 19 (Cauchy Mean Value Theorem)**.** Let f and g be continuous on $[a, b]$ and differentiable on (a, b). Then we have

$$(f(b) - f(a))g'(x) = (g(b) - g(a))f'(x).$$

For some $x \in (a, b)$.

Proof. Define ϕ on $[a, b]$ by

$$\phi = (g(b) - g(a))f - (f(b) - f(a))g.$$

Then we have $\phi(a) = \phi(b) = f(a)g(b) - f(b)g(a)$.

Hence by Rolle's Theorem we have $x \in (a, b)$ such that

$$\phi'(x) = 0$$
$$(f(b) - f(a))g'(x) = (g(b) - g(a))f'(x).$$

$\square$

Theorem 20 (L'Hopitals Rule)**.** Let f and g be continuous at a with $f(a) = g(a) = 0$ and such that $\lim_{x \to a}(\frac{f'}{g'})(x)$ exists.

Then $\lim_{x \to a}(\frac{f}{g})(x)$ exists and is equal to $\lim_{x \to a}(\frac{f'}{g'})(x)$.

Proof. The existence of the limit means that f and g are differentiable on some open interval of a and g' is non-vanishing on this interval. If g vanished more than once on the interval then by Rolle's Theorem g' would vanish on the interval which cannot be. So g vanishes at most once, so we may consider a smaller open interval of a on which g is also non-vanishing.

So by the Cauchy Mean Value Theorem, we have for sufficiently small $h > 0$ that

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

For some $x \in (a, a+h)$.

Now consider an arbitrary $\epsilon > 0$. If we let $\lim_{x \to a}(\frac{f'}{g'})(x) = \alpha$, then for sufficiently small $h > 0$ we have

$$|\frac{f'(x)}{g'(x)} - \alpha| < \epsilon$$
$$|\frac{f(a+h)}{g(a+h)} - \alpha| < \epsilon.$$

Using similar logic applies for the case of small $h < 0$, we have the desired result. ∎

2.4 Inverse Functions

Theorem 21. Let f be continuous and injective on an interval I. Then f is monotone on I.

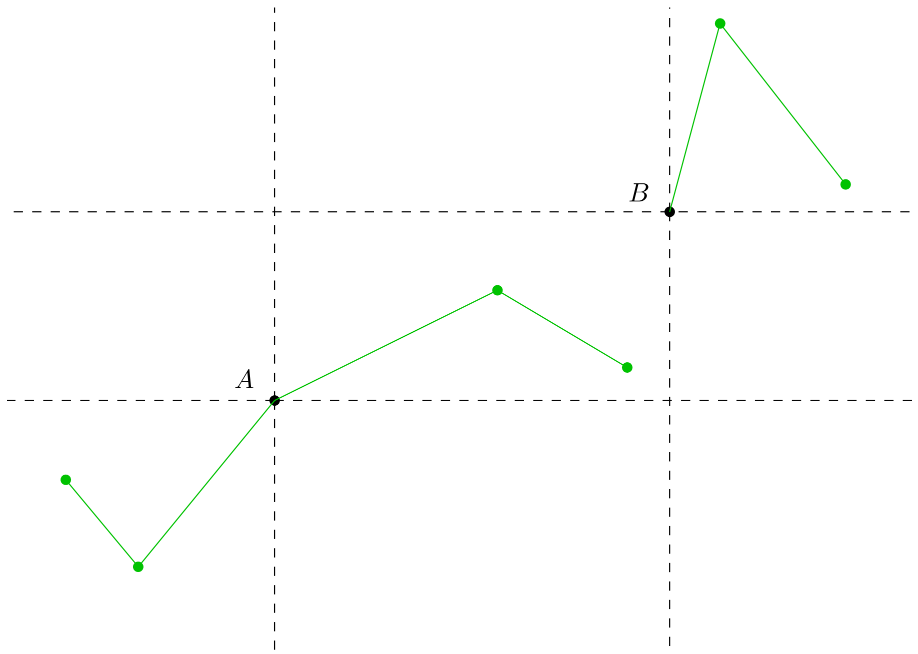

Figure 6

Proof. We proceed by contradiction. Assume we have $a, b, c, d \in I$ such that

$$\begin{aligned} a &< b \\ f(a) &< f(b) \\ c &< d \\ f(c) &> f(d). \end{aligned}$$

We let $A(a, f(a))$ and define the points B, C and D similarly.

Now consider Figure 6. We may divide the plane into nine regions by the lines shown. We need not concern ourselves with the boundaries of the regions, as C and D cannot lie on any boundary as f is an injective function.

We claim that of A, B, C, D there are 3 points X, Y, Z such that $x < y < z$ and $f(y)$ is greater than both of $f(x)$ and $f(z)$ or $f(y)$ is lesser than both of them. Visually this corresponds to three points making a "V" or a "Λ" shape.

If either of C or D lies in any of the regions not on the upper diagonal (the ones with no annotations in the diagram) then we clearly have the desired configuration. Otherwise C and D both lie in one of the diagonal regions. As $c < d$ and $f(c) > f(d)$ they cannot lie in distinct diagonal regions, so we need them in the same region. As shown in the diagram, this also always leads to the desired configuration of points.

This configuration then contradicts injectivity by the Intermediate Value Theorem. □

Indeed, as we have only used the intermediate value property, this result holds for all Darboux functions.

Theorem 22. *Let f be a continuous injection on an interval I. Then f^{-1} is also continuous.*

Proof. As f is continuous, by the Intermediate Value Theorem we have that $f(I)$ is an interval. As f is a continuous injection over an interval, we also have that f is monotone, and so f^{-1} is monotone as well.

For the sake of contradiction, assume that f^{-1} is discontinuous at $a \in f(I)$. Then for some $\epsilon > 0$, for every $\delta > 0$, we have a $|x - a| < \delta$ such that $|f^{-1}(x) - f^{-1}(a)| \geq \epsilon$. As f^{-1} is monotone, this means that on at least one side of a

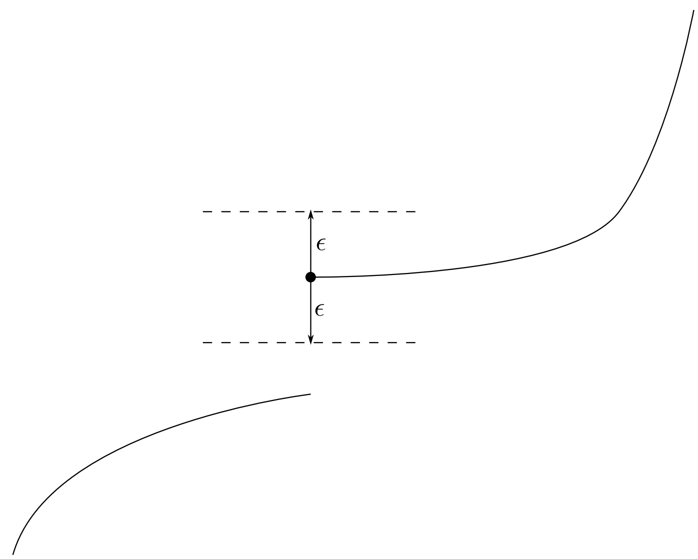

Figure 7: Illustration of a typical monotone discontinuity. Unlike for an arbitrary function, discontinuity can only arise when there is some lower bound on how close the function can be to specific value from one side.

there are no points with $|f^{-1}(x) - f^{-1}(a)| < \epsilon$, as in Figure 7. However this contradicts surjectivity of $f^{-1} : f(I) \to I$. $\square$

Theorem 23. Let f be a continuous injection on the interval I which is differentiable at $a \in I$ with non-vanishing derivative. Then f^{-1} is differentiable at $f(a)$ with derivative

$$(f^{-1})'(f(a)) = \frac{1}{f'(a)}.$$

Proof. The difference quotient we are interested in is

$$\frac{f^{-1}(f(a)+h) - a}{h}.$$

To this end we define $k(h) = f^{-1}(f(a) + h) - a$ for sufficiently small h so that we have

$$f(a) + h = f(a + k(h)).$$

As f is continuous, we have that f^{-1} is continuous, so k is continuous.

We have the difference quotient is

$$\frac{k(h)}{f(a + k(h)) - f(a)}.$$

By the injectivity of f we have that k is non-vanishing everywhere apart from 0. Hence we have the difference quotient as

$$\frac{1}{\left(\frac{f(a+k(h)-f(a)}{k(h)}\right)}.$$

We have that $k(0) = 0$, so by the continuity of k for sufficiently small h, $k(h)$ is arbitrarily small. Hence the difference quotient approaches $\frac{1}{f'(a)}$ as h tends to 0. $\square$

3 Integrals

3.1 The Integral

The integral can be thought of as a sort of "continuous summation" of a function over an interval. Visually this corresponds to the signed area between the graph of the function and the x axis.

Broadly, we may make this rigorous by noting that if we draw rectangles which together are contained in the region or contain the region (such as in Figure 8), then the "area" of the region must be between these values. From this observation, the following definitions are obviously motivated.

Definition 8 (Partition)**.** We define a partition P of $[a, b]$ to be a finite subset of $[a, b]$ containing a and b. We will label the elements as

$$P = \{t_0, t_1, \ldots, t_n\}$$

Where $a = t_0 < t_1 < t_2 < \ldots < t_n = b$.

We say a partition is finer than another if it contains it.

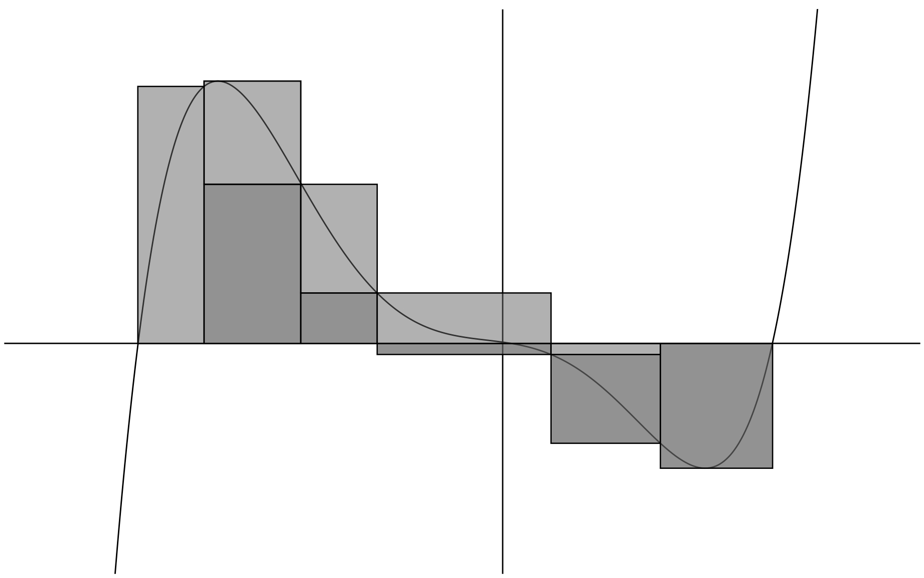

Figure 8

Definition 9 (Upper and Lower Sum)**.** Let f be bounded on $[a, b]$ and P a partition of $[a, b]$. Then we define the upper and lower sums as

$$U(f, P) = \sum_{i=1}^{n} (t_i - t_{i-1}) M_i$$
$$L(f, P) = \sum_{i=1}^{n} (t_i - t_{i-1}) m_i.$$

Where we define

$$m_i = \inf f[t_{i-1}, t_i]$$
$$M_i = \sup f[t_{i-1}, t_i].$$

Theorem 24. Let f be bounded on $[a, b]$ and P, P' arbitrary partitions of $[a, b]$.

1. If P' is finer than P, then $L(f, P') \geq L(f, P)$ and $U(f, P') \leq U(f, P)$.
2. $U(f, P) \geq L(f, P')$.

Proof. The first point follows trivially from the fact that the supremum of a subset cannot exceed the supremum of the original set and similarly for the infimum.

We now show the second point. We note that $P \cup P'$ is finer than P and P', so we have

$$\begin{aligned} U(f, P) &\geq U(f, P \cup P') \\ &\geq L(f, P \cup P') \\ &\geq L(f, P'). \end{aligned}$$

□

As a consequence of this, we have that

$$\sup\{L(f, P)\} \leq \inf\{U(f, P)\}.$$

Clearly the only reasonable case in which we may say there is an integral is when the infimum and supremum coincide, as we may then allow the integral to be this common value.

Definition 10 (Integral)**.** Let f be bounded on $[a, b]$. Then we say f is integrable on $[a, b]$ if

$$\sup\{L(f, P)\} = \inf\{U(f, P)\}.$$

In this case, we say the integral, $\int_a^b f$, is the common value of the infimum and supremum.

We also define the integral

$$\int_b^a f = -\int_a^b f.$$

We now provide an alternative characterisation of integrability. This has little depth and arises from elementary reasoning about infimums and supremums, however it lends itself much more naturally to proofs.

Theorem 25. Let f be bounded on $[a, b]$. Then the following are equivalent

1. f is integrable on $[a, b]$.

2. For every $\epsilon > 0$ we have a partition P of $[a, b]$ such that

$$U(f, P) - L(f, P) < \epsilon.$$

Proof. First assume that f is integrable. Then as $\int_a^b f$ is an infimum and supremum, for every $\epsilon > 0$ we have a partition P of $[a, b]$ such that

$$\int_a^b f - L(f, P) < \frac{\epsilon}{2}$$
$$U(f, P) - \int_a^b f < \frac{\epsilon}{2}.$$

Summing yields

$$U(f,P) - L(f,P) < \epsilon.$$

Now assume the latter hypothesis. Assume for the sake of contradiction that f was not integrable, so we have

$$\sup\{L(f,P)\} < \inf\{U(f,P)\}.$$

Let $\epsilon > 0$ be the difference between the two quantities. Then we have for any partition P of $[a,b]$ that

$$\begin{aligned} U(f,P) - L(f,P) &\geq \inf\{U(f,P)\} - \sup\{L(f,P)\} \\ &= \epsilon. \end{aligned}$$

This contradicts the latter hypothesis. □

We now proceed to prove some fundemental results about integrals that we would expect to be true from their geometrical interpretation.

Theorem 26. Let f be bounded on $[a,b]$ and $a < c < b$.

Then f is integrable on $[a,b]$ precisely when f is integrable on $[a,c]$ and $[c,b]$, and when f is integrable on $[a,b]$ we have

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

Proof. First assume f is integrable on $[a,b]$. Then for any $\epsilon > 0$ we have a partition P of $[a,b]$ such that

$$U(f,P) - L(f,P) < \epsilon.$$

We may assume WLOG that $c \in P$ as otherwise $P \cup \{c\}$ is finer and the inequality still holds. So we may split up P into partitions P_1 and P_2 of $[a, c]$ and $[c, b]$, respectively, so we have

$$(U(f, P_1) + U(f, P_2)) - (L(f, P_1) + L(f, P_2)) < \epsilon$$
$$(U(f, P_1) - L(f, P_1)) + (U(f, P_2) - L(f, P_2)) < \epsilon.$$

Each term in parentheses is nonnegative, so we have

$$U(f, P_1) - L(f, P_1) < \epsilon$$
$$U(f, P_2) - L(f, P_2) < \epsilon.$$

Hence f is integrable on $[a, c]$ and $[c, b]$.

Now assume f is integrable on $[a, c]$ and $[c, b]$. Then for all $\epsilon > 0$, we have partitions P_1 and P_2 of $[a, c]$ and $[c, b]$ resptively such that

$$U(f, P_1) - L(f, P_1) < \frac{\epsilon}{2}$$
$$U(f, P_2) - L(f, P_2) < \frac{\epsilon}{2}.$$

Summing the inequalities yields

$$\begin{aligned}(U(f, P_1) + U(f, P_2)) - (L(f, P_1) + L(f, P_2)) &< \epsilon \\ U(f, P) - L(f, P) &< \epsilon.\end{aligned}$$

Hence f is integrable on $[a, b]$.

Now assume that f is integrable on $[a, b]$. Then for arbitrary partitions P_1

and P_2 of $[a,c]$ and $[c,b]$, respectively we have

$$L(f,P_1) \leq \int_a^c f \leq U(f,P_1)$$
$$L(f,P_2) \leq \int_c^b f \leq U(f,P_2).$$

We note that $P = P_1 \cup P_2$ is also an arbitrary partition of $[a,b]$. Then summing the inequalities yields

$$L(f,P) \leq \int_a^c f + \int_c^b f \leq U(f,P).$$

The only value which satisfies this inequality is $\int_a^b f$, hence we have the desired equality

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

□

Theorem 27. Let f and g be integrable on $[a,b]$, then $f+g$ is integrable on $[a,b]$ and

$$\int_a^b (f+g) = \int_a^b f + \int_a^b g.$$

Proof. Consider an arbitrary $\epsilon > 0$. As f and g are integrable on $[a,b]$ we have partitions of $[a,b]$ which satisfy an inequality of the upper and lower sums. Taking their union we have a partition P of $[a,b]$ such that

$$U(f,P) - L(f,P) < \frac{\epsilon}{2}$$
$$U(g,P) - L(g,P) < \frac{\epsilon}{2}.$$

We also note that

$$
\begin{aligned}
&U(f,P)+U(g,P) \\
&= \sum_{i=1}^{n}(t_i - t_{i-1})(\sup f[t_i - t_{i-1}] + \sup g[t_i - t_{i-1}]) \\
&\geq \sum_{i=1}^{n}(t_i - t_{i-1})(\sup(f+g)[t_i - t_{i-1}]) \\
&= U(f+g,P).
\end{aligned}
$$

By similar logic we also have

$$L(f,P)+L(g,P) \leq L(f+g,P).$$

Subtracting inequalities yields

$$
\begin{aligned}
&U(f+g,P)-L(f+g,P) \\
&\leq (U(f,P)-L(f,P))+(U(g,P)-L(g,P)) \\
&< \epsilon.
\end{aligned}
$$

Hence $f+g$ is integrable on $[a,b]$.

We have for an arbitrary partition P of $[a,b]$ that

$$
\begin{aligned}
L(f+g,P) \leq \int_a^b (f+g) \leq U(f+g,P) \\
L(f,P)+L(g,P) \leq \int_a^b (f+g) \leq U(f,P)+U(g,P).
\end{aligned}
$$

This inequality is satisfied by a unique quantity, hence we have

$$\int_a^b (f+g) = \int_a^b f + \int_a^b g.$$

$\square$

Theorem 28. Let f be continuous on $[a, b]$. Then f is integrable on $[a, b]$.

Proof. As f is continuous on a closed bounded interval we have that f is uniformly continuous. So take an arbitrary $\epsilon > 0$. We then have a $\delta > 0$ such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon.$$

For x, y in $[a, b]$. We now choose a partition P of $[a, b]$ into n equal subintervals such that $\frac{1}{n} < \delta$. So then we have

$$\begin{aligned} &U(f, P) - L(f, P) \\ &= \frac{1}{n}\sum_{i=1}^{n}(M_i - m_i) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\epsilon. = \epsilon. \end{aligned}$$

Hence f is integrable on $[a, b]$. $\square$

Theorem 29. Let f be integrable on $[a, b]$ and define F on $[a, b]$ by

$$F(x) = \int_a^x f.$$

Then F is continuous on $[a, b]$.

Proof. By integrability, we have m and M such that

$$m < f(x) < M.$$

For all x in $[a, b]$.

Consider an arbitrary $x \in [a, b]$ and an arbitraty $\epsilon > 0$. We will consider only when x is not on the boundary of the interval as the boundary cases use similar logic. We have for all h sufficiently small so that $x + h \in [a, b]$ that

$$\begin{aligned} |F(x+h) - F(x)| \\ = |\int_x^{x+h} f| \\ < h(|m| + |M|) \end{aligned}$$

Hence F is continuous at x. □

3.2 The Fundemental Theorem of Calculus

Noting that the derivative is intuitively the notion of a rate of change and the integral is the notion of "summing continuously", it is reasonable to think that under certain conditions the derivative of an integral is the function being integrated and the integral of a derivative is the total change between two points.

We formalise this intuition in the Fundemental Theorem of Calculus.

Theorem 30 (First Fundemental Theorem of Calculus)**.** Let f be integrable on $[a, b]$ and F defined on $[a, b]$ by

$$F(x) = \int_a^x f.$$

Then if f is continuous at $c \in (a, b)$ we have that F is differentiable at c with

$$F'(c) = f(c).$$

Proof. Consider an arbitrary $\epsilon > 0$. Theb by continuity of f at c we have a

$\delta > 0$ such that

$$|x - c| < \delta \implies |f(x) - f(c)| < \epsilon.$$

So for small $h > 0$ we have

$$\begin{aligned} h(f(c) - \epsilon) < \int_c^{c+h} f < h(f(c) + \epsilon) \\ f(c) - \epsilon < \frac{1}{h} \int_c^{c+h} f < f(c) + \epsilon \\ \left| \frac{1}{h} \int_c^{c+h} f - f(c) \right| < \epsilon. \end{aligned}$$

Similar logic yields the same inequality for small $h < 0$. Hence F is differentiable at c with derivative $f(c)$. □

Theorem 31 (Second Fundemental Theorem of Calculus)**.** Let g be differentiable on (a, b) and continuous on $[a, b]$, and g' be integrable on $[a, b]$.

Then we have

$$\int_a^b g' = g(b) - g(a).$$

Proof. Consider an arbitrary partition P of $[a, b]$. By the Mean Value Theorem, for each $1 \leq i \leq n$ we have $x_0 \in (t_{i-1}, t_i)$ such that

$$g'(x_0) = \frac{g(t_i) - g(t_{i-1}}{t_i - t_{i-1}}.$$

We have

$$
\begin{aligned}
&L(g', P) \\
&= \sum_{i=1}^{n}(t_i - t_{i-1}) \inf g'[t_{i-1}, t_i] \\
&\leq \sum_{i=1}^{n}(t_i - t_{i-1}) \frac{g(t_i) - g(t_{i-1})}{t_i - t_{i-1}} \\
&= \sum_{i=1}^{n}(g(t_i) - g(t_{i-1})) \\
&= g(b) - g(a).
\end{aligned}
$$

By similar logic we also have

$$L(g', P) \leq g(b) - g(a) \leq U(g', P).$$

Hence we have the desired equality. □

4 The Logarithm and Exponential

We now define the logarithm and exponential functions. The key motivating characteristic is that they provide homomorphisms from the additive group of reals to the mulitplicative group of positive reals.

Definition 11 (Logarithm)**.** We define the logarithm function log on $\mathbb{R}^+$ by

$$\log(x) = \int_1^x \frac{dx}{x}.$$

Alongside being a homomorphism, we have several other useful properties of the logarithm which we provide below.

Theorem 32. The logarithm function satisfies the following:

1. $\log(xy) = \log(x) + \log(y)$ for any positive x and y.

2. log is an increasing bijection from $\mathbb{R}^+$ to $\mathbb{R}$.

Proof. Consider an arbitrary partition P of $[1, x]$. This corresponds to a arbitrary partition P' of $[y, xy]$ by multiplication by y.

We have

$$\begin{aligned} & L(f, P') \\ = & \sum_{i=1}^{n} (kt_i - kt_{i-1}) \frac{1}{kt_i} \\ = & \sum_{i=1}^{n} (t_i - t_{i-1}) \frac{1}{t_i} \\ = & L(f, P). \end{aligned}$$

Similar logic gives the same equality between the upper sums. Hence we have

$$\begin{aligned} & \log(x) + \log(y) \\ = & \int_1^x \frac{dt}{t} + \int_1^y \frac{dt}{t} \\ = & \int_y^{xy} \frac{dt}{t} + \int_1^y \frac{dt}{t} \\ = & \int_1^{xy} \frac{dt}{t} \\ = & \log(xy). \end{aligned}$$

We have

$$\log'(x) = \frac{1}{x} > 0.$$

Hence log is increasing and hence injective.

We have that $\log(2) > 0$ and

$$\log(2^n) = n\log(2)$$
$$\log(\frac{1}{2^n}) = -n\log(2).$$

So log is bounded neither above nor below. And by continuity this means that log is surjective. $\square$

In light of this we may make the following definition

Definition 12 (Exponential)**.** We define the exponential function $\exp = \log^{-1}$.

We also define Euler's Constant $e = \exp(1)$.

From our results on the logarithm we have that exp is defined on all of $\mathbb{R}$, is increasing, takes on every positive real and most importantly satisfies

$$\exp(x+y) = \exp(x)\exp(y).$$

For every real x and y.

In light of this we may now define exponentiation for arbitrary real exponents.

Definition 13 (Exponentiation)**.** Let $a > 0$ and x an arbitrary real. Then we define

$$a^x = \exp(x\log a).$$

Theorem 33 (Index Laws)**.** Let $a > 0$ and x, y arbitrary reals. Then we have

1.

$$a^{x+y} = a^x a^y.$$

2.

$$(a^x)^y = a^{xy}.$$

Proof. We have

$$\begin{aligned} a^{x+y} &= \exp(x \log a + y \log a) \\ &= \exp(x \log a) \exp(y \log a) \\ &= a^x a^y. \end{aligned}$$

We also have

$$\begin{aligned} (a^x)^y &= (\exp(x \log a))^y \\ &= \exp(y \log \circ \exp(x \log a)) \\ &= \exp(xy \log a) \\ &= a^{xy}. \end{aligned}$$

□

As a consequence of the index laws, we see our definition of exponentiation coincides with the definition for rational exponents. Indeed as every real is arbitrary well approximated by rationals, our definition is the only possible continuous extension of exponentiation to real exponents.

5 Sequences and Series

5.1 Taylor Polynomials

An alternative characterisation of the tangent line to f at a is as the degree 1 polynomial whose zeroeth and first derivative agree with f at a. This of course lends itself to a natural generalisation.

Definition 14 (Taylor Polynomial)**.** Let f be n times diffrentiable at a. Then we define the Taylor polynomial of degree n at a for f, $P_{n,a}$, as the degree n polynomial such that

$$P_{n,a}^{(k)}(a) = f^{(k)}(a)$$

For all $0 \leq k \leq n$.

The definition assumes implicitly that such a polynomial both exists and is unique. However we may trivially show this constructively by showing

$$P_{n,a}(x) = \sum_{j=0}^{n} a_j (x-a)^j.$$

Where

$$a_k = \frac{f_{n,a}^{(k)}(a)}{k!}.$$

The significance of this sequence of polynomials is that they provide extremely good local approximations of f at a. We make the way in which we measure the quality of an approximation in this context precise with the following definition:

Definition 15 (Equal up to Order)**.** We say f and g are equal up to order n

at a if

$$\lim_{x \to a} \frac{f(x) - g(x)}{(x-a)^n} = 0.$$

We see that equalty up to order m is stronger than equality up to order n when $m > n$ as we would intuitively expect. It is also trivial that equality up to order n is an equvialence relation.

Theorem 34. Let f be n times diffrentiable at a. Then f is equal to $P_{n,a}$ up to order n at a.

Proof. We have

$$\begin{aligned} &\lim_{x \to a} \frac{f(x) - P_{n,a}(x)}{(x-a)^n} \\ &= \lim_{x \to a} (x-a)^{-n} (f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x-a)^k) - \frac{f^{(n)}(a)}{n!}. \end{aligned}$$

Repeated applications of l'hopitals rule yields

$$\begin{aligned} \lim_{x \to a} (x-a)^{-n} (f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x-a)^k) \\ = \lim_{x \to a} \frac{f^{(n-1)}(x) - f^{(n-1)}(a)}{n!(x-a)} \\ = \frac{f^{(n)}(a)}{n!} \end{aligned}$$

Hence

$$\lim_{x \to a} \frac{f(x) - P_{n,a}(x)}{(x-a)^n} = 0.$$

□

Indeed as the next result shows, this property alongside $P_{n,a}$ being a polynomial of degree n is actually an alternate characterisation of the Taylor polynomials.

Theorem 35. Let f and g be polynomials of degree n. Then if they are equal up to order n anywhere, $f = g$.

Proof. Let $p = f - g$. Then p is equal up to order n to 0 at some a. Let

$$p(x) = \sum_{k=0}^{n} b_k (x-a)^k.$$

Assume by induction that $b_k = 0$ for $k \leq j - 1 < n$. Then we have

$$p(x) = (x-a)^j \sum_{k=0}^{n-j} b_{k+j} (x-a)^k$$

As p is equal to 0 up to order $j \leq n$ we have

$$\begin{aligned} \lim_{x \to a} \sum_{k=0}^{n-j} b_{k+j} (x-a)^k &= 0 \\ b_j &= 0. \end{aligned}$$

Hence $p = 0$ and so $f = g$. $\square$

Whilst it is helpful to know in an asymptotic sense how well the Taylor Polynomials approximate a function, we now derive some formulae to estimate and bound the error.

Definition 16 (Remainder)**.** Let f be n times diffrentiable at a. Then we define the remainder for the degree n Taylor polynomial at a as

$$R_{n,a} = f - P_{n,a}.$$

Theorem 36. Let f be $n+1$ times diffrentiable on $[a,x]$. Then we have

$$R_{n,a}(x) = \frac{f^{(n+1)}(t)}{n!}(x-t)^n(x-a)$$
$$R_{n,a}(x) = \frac{f^{(n+1)}(t)}{(n+1)!}(x-a)^{n+1}$$

For some $t \in (a,x)$.

If we further assume that $f^{(n+1)}$ is integrable on $[a,x]$ then

$$R_{n,a}(x) = \int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n dt.$$

For some $t \in (a,x)$.

We call these expressions for the remainder the Cauchy, Lagrange and Integral form, respectively.

Proof. We define S on $[a,x]$ by

$$S(t) = R_{n,t}(x).$$

So for any $t \in [a,x]$ we have

$$S(t) = -P_{n,t}(x) + f(x)$$
$$-S'(t) = \sum_{k=0}^{n} \left(\frac{f^{(k+1)}(t)}{k!}(x-t)^k - \frac{kf^{(k)}(t)}{k!}(x-t)^{k-1} \right)$$
$$-S'(t) = \sum_{k=0}^{n} \frac{f^{(k+1)}(t)}{k!}(x-t)^k - \sum_{k=0}^{n-1} \frac{f^{(k+1)}(t)}{k!}(x-t)^k$$
$$S'(t) = -\frac{f^{(n+1)}(t)}{n!}(x-t)^n.$$

By the Mean Value Theorem we have for some $t \in (a, x)$ that

$$\begin{aligned} S'(t) &= \frac{S(x) - S(a)}{x - a} \\ R_{n,a}(x) &= \frac{f^{(n+1)}(t)}{n!}(x-t)^n(x-a). \end{aligned}$$

Let $g(t) = (x-t)^{n+1}$. Then by the Cauchy Mean Value Theorem we have $t \in (a, x)$ such that

$$\begin{aligned} \frac{S'(t)}{g'(t)} &= \frac{S(x) - S(a)}{g(x) - g(a)} \\ R_{n,a}(x) &= \frac{f^{(n+1)}(t)}{(n+1)!}(x-a)^{(n+1)}. \end{aligned}$$

Now assume that $f^{(n+1)}$ is integrable on $[a, x]$. Then we have

$$\begin{aligned} \int_a^x S' &= \int_a^x -\frac{f^{(n+1)}(t)}{n!}(x-t)^n dt \\ R_{n,a}(x) &= \int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n dt. \end{aligned}$$

□

5.2 Taylor Polynomials of Elementary Functions

We now derive the Taylor polynomials and upper bounds for the remainder for various important elementary functions.

In the case of log, exp, sin and cos the derivatives follow a simple pattern and so we may just apply the techniques from the previous subsection.

Repeated diffrentiation of arctan does not follow a simple pattern, however we may use the alternative characterisation of the Taylor polynomials as being equal up to order n to the function.

We have from the formula for geometric series that

$$\frac{1}{1+x^2} = \sum_{k=0}^{n} (-1)^k x^{2k} + \frac{(-1)^{n+1} x^{2n+2}}{1+x^2}.$$

Integrating from 0 to x yields

$$\arctan(x) = \sum_{k=0}^{n} \frac{(-1)^k}{2k+1} x^{2k+1} + (-1)^{n+1} \int_0^x \frac{t^{2n+2}}{1+t^2} dt.$$

We now show that arctan is equal to this polynomial up to order $2n+1$ which shows it is indeed the Taylor polynomial. We have

$$\begin{aligned} \left| \frac{1}{x^{2n+1}} (\arctan(x) - \sum_{k=0}^{n} \frac{(-1)^k}{2k+1} x^{2k+1}) \right| & \\ = \frac{1}{|x|^{2n+1}} \int_0^{|x|} \frac{t^{2n+2}}{1+t^2} dt & \\ < \frac{1}{|x|^{2n+1}} \int_0^{|x|} t^{2n+2} dt & \\ = \frac{1}{2n+3} |x|^2. & \end{aligned}$$

Hence they are indeed equal up to order n at 0.

We now list the Taylor polynomials and the remainder terms of various elementary functions

$$\sin(x) = \sum_{k=0}^{n} \frac{(-1)^{k+1}}{(2k+1)!} x^{2k+1} + \frac{(-1)^{n+1}}{(2n+1)!} \int_0^x (x-t)^{2n+1} \sin(t) dt$$

$$\cos(x) = \sum_{k=0}^{n} \frac{(-1)^k}{(2k)!} x^{2k} + \frac{(-1)^{n+1}}{(2n)!} \int_0^x (x-t)^{2n} \sin(t) dt$$

$$\arctan(x) = \sum_{k=0}^{n} \frac{(-1)^k}{2k+1} x^{2k+1} + (-1)^{n+1} \int_0^x \frac{t^{2n+2}}{1+t^2} dt$$

$$\exp(x) = \sum_{k=0}^{n} \frac{1}{k!} x^k + \frac{1}{n!} \int_0^x e^t (x-t)^n dt$$

$$\log(1+x) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} x^k + (-1)^n \int_0^x \frac{(x-t)^n}{(t+1)^{n+1}} dt.$$

5.3 Sequences

We now define the notion of a sequence and various intuitive properties. Most of these ideas are almost identical to limits of functions.

Definition 17 (Sequence)**.** For an arbitrary set A we define a sequence of elements of A as a function $x : \mathbb{N} \to A$. We write

$$x_n = x(n).$$

In particular a sequence of reals is a mapping $x : \mathbb{N} \to \mathbb{R}$.

Definition 18 (Cluster Point)**.** Let (x_n) be a sequence of reals. Then we say α is a cluster point of (x_n) if for every $\epsilon > 0$ we have an n such that

$$0 < |x_n - \alpha| < \epsilon.$$

Definition 19 (Limit)**.** Let (x_n) be a sequence of reals. Then we say x is the limit of (x_n) if for every $\epsilon > 0$ we have an N such that

$$n \geq N \implies |x_n - x| < \epsilon.$$

Of course the limit is a cluster point however the converse need not hold.

Theorem 37 (Uniqueness of Limits)**.** The limit of a sequence (x_n), x, when it exists, is unique.

Proof. Assume the contrary and say (x_n) approaches both x and y where $x \neq y$. WLOG assume $x < y$.

Then take $\epsilon = \frac{y-x}{2} > 0$. Then for sufficiently large n we have

$$|x_n - x| < \epsilon$$
$$|x_n - y| < \epsilon.$$

So we have

$$x_n < \frac{x+y}{2}$$
$$x_n > \frac{x+y}{2}$$

Which is the desired contradiction. $\square$

Theorem 38 (Monotone Convergence Theorem)**.** Let (x_n) be bounded above and weakly increasing. Then (x_n) is convergent.

Proof. As (x_n) is bounded above we may take

$$\alpha = \sup(x_n).$$

We claim that α is the limit of (x_n). Say this was not the case. Then we have some $\epsilon > 0$ such that we have arbitrarily large n such that

$$|x_n - \alpha| \geq \epsilon.$$

As (x_n) is increasing this means that for every n we have

$$x_n \leq \alpha - \epsilon.$$

Which contradicts α being the least upper bound. □

Of course the analogous result holds for weakly decreasing sequences bounded below.

Theorem 39 (Bolzano-Weierstrass Theorem)**.** Let (x_n) be bounded. Then (x_n) has a cluster point.

Proof. We will call n a peak if

$$m \geq n \implies x_m \leq x_n.$$

If (x_n) has a infinity of peaks then the sequence of peaks forms a weakly decreasing subsequence which converges to some limit, which will be a cluster point of (x_n).

Otherwise we have that for sufficiently large n there are no peaks. Then we may construct a weakly increasing subsequence, as for every n we have an $m > n$ such that $x_m > x_n$, which converges to some limit which will be a cluster point of (x_n). $\square$

Theorem 40 (Sequential Continuity)**.** The function f is continuous at a precisely when for every sequence (x_n) in the the domain of f converging to a we have

$$\lim_{n\to\infty} f(x_n) = f(a).$$

Proof. First assume that f is continuous at a. Consider an arbitrary sequence $(x_n) \to a$. Then for any $\epsilon > 0$ we have an N and $\delta > 0$ such that

$$n \geq N \implies |x_n - a| < \delta$$
$$|x - a| < \delta \implies |f(x) - f(a)| < \epsilon.$$

So we have

$$n \geq N \implies |f(x_n) - f(a)| < \epsilon.$$

Hence $(f(x_n)) \to f(a)$.

Now assume the latter hypothesis. Assume for contradiction that f is discontinuous at a. Then we have an $\epsilon > 0$ such that for every n we have an x_n such that

$$|x_n - a| < \frac{1}{n}$$
$$|f(x_n) - f(a)| \geq \epsilon.$$

The sequence (x_n) then converges to a but $(f(x_n))$ cannot converge to $f(a)$, a contradiction. □

In particular this means that if a function is continuous everywhere we have

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

We now provide an alternative characterisation of convergence of a sequence. The main benefit is that it doesn't require us to actually use the limit of the sequence.

Definition 20 (Cauchy Sequence)**.** Let (x_n) be a real sequence. Then (x_n) is Cauchy if for any $\epsilon > 0$ we have N such that

$$m, n \geq N \implies |a_m - a_n| < \epsilon.$$

Theorem 41 (Cauchy Criterion)**.** The sequence (x_n) is convergent precisely when it is Cauchy.

Proof. First assume (x_n) is convergent. So we have it converges to some α. Now take an arbitrary $\epsilon > 0$. We have an N such that

$$n \geq N \implies |x_n - \alpha| < \frac{\epsilon}{2}.$$

So we have for $m, n \geq N$ that

$$|x_n - \alpha| < \frac{\epsilon}{2}$$
$$|x_m - \alpha| < \frac{\epsilon}{2}.$$

Summing and applying the triangle inequality yields

$$|x_n - x_m| < \epsilon.$$

Now assume the sequence is Cauchy. As a consequence it must be bounded. So by the Bolzano-Weierstrass Theorem we have a cluster point, α. We claim $(x_n) \to \alpha$. Take an arbitrary $\epsilon > 0$.

By the Cauchy property we have an N such that

$$m, n \geq N \implies |a_m - a_n| < \epsilon.$$

As α is a limit point we have $k \geq N$ such that

$$|a_k - \alpha| < \epsilon.$$

Summing and applying the triangle inquality yields

$$n \geq N \implies |a_n - \alpha| < 2\epsilon.$$

□

5.4 Series

Definition 21 (Series)**.** We define a series as a sequence, which we write as

$$\sum_{k=0}^{\infty} x_n$$

to emphasise how we are considering it.

We say the series converges if the sequence of partial sums converges.

We now derive some useful tests for convergence.

Theorem 42 (Comparison Test)**.** Let the sequences (x_n) and (y_n) satisfy

$$0 \le x_n \le y_n$$

For all n. Then if $\sum y_n$ converges so does $\sum x_n$.

Proof. As $\sum y_n$ is convergent, its partial sums are bounded above and so the partial sums of $\sum x_n$ are also bounded above. As (x_n) is nonnegative, the sequence of partial sums is weakly increasing and so by the monotone convergence theorem the series $\sum x_n$ converges. □

Theorem 43 (Ratio Test)**.** Let $x_n > 0$ for all n, and let

$$\lim_{n \to \infty} \frac{x_{n+1}}{x_n} = r.$$

Then if $r < 1$ the series $\sum x_n$ converges and if $r > 1$ the series $\sum x_n$ diverges.

Proof. First assume $r > 1$. Take $\epsilon = \frac{r-1}{2} > 0$. Then we have an N such that

$$n \ge N \implies \frac{x_{n+1}}{x_n} \ge \frac{r+1}{2}.$$

Hence we have

$$x_{N+n} \ge x_N \left(\frac{r+1}{2}\right)^n$$

This means the terms do not tend to vanish, meaning the series cannot converge.

The case of $r < 1$ uses similar logic. □

Theorem 44 (Integral Test)**.** Let f be nonnegative and weakly decreasing on $[1, \infty)$.

Then the series $\sum f(n)$ converges precisely when the integral $\int_1^\infty f$ exists.

Proof. First assume the integral exists. Then the sequence $(\int_1^n f)$ is bounded above. We have by monotonicty that

$$\begin{aligned}\int_1^n f &= \sum_{k=1}^{n-1} \int_k^{k+1} f \\ &\geq \sum_{k=1}^{n-1} f(k+1).\end{aligned}$$

Hence the partial sums are bounded above and so the series converges.

Now assume that the series converges. By monotonicty $\int_1^x f$ exists for all $x \geq 0$ and the integral is weakly increasing with x as f is nonnegative. Let n be an integer larger than x. By montonicity we have

$$\begin{aligned}&\int_1^x f \\ &\leq \int_1^n f \\ &= \sum_{k=1}^{n-1} \int_k^{k+1} f \\ &\leq \sum_{k=1}^{n-1} f(k)\end{aligned}$$

Which is bounded above by convergence. Hence the integral exists. $\square$