MA2201: Analysis II

Differentiation

Spring 2021

Satvik Saha
19MS154

*Indian Institute of Science Education and Research, Kolkata,
Mohanpur, West Bengal, 741246, India.*

The origins of differential calculus lie in our attempts to approximate various functions using linear ones. Suppose that we have been given a curve described by the function f, and we want to *locally* approximate the function around a point x using a straight line. In other words, for a small shift h, we want to write

$$f(x+h) \approx f(x) + kh.$$

Here, k is the slope of the straight line. In order to obtain k, we can rearrange the above to get

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

As we pick smaller and smaller neighbourhoods of x, we want our approximation to get better and better. Thus, if such an approximation is possible, then the value of k must stabilize. This means that the limit

$$k = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

must exist. Note that this immediately forces the continuity of f, since

$$\lim_{h \to 0} f(x+h) - f(x) = \lim_{h \to 0} h \cdot \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = 0k = 0,$$

whereby $\lim_{x \to a} f(x) = f(a)$. Splitting the limit is justified because the individual limits exist. If such a limit k exists, we call it the derivative of f at x, denoted $f'(x)$. We are now able to write

$$f(x+h) \approx f(x) + f'(x)h.$$

Definition 2.1 (Derivative)**.** The derivative of a function $f\colon [a,b] \to \mathbb{R}$ at a point $x \in [a,b]$ is defined as

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

if such a limit exists. Note that we only demand a one-sided limit if x is an endpoint. If the derivative of f exists at every point in $[a,b]$, we say that f is differentiable on $[a,b]$.

Example. Consider the map $x \mapsto x^n$, where $n \in \mathbb{N}$. Using the binomial theorem, we can write

$$(x+h)^n = x^n + nx^{n-1}h + \cdots + h^n,$$

which means that

$$\frac{d}{dx}x^n = \lim_{h\to 0}\frac{1}{h}\left[(x+h)^n - x^n\right] = \lim_{h\to 0}\left[nx^{n-1} + \binom{n}{2}x^{n-2}h + \cdots + h^{n-1}\right] = nx^{n-1}.$$

Note that the process of differentiation we described can be generalised to multivariable functions. The idea is to locally approximate a function with an affine function.

Theorem 2.1. *If $f\colon (a,b) \to \mathbb{R}$ is differentiable on (a,b), then it is also continuous on (a,b).*

Theorem 2.2. *Let $f\colon I \to \mathbb{R}$ be a continuous function. Then,*

1. *f maps compact sets to compact sets.*
2. *f maps connected sets to connected sets.*

Corollary 2.2.1. *A continuous function $f\colon I \to \mathbb{R}$ maps intervals to intervals.*

Corollary 2.2.2. *A continuous function $f\colon [a,b] \to \mathbb{R}$ attains its minimum and maximum on $[a,b]$.*

Definition 2.2. Given $f\colon (a,b) \to \mathbb{R}$, a point $c \in (a,b)$ is said to be a point of local maximum if there exists a neighbourhood I_c of c such that

$$f(c) > f(x),$$

for all $x \in I_c \setminus \{c\}$. There is an analogous definition for a local minimum.

Theorem 2.3. *If $f\colon (a,b) \to \mathbb{R}$ is differentiable and $c \in (a,b)$ is a point of local minimum or maximum, then $f'(c) = 0$.*

Remark. The converse is not true. Note that the derivative of $x \mapsto x^3$ vanishes at $x = 0$, but that is not a local minimum or maximum.

Proof. Let c be a local minimum or maximum of f, but suppose that $f'(c) \neq 0$. Define the function

$$g\colon (a,b) \to \mathbb{R}, \qquad g(x) = \begin{cases} (f(x) - f(c))/(x-c), & \text{if } x \neq c \\ f'(c), & \text{if } x = c \end{cases}$$

We note that g is continuous. Also, $f'(c) = g(c) \neq 0$. If $g(c) > 0$, there exists a neighbourhood $I_\delta = (c-\delta, c+\delta)$ such that for all $x \in I_\delta$, $g(x) > 0$, from the continuity of g. This means that on I_c,

$$\frac{f(x)-f(c)}{x-c} > 0,$$

which gives $f(x) > f(c)$ on $(c, c+\delta)$ and $f(x) < f(c)$ on $(c-\delta, c)$. This means that c cannot be a local minimum, nor a local maximum. There is an analogous case assuming $g(c) < 0$, which leads to the same contradiction. Thus, we must have $f'(c) = g(c) = 0$. □

Theorem 2.4. *If $f\colon (a,b) \to \mathbb{R}$ is twice differentiable, and $c \in (a,b)$ is such that $f'(c) = 0$ and $f''(c) < 0$, then c is a point of local maximum. If $f'(c) = 0$ and $f''(c) > 0$, then c is a point of local minimum.*

Theorem 2.5 (Rolle's Theorem)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous, and differentiable on (a,b), with $f(a) = f(b)$. Then, there exists $c \in (a,b)$ such that $f'(c) = 0$.*

Proof. Set $f(a) = f(b) = \kappa$. From the continuity of f, note that the image of the closed interval $[a,b]$ is another closed interval $[\alpha, \beta]$. This means that $\alpha \leq \kappa \leq \beta$. Note that if $\alpha = \beta = \kappa$, then the function f is identically equal to the constant κ, hence $f'(x) = 0$ everywhere on $[a,b]$. By the continuity of f, it must attain its maximum and minimum on $[a,b]$. If $\beta > \kappa$, then the maximum is al least β and is hence not attained at the endpoints, which means that the point of maximum lies in (a,b). If $\alpha < \kappa$, then the same argument shows that f attains a minimum in (a,b). Thus, in either case, we have found $c \in (a,b)$ which is either a maximum or minimum of f, i.e. $f'(c) = 0$. □

Theorem 2.6 (Mean Value Theorem)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous, and differentiable on (a,b). Then, there exists $c \in (a,b)$ such that*

$$f(b) - f(a) = f'(c)\,(b-a).$$

Proof. Apply Rolle's Theorem on the function defined as

$$g\colon [a,b] \to \mathbb{R}, \qquad g(x) = f(x) - f(a) - \frac{f(b)-f(a)}{b-a} \cdot (x-a).$$

Note that g is continuous on $[a,b]$, differentiable on (a,b), and $g(a) = g(b) = 0$. □

Theorem 2.7. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be differentiable, and $f'(x) > 0$ for all $x \in \mathbb{R}$. Then, f is strictly increasing on $\mathbb{R}$.*

Proof. Let $x_2 > x_1$. By the mean value theorem, we pick $c \in (x_1, x_2)$ such that

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) > 0.$$ □

Remark. The converse is not true. The map $x \mapsto x^3$ is strictly increasing, but its derivative vanishes at 0.

Theorem 2.8 (Chain rule)**.** *Let f and g be differentiable on $\mathbb{R}$. Then, $f \circ g$ is also differentiable, with*

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

Proof. Fix $a \in \mathbb{R}$. Define the functions

$$\varphi \colon \mathbb{R} \to \mathbb{R}, \qquad \varphi(x) = \begin{cases} (g(x) - g(a))/(x-a) & \text{if } x \neq a \\ g'(a), & \text{if } x = a \end{cases},$$

$$\psi \colon \mathbb{R} \to \mathbb{R}, \qquad \psi(y) = \begin{cases} (f(y) - f(b))/(y-b) & \text{if } y \neq b \\ f'(b), & \text{if } y = b \end{cases}.$$

Note that φ and ψ are continuous. Also, when $x \neq a$, we have

$$g(x) - g(a) = \varphi(x)(x-a).$$

Set $b = g(a)$, and write

$$f(g(x)) - f(g(a)) = \psi(g(x))(g(x) - g(a)) = \psi(g(x))\,\varphi(x)(x-a).$$

Setting $h = f \circ g$, we have

$$\frac{h(x) - h(a)}{x - a} = \psi(g(x))\varphi(x).$$

Taking limits $x \to a$, we use the continuity of φ, ψ and g to conclude that the derivative of h is indeed defined at a, and

$$h'(a) = \psi(g(a))\,\varphi(a) = f'(g(a))\,g'(a).$$ □

Definition 2.3 (Intermediate Value Property)**.** Let $f \colon (a,b) \to \mathbb{R}$ be such that for all $c, d \in (a,b)$ such that $f(c) < f(d)$ and $\lambda \in (f(c), f(d))$, there exists $x_0 \in (a,b)$ such that $f(x_0) = \lambda$. Then, we say that f has the intermediate value property.

Theorem 2.9 (Intermediate Value Theorem)**.** *All continuous functions $f \colon (a,b) \to \mathbb{R}$ have the intermediate value property.*

Theorem 2.10. *Let $f \colon (a,b) \to \mathbb{R}$ be differentiable. Then, f' satisfies the intermediate value property.*

Proof. Let $c, d \in (a,b)$ and let $\lambda \in \mathbb{R}$ such that $\lambda \in (f'(c), f'(d))$. We wish to find $x_0 \in (a,b)$ such that $f'(x_0) = \lambda$. Define

$$g \colon (a,b) \to \mathbb{R}, \qquad g(x) = f(x) - \lambda x.$$

Note that $g'(x) = f'(x) - \lambda$, so $g'(c) < 0$ and $g'(d) > 0$. Thus, g is decreasing near c and increasing near d, so we can find $t_1, t_2 \in (c,d)$ such that $g(t_1) < g(c)$ and $g(t_2) < g(d)$. This means that g has no local minimum at c nor d. From the continuity of g, there exists $x_0 \in [c,d]$ such that $g(x_0) = \inf_{[c,d]} g(x)$. We have already shown that x_0 is neither c, nor d, so $x_0 \in (c,d)$. Hence, $g'(x_0) = 0$, which gives $f'(x_0) = \lambda$. □

Lemma 2.11. *If $f\colon (a,b) \to (c,d)$ is surjective, continuous and strictly increasing, then f is invertible with a continuous inverse.*

Theorem 2.12 (Inverse function theorem)**.** *Let $f\colon (a,b) \to (c,d)$ be surjective and differentiable, with $f'(x) \neq 0$ everywhere. Then, f is invertible, with a differentiable inverse whose derivative is given by*

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

Proof. Given $f'(x) \neq 0$ on (a,b). Then intermediate value property gives either $f'(x) > 0$ for all $x \in (a,b)$, or $f'(x) < 0$. Without loss of generality, assume the former. This means that f is strictly increasing on (a,b), continuous, and surjective. Our lemma gives the existence of a continuous inverse f^{-1}.

Let $y \in (c,d)$, and let $x = f^{-1}(y)$. From the continuity of f^{-1}, we can always write $f^{-1}(y+\kappa) = x + h$. Thus,

$$\lim_{\kappa\to 0} \frac{f^{-1}(y+\kappa) - f^{-1}(y)}{\kappa} = \lim_{\kappa\to 0} \frac{x+h-x}{\kappa} = \lim_{\kappa\to 0} \frac{h}{\kappa}.$$

Note that $h \to 0$ as $\kappa \to 0$. Thus, this limit can be written as

$$(f^{-1})'(y) = \lim_{h\to 0} \frac{h}{f(x+h) - f(x)} = \frac{1}{f'(x)}.$$

□

Corollary 2.12.1. *Let f be continuously differentiable on $\mathbb{R}$, with $f'(x_0) \neq 0$ for some $x_0 \in \mathbb{R}$. Then, there exists some neighbourhood of x_0 on which f is invertible, with a continuously differentiable inverse.*

Theorem 2.13. *Let $f_n \to f$ pointwise and $\{f_n'\}$ converge uniformly on some interval (a,b). Then, $f_n \to f$ uniformly.*

Proof. We show that $\{f_n\}$ is uniformly Cauchy on E. Note that for some fixed t, we can write

$$|f_n(x) - f_m(x)| \leq |(f_n(x) - f_m(x)) - (f_n(t) - f_m(t))| + |f_n(t) - f_m(t)|.$$

Using the Mean Value Theorem, the first term can be bounded as

$$|(f_n(x) - f_m(x)) - (f_n(t) - f_m(t))| = (f_n' - f_m')(x_0)|x-t|,$$

where x_0 is between x and t. From the pointwise convergence of $f_n \to f$, we have

$$|f_n(t) - f_m(t)| < \frac{\epsilon}{2}$$

for all $n, m \geq N_t$. The uniform convergence of $\{f_n'\}$ means that we can find N_0 such that

$$|f_n'(x_0) - f_m'(x_0)| < \frac{\epsilon}{2(b-a)}$$

for all $n, m > N_0$. Thus, for all $x \in [a,b]$, and $n, m \geq N_t + N_0$, we have

$$|f_n(x) - f_m(x)| < \frac{\epsilon}{2(b-a)} \cdot (b-a) + \frac{\epsilon}{2} = \epsilon.$$

This means that $\{f_n\}$ is uniformly Cauchy on $[a,b]$, which gives the uniform convergence of $\{f_n\}$. $\square$

Remark. We only needed to use the pointwise convergence of $\{f_n\}$ at one point t. By using pointwise convergence everywhere, we can allow for unbounded intervals, or the entirety of $\mathbb{R}$.

Theorem 2.14. *Let $\{f_n\}$ be a sequence of differentiable functions on some bounded interval (a,b) such that $f_n \to f$ pointwise and $\{f_n'\}$ converges uniformly on every $[\alpha, \beta] \subset (a,b)$. Then, f is differentiable and $f_n' \to f'$.*

Remark. We allow a, b to be $\pm\infty$.

Proof. Let $x_0 \in (a,b)$. We wish to show that the following limit exists.

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Define $\varphi \colon (a,b) \setminus \{x_0\} \to \mathbb{R}$,

$$\varphi(x) = \frac{f(x) - f(x_0)}{x - x_0}.$$

Also define the functions $\varphi_n \colon (a,b) \to \mathbb{R}$,

$$\varphi_n(x) = \begin{cases} (f_n(x) - f_n(x_0))/(x - x_0) & \text{if } x \neq x_0, \\ f_n'(x_0) & \text{if } x = x_0. \end{cases}$$

Note that φ_n are continuous, from the continuity of each f_n. When $x \neq x_0$, we see that $\varphi_n(x) \to \varphi(x)$. For $x = x_0$, we know that f_n' converges hence $\varphi_n(x_0)$ also converges. This gives us pointwise convergence.

We want to show that $\{\varphi_n\}$ converges uniformly. Fix $[\alpha, \beta] \subset (a,b)$ such that $x_0 \in (\alpha, \beta)$. For $x \neq x_0$, we have

$$|\varphi_n(x) - \varphi_m(x)| = \left| \frac{(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))}{x - x_0} \right|.$$

Using the Mean Value Theorem on $g = f_n - f_m$, we choose c between x and x_0 such that $(x - x_0)g'(c) = g(x) - g(x_0)$. Thus,

$$|\varphi_n(x) - \varphi_m(x)| = |f_n'(c) - f_m'(c)| < \epsilon$$

for all $m, n \geq N$ for some N, given by the uniform convergence of $\{f_n'\}$. This shows that $\{\varphi_n\}$ also converges uniformly on $[\alpha, \beta]$. Note that when $x = x_0$, $|f_n'(x_0) - f_m'(x_0)|$ is similarly bounded.

Now that $\{\varphi_n\}$ converges uniformly, we know that the limit function is continuous. Since it converges pointwise to φ on $x \neq x_0$ and to $\lim_{n\to\infty} f_n'(x_0)$ when $x = x_0$, continuity gives the existence of the desired limit and

$$\lim_{n \to \infty} f_n'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0),$$

which gives the differentiability of f. Also note that $f_n' \to f'$. $\square$

Lemma 2.15 (Abel's Lemma). *Let*

$$\sum_{x=0}^{\infty} a_n(x-a)^n$$

be a convergent power series on $(a-R, a+R)$*. If the series converges absolutely for some* $x = a+c$ *within that interval, then it must converge uniformly on any closed interval* $[\alpha, \beta] \subset (a-c, a+c)$*.*

Proof. Note that for all $x \in [\alpha, \beta]$,

$$|a_n(x-a)^n| \leq |a_n||c|^n$$

which gives the uniform convergence of $\sum_{n=1}^{\infty} a_n(x-a)^n$ on $[\alpha, \beta]$ by the Weierstrass M-test. □

Lemma 2.16. *If a power series converges absolutely on* $(a-R, a+R)$*, then it is differentiable, with the derivative being*

$$\sum_{n=1}^{\infty} n a_n(x-a)^{n-1}.$$

Proof. The absolute convergence of the power series gives its uniform convergence on every compact subset of $(a-R, a+R)$. Note that this gives the continuity of the power series. Now, note that

$$\limsup_{n\to\infty} |na_n|^{1/n} = \lim_{n\to\infty} n^{1/n} \cdot \limsup_{n\to\infty} |a_n|^{1/n} = \limsup_{n\to\infty} |a_n|^{1/n},$$

so the series of derivatives of terms of the power series converges absolutely on the same domain. This again gives the uniform convergence of this series of derivatives of terms. Abel's Lemma gives uniform convergence on all compact subsets. Thus, by the previous theorem, our power series is differentiable, with the derivative equal to the series of derivatives of terms. □

Corollary 2.16.1. *A power series is infinitely differentiable on its interval of convergence.*

Example. Consider $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto \sin x$. We want to show that $f'(x) = \cos x$. Write f as a power series,

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

This converges absolutely on $\mathbb{R}$. Our lemma gives

$$f'(x) = \sum_{n=0}^{\infty} \frac{(-1)^{2n} x^{2n}}{(2n)!} = \cos x.$$

Example. Consider the function

$$f\colon \mathbb{R} \to \mathbb{R}, \qquad f(x) = \begin{cases} e^{-1/x}, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases}$$

This function is differentiable everywhere. This is easily seen when $x \neq 0$. For $x = 0$, the left hand limit is simply

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

and from the positive side,

$$\lim_{x \to 0^+} \frac{f(x) - f(0)}{x} = \lim_{x \to 0^+} \frac{e^{-1/x}}{x} = 0.$$

Hence, $f'(0) = 0$. Indeed,

$$\lim_{x \to 0^+} \frac{e^{-1/x}}{p(x)} = 0$$

for any polynomial $p(x)$, which means that f is infinitely differentiable, with $f^{(n)}(0) = 0$. This means that if f has a power series centred at $x = 0$, all its coefficients must be identically zero. Thus, f has no power series around $x = 0$.

Definition 2.4. We notate the n^{th} derivative of a function f as

$$f^{(n)} = \frac{d^n f}{dx^n} = \underbrace{\frac{d}{dx}\left(\frac{d}{dx}\left(\cdots \frac{df}{dx} \cdots\right)\right)}_{n \text{ times}}.$$

Example. Consider

$$f\colon \mathbb{R} \to \mathbb{R}, \qquad x \mapsto x^m.$$

Differentiating n times, we write

$$f^{(n)}(x) = \frac{d^n}{dx^n} x^m = \begin{cases} m(m-1)\dots(m-n+1)x^{m-n} & \text{if } n \leq m, \\ 0 & \text{if } n > m. \end{cases}$$

Example. Consider

$$f\colon \mathbb{R} \to \mathbb{R}, \qquad x \mapsto \sin x.$$

Differentiating n times, we write

$$f^{(n)}(x) = \frac{d^n}{dx^n} \sin x = \sin\left(x + \frac{n\pi}{2}\right).$$

Theorem 2.17. *Let f be differentiable n times at x, and let it be of the form*

$$f = g \cdot h$$

where g and h are also differentiable n times at x. Then,

$$f^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} g^{(k)}(x) h^{(n-k)}(x).$$

Proof. For $n = 1$, this is simply the product rule. Suppose that this is true for some $n \geq 1$. Then,

$$f^{(n+1)}(x) = \frac{d}{dx} f^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} \frac{d}{dx} \left[g^{(k)}(x) h^{(n-k)}(x) \right].$$

Now,

$$\frac{d}{dx} \left[g^{(k)}(x) h^{(n-k)}(x) \right] = g^{(k+1)}(x) h^{(n-k)}(x) + g^{(k)}(x) h^{(n-k+1)}(x).$$

Plugging this back in and shifting indices,

$$f^{(n+1)}(x) = \sum_{k=1}^{n+1} \binom{n}{k-1} g^{(k)}(x) h^{(n-k+1)}(x) + \sum_{k=0}^{n} \binom{n}{k} g^{(k)}(x) h^{(n-k+1)}(x).$$

Separating the first and last terms and using

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$$

completes the proof by induction. □

Lemma 2.18. *Suppose that $p(x)$ is a polynomial of degree n, and we have been given the values*

$$b_i = \frac{p^{(n)}(a)}{n!}$$

for all $i = 0, \ldots, n$. Then, the polynomial is uniquely determined, as

$$p(x) = b_0 + b_1(x-a) + \cdots + b_n(x-a)^n.$$

Definition 2.5 (Taylor polynomial)**.** Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function such that $f', f'', \ldots, f^{(n)}$ are also continuous. Then, for $x_0 \in (a,b)$, the polynomial

$$p(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)(x-x_0)^k}{k!}$$

is called the Taylor polynomial of f of degree n about the point x_0.

Remark. Note that

$$p(x_0) = f(x_0), \quad p'(x_0) = f'(x_0), \quad p''(x_0) = f''(x_0), \quad \ldots \quad p^{(n)}(x_0) = f^{(n)}(x_0).$$

Definition 2.6. A polynomial p is said to approximate a function f up to the order n near a point x_0 if for every $\epsilon > 0$, there exists a δ neighbourhood of x_0 where for all $k = 0, \dots, n$,

$$|f^{(k)}(x) - p^{(k)}(x)| < \epsilon$$

Lemma 2.19. *The Taylor polynomial of degree n of a function f approximates it up to order n.*

Proof. Note that

$$|f^{(k)}(x) - p^{(k)}(x)| \leq |f^{(k)}(x) - f^{(k)}(x_0)| + |f^{(k)}(x_0) - p^{(k)}(x_0)| + |p^{(k)}(x_0) - p^{(k)}(x)|.$$

For the Taylor polynomial, the central term is zero. The continuity of $f^{(k)}$ and $p^{(k)}$ allow us to control the remaining terms, giving the desired result. □

Definition 2.7 (Remainder)**.** Let p approximate f up to order n. Then, the remainder term is defined on the interval of approximation as

$$R_{n+1}(x) = f(x) - p(x).$$

Definition 2.8 (Big O and small o notation)**.** Let f and g be two functions on a neighbourhood of x_0. We say that $f \sim O(g)$ near x_0 if there exists $M > 0$ such that

$$\frac{|f(x)|}{|g(x)|} \leq M$$

for all points x near x_0.

We say that $f \sim o(g)$ near x_0 if

$$\lim_{x \to x_0} \frac{|f(x)|}{|g(x)|} = 0.$$

Theorem 2.20 (Taylor's theorem)**.** *Let $f\colon [a, b] \to \mathbb{R}$ be such that $f, f', \dots, f^{(n+1)}$ are continuous. Then,*

$$f(x) = p(x) + R_{n+1}(x)$$

where p is the Taylor polynomial of degree n of f around some point $x_0 \in (a, b)$, and R_{n+1} is defined as

$$R_{n+1}(x) = \frac{f^{(n+1)}(c)(x - x_0)^{n+1}}{(n + 1)!}$$

for some c between x and x_0.

Remark. The former form of R_{n+1} is called Lagrange's form. The following is Cauchy's form; for $0 < \theta < 1$,

$$R_{n+1}(x) = \frac{f^{(n+1)}(x_0 + \theta(x - x_0))(x - x_0)^{n+1}(1 - \theta)^n}{n!}$$