Mathematical Inequalities using Taylor Series

Hemanta K. Maji

January 8, 2018

1 Overview

We begin by recalling the Rolle's Theorem.[1] Using this result, we shall derive the Lagrange Form of the Taylor's Remainder Theorem. Subsequently, we shall derive several mathematical inequalities as a corollary of this result. For example,

1. We shall use the Taylor's Remainder Theorem to upper and lower bound exponential functions using polynomials.
2. We shall use the Taylor' Remainder Theorem to obtain the Jensen's Inequality.
   - From the Jensen's Inequality, we shall derive standard mathematical inequalities like the AM-GM-HM inequality, Cauchy-Schwarz inequality, Young's inequality, and Hölder's inequality,
3. In the future, we shall use the Taylor's Remainder Theorem to prove the Bonami-Gross-Beckner Hypercontractivity Theorem in Fourier Analysis.

2 Rolle's Theorem

Theorem 1 (Rolle's Theorem)**.** *Let f be a real function that is differentiable. Suppose $a < b$ and we have $f(a) = f(b)$. Then, there exists $c \in [a, b]$ such that $f'(c) = 0$.*

Intuitively, if the a function achieves equal values at the extreme points of the interval (a, b) then somewhere in the middle its slope is 0, i.e., it achieves an extremum (or, the function is a constant function in the range $[a, b]$).

2.1 Mean Value Theorem

We can prove a generalization of this result, namely the Mean value theorem, using Rolle's Theorem.

Theorem 2 (Mean value theorem)**.** *Let f be a real function that is differentiable in the range $[a, a + \varepsilon]$. Then, there exists $\theta \in (0, 1)$ such that*

$$f'(a + \theta\varepsilon) = \frac{f(a + \varepsilon) - f(a)}{\varepsilon}$$

[1] We emphasize that we shall *not* prove this result, but use this result to derive interesting theorems. Students who are interested in understanding the proof of this theorem are referred to the proof of Rolle's theorem, Mean-value theorem, and Cauchy's Mean-value theorem using the Extreme value theorem.

Note that if $f(x) = f(x+\varepsilon)$, then the mean value theorem is equivalent to the Rolle's theorem. We can, in fact, prove the mean value theorem using the Rolle's theorem.

Proof. Consider the following function

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

Since the function f is differentiable, the function g is also differentiable. In fact, we have

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

Note that $g(a) = f(a)$, and $g(a+\varepsilon) = f(a)$. So, we can apply the Rolle's Theorem to conclude that there exists $\delta \in (0,\varepsilon)$ such that $g'(a+\delta) = 0$. That implies that

$$\begin{aligned} & g'(a+\delta) = 0 \\ \iff \quad & f'(a+\delta) - \frac{f(a+\varepsilon) - f(a)}{\varepsilon} = 0 \\ \iff \quad & f'(a+\delta) = \frac{f(a+\varepsilon) - f(a)}{\varepsilon} \end{aligned}$$

Define $\theta = \delta/\varepsilon$, and we get the result that there exists $\theta \in (0,1)$ such that

$$f'(a+\theta\varepsilon) = \frac{f(a+\varepsilon) - f(a)}{\varepsilon}$$

This completes the proof of the mean value theorem using the Rolle's theorem. □

Intuition. An equivalent manner of expressing $f'(a+\theta\varepsilon) = \frac{f(a+\varepsilon)-f(a)}{\varepsilon}$ is

$$f(a+\varepsilon) = f(a) + f'(a+\theta\varepsilon)\varepsilon$$

This expressing states that $f(a+\varepsilon)$ is roughly $f(a)$ plus a linear-term in ε. Can we approximate $f(a+\varepsilon)$ using higher degree polynomials in ε? The Taylor series is an approach towards this objective.

2.2 Generalized Rolle's Theorem

In this section we shall derive a generalized form of Rolle's Theorem that shall help us prove the Lagrange form of the Taylor's Remainder Theorem. In the sequel, we shall refer to the k-th order derivative of f as $f^{(k)}$. Moreover, we shall use $f^{(0)}$ to represent the function f.

Theorem 3 (Generalized Rolle's Theorem)**.** *Let f be a function that is differentiable k times. Suppose $f(a) = f(a+\varepsilon)$ and $f^{(1)}(a) = f^{(2)}(a) = \cdots = f^{(k)}(a) = 0$. Then, there exists $\theta \in (0,1)$ such that $f^{(k+1)}(a+\theta\varepsilon) = 0$.*

Proof. The proof shall proceed by repeated application of the mean value theorem. Note that $f(a) = f(a+\varepsilon)$, so there exists $\theta_1 \in (0,1)$ such that $f^{(1)}(a+\theta_1\varepsilon) = 0$.

Now, note that $f^{(1)}(a) = f^{(1)}(a+\theta_1\varepsilon)$, so there exists $\theta_2 \in (0,1)$ such that $f^{(2)}(a+\theta_2\cdot\theta_1\varepsilon) = 0$.

Iteratively applying this approach, we deduce that there exists $\theta_1,\dots,\theta_{k+1} \in (0,1)$ such that

$$f^{(k+1)}(a+\theta_{k+1}\cdots\theta_1\varepsilon) = 0$$

We define $\theta = \theta_{k+1}\cdots\theta_1$ and we get the result. □

3 Lagrange form of the Taylor's Remainder Theorem

Theorem 4 (Lagrange form of the Taylor's Remainder Theorem)**.** *Let f be a real function that is differentiable $(k+1)$ times. Then, there exists $\theta \in (0,1)$ such that the following holds true.*

$$f(a+\varepsilon) = \sum_{i=0}^{k} f^{(i)}(a)\frac{\varepsilon^i}{i!} + f^{(k+1)}(a+\theta\varepsilon)\frac{\varepsilon^{k+1}}{(k+1)!}$$

Note that the above theorem for $k = 0$ is identical to the mean value theorem. Let us prove the theorem for $k = 1$ using the generalized Rolle's theorem. We shall later generalize this proof to prove the general case of arbitrary k.

3.1 Toy Example of $k = 1$

So, we want to show that there exists θ such that

$$f(a+\varepsilon) = f(a) + f^{(1)}(a)\frac{\varepsilon}{1!} + f^{(2)}(a+\theta\varepsilon)\frac{\varepsilon^2}{2!}$$

Consider the polynomial

$$\alpha(x) = f(a) + f^{(1)}(a)\frac{(x-a)}{1!}$$

Note that the polynomial α has the following properties

1. $\alpha(a) = f(a)$, and
2. $\alpha^{(1)}(a) = f^{(1)}(a)$

Next, consider the polynomial h defined as follows

$$h(x) = \alpha(x) + \frac{f(a+\varepsilon) - \alpha(a+\varepsilon)}{\varepsilon^2}(x-a)^2$$

Note that the polynomial h has the following properties

1. $h(a) = f(a)$
2. $h^{(1)}(a) = f^{(1)}(a)$, and
3. $h(a+\varepsilon) = f(a+\varepsilon)$

Finally, consider the polynomial g defined as follows

$$g(x) = f(x) - h(x)$$

Note that the polynomial g has the following properties

1. $g(a) = g(a+\varepsilon)$,
2. $g^{(1)}(a) = 0$

Now, we can apply the generalized Rolle's theorem for $k = 1$ to conclude that there exists $\theta \in (0, 1)$ such that

$$g^{(k+1)}(a + \theta\varepsilon) = 0$$

This is equivalent to

$$f^{(k+1)}(a + \theta\varepsilon) = h^{(k+1)}(a + \theta\varepsilon) \tag{1}$$

To calculate $h^{(k+1)}(x)$, note that α is a degree k polynomial in x. So, we have $\alpha^{(k+1)}(x) = 0$. Therefore, we have $h^{(k+1)}(x) = (k+1)!\frac{f(a+\varepsilon)-\alpha(a+\varepsilon)}{\varepsilon^2}$.

This reduces Equation 1 to

$$f^{(k+1)}(a + \theta\varepsilon) = (k+1)!\frac{f(a+\varepsilon) - \alpha(a+\varepsilon)}{\varepsilon^{k+1}}$$

Rearranging, we get

$$f(a+\varepsilon) = \alpha(a+\varepsilon) + f^{(k+1)}(a+\theta\varepsilon)\frac{\varepsilon^{k+1}}{(k+1)!}$$

Note that we have

$$\alpha(a+\varepsilon) = f(a) + f^{(1)}\frac{\varepsilon}{1!}$$

Given this modular approach to the construction of the polynomials α, h, and g, we can easily prove the Lagrange form of the Taylor's Remainder theorem.

3.2 Full Proof for General k

Define the following polynomial

$$\alpha(x) = \sum_{i=0}^{k} f^{(i)}(a)\frac{(x-a)^i}{i!}$$

Note that we have the following property. For all $i \in \{0, \dots, k\}$, we have $\alpha^{(i)}(a) = f^{(i)}(a)$.

Define the polynomial

$$h(x) = \alpha(x) + \frac{f(a+\varepsilon) - \alpha(a+\varepsilon)}{\varepsilon^{k+1}}(x-a)^{k+1})$$

This polynomial has the property that

1. For all $i \in \{0, \dots, k\}$, we have $h^{(i)}(a) = f^{(i)}(a)$, and
2. $h(a+\varepsilon) = f(a+\varepsilon)$.

Next, define the polynomial $g(x) = f(x) - h(x)$. The polynomial g has the property that

1. For all $i \in \{0, \dots, k\}$, we have $g^{(i)}(a) = 0$, and
2. $g(a) = g(a+\varepsilon) = 0$.

Finally, we apply the general Rolle's theorem to conclude that there exists $\theta \in (0,1)$ such that

$$g^{(k+1)}(a+\theta\varepsilon) = 0$$

This statement is equivalent to

$$f^{(k+1)}(a+\theta\varepsilon) = h^{(k+1)}(x+\theta\varepsilon) \tag{2}$$

As argued in the previous section, we have

$$h^{(k+1)}(x+\theta\varepsilon) = (k+1)!\frac{f(a+\varepsilon)-\alpha(a+\varepsilon)}{\varepsilon^{k+1}}$$

Substituting this value in Equation 2 and rearranging, we get

$$\begin{aligned} f(a+\varepsilon) &= \alpha(a+\varepsilon) + f^{(k+1)}(a+\theta\varepsilon)\frac{\varepsilon^{k+1}}{(k+1)!} \\ &= \sum_{i=0}^{k} f^{(i)}(a)\frac{\varepsilon^i}{i!} + f^{(k+1)}(x+\theta\varepsilon)\frac{\varepsilon^{k+1}}{(k+1)!} \end{aligned}$$

4 Application 1: Jensen's Inequality

Definition 1 (Convex Function). *A function f is convex in the range $[a,b]$ is $f^{(2)}$ is positive in the range (a,b).*

A concave function has $f^{(2)}$ negative in the range (a,b). For example, $f(x)=x^2$, $f(x)=1/x$, and $f(x)=\exp(x)$ are a few representative example of convex functions. On the other hand, $f(x)=\sqrt{x}$, and $f(x)=\log x$ are a few representative example of concave functions. Note that if f is convex then the function $-f$ is concave; and vice-versa.

Jensen's Inequality is a general inequality that holds for any convex function.

Theorem 5 (Jensen's Inequality). *If f'' is positive in the range $[a,b]$, then we have*

$$\frac{f(a)+f(b)}{2} \geqslant f\left(\frac{a+b}{2}\right)$$

Moreover, equality holds if and only if $a=b$.

For a concave function, the direction of the inqueality is reversed. Let us prove Jensen's Inequality using the Lagrange form of the Taylor's remainder theorem.

Proof. Let $\mu = \frac{a+b}{2}$ and $\varepsilon = \frac{b-a}{2}$. So, we want to prove that

$$\frac{f(\mu-\varepsilon)+f(\mu+\varepsilon)}{2} \geqslant f(\mu)$$

Consider the following function

$$g(x) = \frac{f(\mu+x)+f(\mu-x)}{2}$$

So, to prove the Jensen's inequality, we need to show that $g(\varepsilon) \geqslant g(0)$.

Let us apply the Lagrange form of the Taylor's remainder theorem to the function g at $a = 0$

$$g(\varepsilon) = g(0) + g^{(1)}(0)\frac{\varepsilon}{1!} + g^{(2)}(\theta\varepsilon)\frac{\varepsilon^2}{2!}$$

Note that $g^{(1)}(x) = \frac{f^{(1)}(\mu+x)-f^{(1)}(\mu-x)}{2}$, and $g^{(1)}(0) = 0$. Further, we have $g^{(2)}(x) = \frac{f^{(2)}(\mu+x)+f^{(2)}(\mu-x)}{2}$. Note that when $x \in [0, \varepsilon]$, both $f^{(2)}(\mu + x)$ and $f^{(2)}(\mu - x)$ are positive. So, we get

$$g(\varepsilon) = g(0) + \overbrace{g^{(1)}(0)\frac{\varepsilon}{1!}}^{=0} + \overbrace{g^{(2)}(\theta\varepsilon)\frac{\varepsilon^2}{2!}}^{\geqslant 0}$$

This proves that $g(\varepsilon) \geqslant g(0)$, and equality holds if and only if $\varepsilon = 0$, i.e., $a = b$. $\square$

We can use this form of the Jensen's inequality to prove the following general Jensen's inequality.

Theorem 6. *Let f be a convex function. Let $\mathbb{X}$ be a random variable over the sample space $\Omega \subseteq \mathbb{R}$ Then, the following holds*

$$\mathbb{E}\left[f(\mathbb{X})\right] \geqslant f\left(\mathbb{E}\left[\mathbb{X}\right]\right)$$

And equality holds if and only if the support of $\mathbb{X}$ has a unique element.

4.1 AM-GM-HM Inequality

We are interested to show the following result.

Theorem 7 (AM-GM)**.** *For positive a, b, we have $\frac{a+b}{2} \geqslant \sqrt{ab}$. Equality holds if and only if $a = b$.*

Prove this theorem by applying Jensen's inequality to the function $f(x) = \log x$.

The GM-HM inequality states that the geometric mean $\sqrt{ab}$ is great-than-equal-to the harmonic mean $\left(\frac{\frac{1}{a}+\frac{1}{b}}{2}\right)^{-1}$. This is a corollary of the AM-GM inequality.

4.2 Cauchy-Schwarz Inequality

We are interested to show the following result.

Theorem 8 (Cauchy-Schwarz Inequality)**.** *Let a_1, a_2, b_1, b_2 be positive reals. Then, the following holds*

$$(a_1b_1 + a_2b_2) \leqslant \left(a_1^2 + a_2^2\right)^{1/2}\left(b_1^2 + b_2^2\right)^{1/2}$$

Equality holds if and only if $\frac{a_1}{b_1} = \frac{a_2}{b_2}$.

We can rearrange the target result as follows

$$\left(1 + \frac{a_2}{a_1}\cdot\frac{b_2}{b_1}\right) \leqslant \left(1 + \left(\frac{a_2}{a_1}\right)^2\right)^{1/2}\left(1 + \left(\frac{b_2}{b_1}\right)^2\right)^{1/2}$$

Substitutive, $(a_2/a_1)^2 = a$ and $(b_2/b_1)^2 = b$, we get

$$(1 + \sqrt{ab}) \leqslant (1 + a)^{1/2}(1 + b)^{1/2}$$

Apply Jensen's inequality to the function $f(x) = \log\left(1 + \exp(x)\right)$ to prove the Cauchy-Schwarz inequality.

4.3 Young's Inequality

Let p, q be positive reals such that

$$\frac{1}{p}+\frac{1}{q}=1$$

Note that both p and q have to be greater than 1. The numbers p and q are referred to as the Hölder conjugate of each other.

Theorem 9 (Young's Inequality)**.** *Let p and q be Hölder conjugates. Then, the following holds*

$$ab \leqslant \frac{a^p}{p}+\frac{b^q}{q}$$

Equality holds if and only if $a^p = b^q$.

Apply Jensen's inequality to $f(x) = \log(x)$ to prove this inequality. Note that $p = q = 2$ yields the AM-GM inequality.

4.4 Hölder's Inequality

Theorem 10 (Hölder's Inequality)**.** *Let p and q be Hölder conjugates. Let a_1, a_2, b_1, b_2 be positive reals. Then, the following holds*

$$(a_1 b_1 + a_2 b_2) \leqslant \left(a_1^p + a_2^p\right)^{1/p} \left(b_1^q + b_2^q\right)^{1/q}$$

Prove this theorem using the Jensen's inequality. When does equality hold?
Note that $p = q = 2$ corresponds to the Cauchy-Schwarz inequality.

5 Application 2: Approximating $\exp(-x)$ and $\ln(1-x)$

5.1 Approximating $\exp(-x)$ using polynomials

Let $f(x) = \exp(-x)$ and $a = 0$. Note that $f^{(i)}(x) = (-1)^i \exp(-x)$. So, we can conclude that, for every $k \geqslant 0$, we have

$$\exp(-\varepsilon) = f(\varepsilon) = \left(\sum_{i=0}^{k} \frac{(-\varepsilon)^i}{i!}\right) + (-1)^{k+1} \frac{\exp(-\theta\varepsilon)\varepsilon^{k+1}}{(k+1)!}$$

If k is even, then the remainder

$$(-1)^{k+1} \frac{\exp(-\theta\varepsilon)\varepsilon^{k+1}}{(k+1)!}$$

is negative. So, for even k, we conclude that

$$\exp(-\varepsilon) \leqslant \sum_{i=0}^{k} \frac{(-\varepsilon)^i}{i!}$$

On the other hand, if k is odd, then the remainder

$$(-1)^{k+1}\frac{\exp(-\theta\varepsilon)\varepsilon^{k+1}}{(k+1)!}$$

is positive. So, for odd k, we conclude that

$$\exp(-\varepsilon) \geqslant \sum_{i=0}^{k} \frac{(-\varepsilon)^i}{i!}$$

5.2 Approximating $\ln(1-x)$ using polynomials

Let $f(x) = \ln(1-x)$ and $a = 0$. Note that $f^{(i)}(x) = -\frac{(i-1)!}{(1-x)^i}$, for $i \geqslant 1$. When approximating $\ln(1-\varepsilon)$, the remainder

$$-\frac{\varepsilon^{k+1}}{(k+1)(1-\theta\varepsilon)^{k+1}}$$

is always negative. So, for all $k \geqslant 0$, we get that

$$\ln(1-\varepsilon) \leqslant \sum_{i=1}^{k} \frac{-\varepsilon^i}{i}$$

It is left as an **exercise** to prove that for $0 \leqslant \varepsilon \leqslant 1/2$, the following lower-bound holds

$$\ln(1-\varepsilon) \geqslant -\frac{-\varepsilon^k}{k} + \sum_{i=1}^{k} \frac{-\varepsilon^i}{i}$$