Lecture 01: Mathematical Inequalities

In today's lecture, we shall cover some techniques to prove fundamental mathematical inequalities. We shall rely on the Lagrange form of the Taylor's Remainder Theorem to prove these results. We emphasize that we shall not prove the theorem itself. The course website provides an additional resource that presents the proof of this result. Interested students are encouraged to go over that proof.

Overview II

We shall, use this theorem to prove the following mathematical inequalities.

1. Jensen's Inequality,
   1. AM-GM-HM Inequality
   2. Cauchy-Schwarz Inequality
   3. Young's Inequality
   4. Hölder's Inequality
2. Approximating $\exp(-x)$ and $\ln(1-x)$ using polynomials, and
3. (In the future, we shall cover) Bonami-Beckner-Gross Hypercontractivity Inequality

Lagrange Form of the Taylor's Remainder Theorem I

Let us begin by recalling the Taylor's Theorem

Theorem (Taylor's Theorem)

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

For example

1. Using $f(x) = \exp(-x)$ and $a = 0$, we get

$$\exp(-\varepsilon) = 1 - \frac{\varepsilon}{1!} + \frac{\varepsilon^2}{2!} - \frac{\varepsilon^3}{3!} + \cdots$$

2. Using $f(x) = \ln(1-x)$ and $a = 0$, we get

$$\ln(1-\varepsilon) = -\frac{\varepsilon}{1} - \frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3} - \cdots$$

Motovation. Suppose we truncate the infinite Taylor series at the $f^{(k)}\frac{\varepsilon^k}{k!}$ term.

1. Is the truncated series an "overestimation" or an "underestimation"?
2. How good is the quality of approximation?

The Lagrange form of the Taylor Remainder Theorem will help answer this question.

Lagrange Form of the Taylor's Remainder Theorem III

Theorem (Lagrange Form of the Taylor Remainder Theorem)

For every a and ε, there exists $\theta \in (0,1)$ such that

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

We refer to the term $R = f^{(k+1)}(a+\theta\varepsilon)\frac{\varepsilon^{k+1}}{(k+1)!}$ as the remainder.

- If the remainder is positive, then the truncation is an underestimation. If the remainder is negative, then the truncation is an overestimation.
- The absolute value of the remainder determines the quality of approximation.

Example Problems I

Problem 1. Let $f(x) = \exp(-x)$ and $a = 0$. For $k \geqslant 0$, define $p_k(\varepsilon) = \sum_{i=0}^{k} \frac{(-\varepsilon)^i}{i!}$.
For example, we have $p_0(\varepsilon) = 1$, $p_1(\varepsilon) = 1 - \varepsilon$, $p_2(\varepsilon) = 1 - \varepsilon + \varepsilon^2/2$, and $p_3(\varepsilon) = 1 - \varepsilon + \varepsilon^2/2 - \varepsilon^3/6$, and so on.
For $0 \leqslant \varepsilon \leqslant 1$, apply the Taylor's Remainder Theorem to deduce the following.

1. If k is odd then we have $\exp(-\varepsilon) \geqslant p_k(\varepsilon)$.
2. If k is even then we have $\exp(-\varepsilon) \leqslant p_k(\varepsilon)$.
3. Prove that the absolute value of the remainder when we estimate $\exp(-\varepsilon)$ by $p_k(\varepsilon)$ is at most $\varepsilon^{k+1}/(k+1)!$.

Use the code at Desmos to experiment and develop intuition.

Problem 2. Let $f(x) = \ln(1-x)$ and $a = 0$. For $k \geqslant 0$, define $p_k(\varepsilon) = \sum_{i=1}^{k} \frac{-\varepsilon^i}{i}$.
For example $p_0(\varepsilon) = 0$, $p_1(\varepsilon) = -\varepsilon$, $p_2(\varepsilon) = -\varepsilon - \varepsilon^2/2$, $p_3(\varepsilon) = -\varepsilon - \varepsilon^2/2 - \varepsilon^3/3$, and so on.
For $0 \leqslant \varepsilon \leqslant 1$, apply the Taylor's Remainder Theorem to deduce the following.

1. We have $\ln(1-\varepsilon) \leqslant p_k(\varepsilon)$, for all $k \geqslant 0$.
2. What is the magnitude of the remainder?
3. How will you get a lower bound of $\ln(1-\varepsilon)$?

Use the code at Desmos to experiment and develop intuition.

A High-level Intuitive Summary

- We are using polynomials to estimate any function f
- The "behavior of f" at $(a + \varepsilon)$ is guided by the "properties of f" at the point a!

Definition (Convex Function)

A function f is convex in the range $[a, b]$ if $f^{(2)}$ is positive in $[a, b]$.

For example, the following functions are convex

1. $f(x) = x^2$
2. $f(x) = \exp(x)$
3. $f(x) = \exp(-x)$
4. $f(x) = 1/x$, in $(0, \infty)$

Think: How to define convexity of functions of multiple variables?

Jensen's Inequality, intuitively, states the following. Suppose f is a convex function. The secant joining any two points on the curve of f lies above the curve of f.

Theorem (Jensen's Inequality)

For a convex f, we have

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

Equality holds if and only if $a = b$.

In general, if $\mathbb{X}$ is a probability distribution over a sample space Ω then

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

Jensen's Inequality II

- We can use the Lagrange Form of the Taylor's remainder theorem to prove the Jensen's inequality
- A function f is concave if the function $-f$ is convex. For example, the function $\ln x$, $\ln(1-x)$ in the range $[0,1)$, $\sqrt{x}$ in the range $[0,\infty)$, and $1/x$, in the range $(-\infty,0)$ are concave function.
- **Think:** What is Jensen's inequality for concave functions?

Example

- Suppose $f(x) = x^2$. Note that f is convex.
- So, we get the following inequality. For all a, b, we have

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

Equality holds if and only if $a = b$.

Example Problems I

Use Jensen's Inequality to prove the following mathematical inequalities.

1. **AM-GM Inequality.** For positive a, b, we have

$$\frac{a+b}{2} \geqslant \sqrt{ab}$$

   Equality holds if and only if $a = b$.
   Consider the function $f(x) = \ln x$ to prove this inequality.

2. **Cauchy-Schwarz Inequality.** For positive a_1, a_2, b_1, b_2, we have

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

   Equality holds if and only if $\frac{a_1}{b_1} = \frac{a_2}{b_2}$.
   Consider the function $f(x) = \ln\left(1 + \exp(x)\right)$.

Example Problems II

3. **Young's Inequality.** Let $p, q \geqslant 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$. Such a pair of p and q is referred to as Hölder conjugates. For positive a, b, we have

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

   Equality holds if and only if $a^p = b^q$.
   Consider the function $f(x) = \ln x$.

4. **Hölder's Inequality.** For Hölder conjugates p and q, the following holds for positive a_1, a_2, b_1, b_2.

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

   What is the equality characterization? What function $f(x)$ will you consider?

Examples

Example 1: Bound exp using Polynomials I

- Our objective is to bound $\exp(-x)$ using polynomials in x when x is in the set $[0, 1]$. We shall use Lagrange form of the Taylor's Remainder Theorem to prove these bounds
- First, let us recall what the Lagrange's form of the Taylor's Remainder Theorem states. Suppose f is a "well-behaved" function. Let $f^{(i)}$ represent the i-th derivative of f (here $f^{(0)}$ represents the function f itself). For any choice of a, k, ε, there exists $\theta \in [0, 1]$ such that the following identity holds

$$f(a+\varepsilon) = \overbrace{\left(\sum_{i=0}^{k} f^{(k)}(a)\frac{\varepsilon^k}{k!}\right)}^{\text{Estimate}} + \overbrace{f^{(k+1)}(a+\theta\varepsilon)\frac{\varepsilon^{k+1}}{(k+1)!}}^{\textit{Remainder}}$$

We emphasize that the value of θ depends on the values of a, k, ε. The sign of the remainder determines whether the estimate is an overestimation or an underestimation of the value $f(\varepsilon)$.

- As a corollary, when $a = 0$, the above statement yields the following result. For any choice of k, ε, there exists $\theta \in [0, 1]$ such that

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

- We shall use $f(x) = \exp(-x)$
- Claim: $f^{(i)}(x) = (-1)^i \exp(-x)$ (you can use induction to prove this claim)
- So, we have $f^{(i)}(0) = (-1)^i$

Example 1: Bound exp using Polynomials III

- **Case** $k = 1$. Let us apply the Lagrange form of the Taylor's Remainder theorem to $f(x) = \exp(-x)$, and for the choice of $a = 0$ and $k = 1$. So, for every ε there exists $\theta \in [0, 1]$ such that

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

This expression is equivalent to

$$\exp(-\varepsilon) = 1 - \varepsilon + \overbrace{\exp(-\theta\varepsilon)\frac{\varepsilon^2}{2!}}^{\text{Remainder}}$$

Note that the remainder is positive. So, we have $\exp(-\varepsilon) \geqslant 1 - \varepsilon$.
We have our first underestimation of $\exp(-x)$ using polynomials in x.

Example 1: Bound exp using Polynomials IV

- **Case** $k = 2$**.** Let us use $k = 2$ now. So, for every ε there exists $\theta \in [0, 1]$ such that

$$f(\varepsilon) = f(0) + f^{(1)}(0)\frac{\varepsilon}{1!} + f^{(2)}(\varepsilon)\frac{\varepsilon^2}{2!} + f^{(3)}(\theta\varepsilon)\frac{\varepsilon^3}{3!}$$

This expression is equivalent to

$$\exp(-\varepsilon) = 1 - \varepsilon + \varepsilon^2/2 - \exp(-\theta\varepsilon)\frac{\varepsilon^3}{3!}$$

Note that the remainder is negative. So, we have
$\exp(-\varepsilon) \leqslant 1 - \varepsilon + \varepsilon^2/2$
We have our first overestimation of $\exp(-x)$ using polynomials in x.

- In general, if k is odd we get an underestimation

$$\exp(-\varepsilon) \geqslant 1 - \varepsilon + \varepsilon^2/2 - \cdots - \varepsilon^k/k!$$

If k is even, we get the overestimation

$$\exp(-\varepsilon) \leqslant 1 - \varepsilon + \varepsilon^2/2 - \cdots + \varepsilon^k/k!$$

Example 2: AM-GM Inequality I

- Our objective is to prove the AM-GM inequality using Jensen's Inequality
- Let us recall the AM-GM inequality. In the simplest form, it states that for any $a, b \geqslant 0$, we have

$$\frac{a+b}{2} \geqslant \sqrt{ab},$$

and equality holds if and only if $a = b$. Note that this statement already implies that the inequality if "strict" if $a \neq b$.

- In general, let $a_1, \ldots a_n \geqslant 0$ be n real numbers. Let $p_1, \ldots, p_n$ define a probability distribution (this implies that $p_i \geqslant 0$ and $\sum_{i=1}^n p_i = 1$). The general AM-GM inequality states that

$$\sum_{i=1}^{n} p_i a_i \geqslant \prod_{i=1}^{n} a_i^{p_i}$$

Furthermore, equality holds if and only if $a_1 = a_2 = \cdots = a_n$. Note that the simplest form of the AM-GM inequality is the restriction of this statement to $n = 2$ and $p_1 = p_2 = 1/2$.

- Let us try to "play around" with the AM-GM inequality to find the appropriate function f on which we shall apply the Jensen's inequality. We need to prove

$$\sum_{i=1}^{n} p_i a_i \geqslant \prod_{i=1}^{n} a_i^{p_i}$$

Note that we can write a_i as $\exp(\ln(a_i))$. So, the AM-GM inequality is equivalent to proving

$$\sum_{i=1}^{n} p_i a_i \geqslant \prod_{i=1}^{n} a_i^{p_i} = \prod_{i=1}^{n} \exp(\ln(a_i))^{p_i} = \prod_{i=1}^{n} \exp(p_i \ln(a_i)) = \exp\left(\sum_{i=1}^{n} p_i \ln(a_i)\right)$$

Example 2: AM-GM Inequality IV

- Since ln is monotone, we can take ln on both sides and it is equivalent to proving

$$\ln\left(\sum_{i=1}^{n} p_i a_i\right) \geqslant \sum_{i=1}^{n} p_i \ln(a_i)$$

- Look, now the inequality that we need to prove involves expressions of the form

$$f\left(\sum_{i=1}^{n} p_i a_i\right) \text{ and } \sum_{i=1}^{n} p_i f(a_i)$$

- So, we apply Jensen's Inequality to the function $f(x) = \ln(x)$ (which is convex downwards) and obtain the inequality. Equality holds if and only if all points coincide, that is, $a_1 = a_2 = \cdots = a_n$.

Example 3: Cauchy–Schwarz Inequality I

- Suppose we have $a_1, \ldots, a_n, b_1, \ldots, b_n \neq 0$. Cauchy–Schwarz inequality states that

$$\left|\sum_{i=1}^{n} a_i b_i\right| \leqslant \left(\sum_{i=1}^{n} a_i^2\right)^{1/2} \left(\sum_{i=1}^{n} b_i^2\right)^{1/2}$$

  And, inequality holds if and only if $\frac{a_1}{b_1} = \frac{a_2}{b_2} = \cdots = \frac{a_n}{b_n}$.
- As in the previous example we shall manipulate the Cauchy–Schwarz inequality into an equivalent inequality that we can prove using the Jensen's inequality. However, this manipulation is tricky in this case. The first hint regarding what points we should be using is given by the equality condition, which states that $\frac{a_i}{b_i}$ is constant. So, we should try to rewrite the Cauchy–Schwarz inequality so that the expression $\frac{a_i}{b_i}$ shows up.

Example 3: Cauchy–Schwarz Inequality II

- The Cauchy–Schwarz inequality is equivalent to

$$\left|\sum_{i=1}^{n} b_i^2 \cdot \left(\frac{a_i}{b_i}\right)\right| \leqslant \left(\sum_{i=1}^{n} b_i^2 \cdot \left(\frac{a_i}{b_i}\right)^2\right)^{1/2} \left(\sum_{i=1}^{n} b_i^2\right)^{1/2}$$

- Note that the left hand side has the points as $\frac{a_i}{b_i}$. However, there is a slight problem. The corresponding coefficients b_i^2 do not define a probability (although the values are positive, they might not add up to 1). So, we divide both sides of the expression by $B = \sum_{j=1}^{n} b_j^2$. This manipulation, yields the following equivalent expression

$$\left|\sum_{i=1}^{n} \frac{b_i^2}{B} \cdot \left(\frac{a_i}{b_i}\right)\right| \leqslant \frac{1}{B}\left(\sum_{i=1}^{n} b_i^2 \cdot \left(\frac{a_i}{b_i}\right)^2\right)^{1/2} \sqrt{B} = \left(\sum_{i=1}^{n} \frac{b_i^2}{B} \cdot \left(\frac{a_i}{b_i}\right)^2\right)^{1/2}$$

Example 3: Cauchy–Schwarz Inequality III

- Let us define $p_i = b_i^2/B$ and $x_i = a_i/b_i$. This substitution makes the Cauchy–Schwarz inequality equivalent to

$$\left|\sum_{i=1}^{n} p_i x_i\right| \leqslant \left(\sum_{i=1}^{n} p_i x_i^2\right)^{1/2}$$

- Both sides of the inequality are positive, so we can square both sides and get an equivalent inequality

$$\left(\sum_{i=1}^{n} p_i x_i\right)^2 \leqslant \sum_{i=1}^{n} p_i x_i^2$$

- If we prove the above inequality then we have proven Cauchy–Schwarz inequality. We shall use $f(x) = x^2$ (convex upwards function) and apply Jensen's inequality to prove this inequality. Furthermore, equality holds if and only if all points $x_i = a_i/b_i$ are identical.
- **Exercise:** Prove the Hölder's inequality that states the following. Let $a_1, \ldots, a_n, b_1, \ldots, b_n > 0$. Let p, q be positive reals such that $\frac{1}{p} + \frac{1}{q} = 1$.

$$\sum_{i=1}^{n} a_i b_i \leqslant \left(\sum_{i=1}^{n} a_i^p \right)^{1/p} \left(\sum_{i=1}^{n} b_i^q \right)^{1/q}$$

Equality holds if and only if a_i^p / b_i^q is identical for all $i \in \{1, \ldots, n\}$.