

# MATH 4318: Analysis II

Frank Qiang  
Instructor: Zhiwu Lin

Georgia Institute of Technology  
Spring 2024

# Contents

<b>1</b>	<b>Jan. 9 — The Derivative</b>	<b>3</b>
1.1	Defining the Derivative . . . . .	3
1.2	Immediate Properties . . . . .	3
1.3	Rules for Differentiation . . . . .	4
<b>2</b>	<b>Jan. 11 — The Mean Value Theorem</b>	<b>6</b>
2.1	The Mean Value Theorem . . . . .	6
2.2	Applications . . . . .	8
2.3	Cauchy's Mean Value Theorem . . . . .	9
<b>3</b>	<b>Jan. 16 — Taylor's Theorem</b>	<b>10</b>
3.1	Darboux's Lemma . . . . .	10
3.2	L'Hôpital's Rule . . . . .	10
3.3	Taylor's Theorem . . . . .	12
<b>4</b>	<b>Jan. 18 — Taylor Polynomials</b>	<b>14</b>
4.1	Common Taylor Polynomials . . . . .	14
4.2	Combining Taylor Polynomials . . . . .	14
4.3	Applications for Taylor Polynomials . . . . .	16
4.3.1	Finding Limits . . . . .	16
4.3.2	Estimation . . . . .	17
<b>5</b>	<b>Jan. 23 — The Riemann Integral</b>	<b>19</b>
5.1	The Anti-Derivative . . . . .	19
5.2	The Riemann Integral . . . . .	20
5.3	Properties of the Riemann Integral . . . . .	22
<b>6</b>	<b>Jan. 25 — Riemann Integrability</b>	<b>24</b>
6.1	Conditions for Integrability . . . . .	24
6.2	The Fundamental Theorem of Calculus . . . . .	27
<b>7</b>	<b>Jan. 30 — Riemann Integrability, Part 2</b>	<b>29</b>
7.1	Conditions for an Anti-Derivative . . . . .	29
7.2	More Conditions for Integrability . . . . .	31
<b>8</b>	<b>Feb. 1 — Riemann Integrability, Part 3</b>	<b>33</b>
8.1	Even More Conditions for Integrability . . . . .	33
<b>9</b>	<b>Feb. 6 — Exchange of Limit Operations</b>	<b>37</b>
9.1	Motivation . . . . .	37
9.2	Exchange of the Limit and Integral . . . . .	37

9.3	Exchange of the Limit and Derivative . . . . .	38
9.4	Infinite Series . . . . .	39
<b>10 Feb. 8</b>	<b>— Infinite Series</b>	<b>42</b>
10.1	Lots of Convergence Tests . . . . .	42
10.1.1	The Comparison Test . . . . .	42
10.1.2	The Root Test . . . . .	44
<b>11 Feb. 15</b>	<b>— Absolute Convergence</b>	<b>46</b>
11.1	Example of the Root Test . . . . .	46
11.2	The Integral Test . . . . .	46
11.3	Absolute and Conditional Convergence . . . . .	49
11.3.1	The Alternating Series Test . . . . .	50
<b>12 Feb. 20</b>	<b>— Series of Functions</b>	<b>51</b>
12.1	Rearrangements of Series . . . . .	51
12.2	Series of Functions . . . . .	52
12.3	Exchange of the Limit and Infinite Series . . . . .	54
<b>13 Feb. 22</b>	<b>— Power Series</b>	<b>56</b>
13.1	Power Series . . . . .	56
13.2	Radius of Convergence . . . . .	57
<b>14 Feb. 27</b>	<b>— Taylor Series</b>	<b>61</b>
14.1	Taylor Series . . . . .	61
14.2	Convergence of Taylor Series . . . . .	62
14.3	Metric Spaces . . . . .	63
14.4	Existence and Uniqueness Problem for ODEs . . . . .	64
14.5	The Contraction Mapping Principle . . . . .	64
<b>15 Feb. 29</b>	<b>— Contraction Mapping</b>	<b>66</b>
15.1	Newton's Method . . . . .	66
15.2	Existence and Uniqueness for ODEs . . . . .	67
15.3	Implicit Function Theorem . . . . .	68
<b>16 Mar. 5</b>	<b>— The Derivative in <math>\mathbb{R}^n</math></b>	<b>69</b>
16.1	Partial Derivatives . . . . .	69
16.2	Differentiability . . . . .	70
<b>17 Mar. 7</b>	<b>— Vector-Valued Functions</b>	<b>74</b>
17.1	The Derivative for Vector-Valued Functions . . . . .	74
17.2	Implicit Function Theorem in $\mathbb{R}^n$ . . . . .	75
17.3	Higher Order Derivatives . . . . .	76
<b>18 Mar. 12</b>	<b>— More on the Implicit Function Theorem</b>	<b>77</b>
18.1	Example of Unequal Mixed Partial Derivatives . . . . .	77
18.2	Revisiting the Implicit Function Theorem . . . . .	77
18.3	Lagrange Multipliers . . . . .	78

# Lecture 1

## Jan. 9 — The Derivative

### 1.1 Defining the Derivative

**Definition 1.1.** Let  $f$  be a real-valued function on an open interval  $U \subseteq \mathbb{R}$ . Let  $x_0 \in U$ , we say  $f$  is *differentiable* at  $x_0$  if

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists. If it does, then this limit, denoted by  $f'(x_0)$ , is called the *derivative* of  $f$  at  $x_0$ .

**Remark.** By definition, for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| \leq \epsilon$$

if  $|x - x_0| < \delta$  and  $x \in U$ . Multiplying both sides by  $|x - x_0|$  yields

$$|f(x) - f(x_0) - f'(x_0)(x - x_0)| \leq \epsilon|x - x_0|.$$

In other words,

$$|f(x) - \varphi(x)| \leq \epsilon|x - x_0|$$

where  $\varphi(x) = f(x_0) + f'(x_0)(x - x_0)$ . In other words,  $\varphi(x)$  is a first-order approximation of  $f(x)$  near  $x_0$ . Geometrically, this is approximating the graph of  $y = f(x)$  by the tangent line  $y = \varphi(x)$ .

### 1.2 Immediate Properties

**Proposition 1.1.** Let  $U \subseteq \mathbb{R}$  be an open set and  $f : U \rightarrow \mathbb{R}$ . If  $f$  is differentiable at  $x_0 \in U$ , then  $f$  is continuous at  $x_0$ .

*Proof.* Pick any  $\epsilon_0 > 0$ . Then there exists  $\delta_0 > 0$  such that whenever  $|x - x_0| < \delta_0$  and  $x \in U$ ,

$$|f(x) - f(x_0) - f'(x_0)(x - x_0)| \leq \epsilon_0|x - x_0|.$$

By the triangle inequality,

$$|f(x) - f(x_0)| \leq \epsilon_0|x - x_0| + |f'(x_0)||x - x_0| = (\epsilon_0 + |f'(x_0)|)|x - x_0|.$$

Now for any  $\epsilon > 0$ , choose  $\delta = \min\{\delta_0, \epsilon/(\epsilon_0 + |f'(x_0)|)\}$ . Then

$$|f(x) - f(x_0)| \leq (\epsilon_0 + |f'(x_0)|)|x - x_0| < (\epsilon_0 + |f'(x_0)|)\delta \leq \epsilon$$

whenever  $|x - x_0| < \delta$  and  $x \in U$ . Thus  $f$  is continuous at  $x_0$ . □

**Example 1.1.1.** Take the function

$$f(x) = \begin{cases} x \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Note that  $f$  is continuous on  $\mathbb{R}$ . For  $x \neq 0$ , continuity is clear since both  $x$  and  $\sin(1/x)$  are continuous. At  $x = 0$ , we have

$$\lim_{x \rightarrow 0} f(x) = \lim_{x \rightarrow 0} x \sin(1/x) = 0 = f(0)$$

since  $|x \sin(1/x)| \leq |x|$  for all  $x \in \mathbb{R}$ , so  $f$  is also continuous at  $x = 0$ . However,  $f$  is not differentiable at  $x = 0$ . Consider the limit

$$\lim_{x \rightarrow 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \rightarrow 0} \sin(1/x),$$

which does not exist since  $\sin(1/x)$  oscillates. So  $f$  is not differentiable at  $x = 0$ .

**Example 1.1.2.** Take the function  $f(x) = |x|$ , which is continuous everywhere on  $\mathbb{R}$ . However,  $f$  is not differentiable at  $x = 0$ , since

$$\lim_{x \rightarrow 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \rightarrow 0} \frac{|x|}{x}.$$

Note that

$$\frac{|x|}{x} = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0, \end{cases}$$

so the limit does not exist as  $x \rightarrow 0$ . Thus  $f$  is not differentiable at  $x = 0$ .

**Remark.** For the previous example, we can however define the *left (right) derivative* by

$$f'_-(x_0) = \lim_{x \rightarrow x_0^-} \frac{f(x) - f(x_0)}{x - x_0} \quad \text{and} \quad f'_+(x_0) = \lim_{x \rightarrow x_0^+} \frac{f(x) - f(x_0)}{x - x_0}.$$

If  $f$  is differentiable, then  $f'_-(x_0) = f'_+(x_0)$ . In the previous example,  $f'_-(0) = -1$  and  $f'_+(0) = 1$ . For the first example however, even  $f'_\pm(0)$  does not exist.

**Remark.** In one dimension, the existence of the derivative implies that the function is differentiable (the function is approximated by a linear function). However, in multiple dimensions, the existence of partial derivatives does not imply differentiability.

## 1.3 Rules for Differentiation

**Proposition 1.2.** Let  $U \subseteq \mathbb{R}$  be open and  $f, g : U \rightarrow \mathbb{R}$  be differentiable. Then

1.  $(f + g)'(x_0) = f'(x_0) + g'(x_0)$
2.  $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$
3. if  $g(x_0) \neq 0$ , then  $(f/g)'(x_0) = (f'(x_0)g(x_0) - f(x_0)g'(x_0))/(g(x_0)^2)$ .

*Proof.* Find in textbook (Rosenlicht). □

**Proposition 1.3.** We have  $\frac{d}{dx}(c) = 0$ ,  $\frac{d}{dx}(x) = 1$ , and  $\frac{d}{dx}(x^n) = nx^{n-1}$  for all  $n \in \mathbb{N}$ .

*Proof.* We prove the last claim (the power rule) for  $n \geq 1$  by induction. The base case  $n = 1$  is the first claim which is true. Now suppose that the result holds for any  $n \leq k \in \mathbb{N}$ , and we show that it remains true for  $n = k + 1$ . By the product rule, we have

$$\frac{d}{dx}(x^{k+1}) = \frac{d}{dx}(x \cdot x^k) = \frac{d}{dx}(x) \cdot x^k + x \cdot \frac{d}{dx}(x^k) = x^k + xkx^{k-1} = (k+1)x^k.$$

Thus by induction this holds for all  $n \geq 1$ . We can do negative integers by the quotient rule.  $\square$

**Remark.** The power rule actually holds for any  $n \in \mathbb{R}$ .

**Proposition 1.4** (Chain rule). *Let  $U$  and  $V$  be open sets of  $\mathbb{R}$  and let  $f : U \rightarrow V, g : V \rightarrow \mathbb{R}$  be differentiable. Let  $x_0 \in U$  be such that  $f'(x_0)$  and  $g'(f(x_0))$  exist. Then  $(g \circ f)'(x_0)$  exists and*

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

*Proof.* For any fixed  $y_0$  for which  $g'(y_0)$  exists, set

$$A(y, y_0) = \begin{cases} (g(y) - g(y_0))/(y - y_0) & \text{if } y \in V \text{ and } y \neq y_0 \\ g'(y_0) & \text{if } y = y_0. \end{cases}$$

Then  $A$  is continuous at  $y_0$ . To find  $(g \circ f)'(x_0)$ , observe that

$$\begin{aligned} \lim_{x \rightarrow x_0} \frac{g(f(x)) - g(f(x_0))}{x - x_0} &= \lim_{x \rightarrow x_0} \frac{A(f(x), f(x_0))(f(x) - f(x_0))}{x - x_0} \\ &= \lim_{x \rightarrow x_0} A(f(x), f(x_0)) \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = g'(f(x_0))f'(x_0), \end{aligned}$$

by the continuity of  $A$  at  $f(x_0)$  and the differentiability of  $f$  at  $x_0$ .  $\square$

**Remark.** The rough idea of what we did here is

$$\begin{aligned} \lim_{x \rightarrow x_0} \frac{g(f(x)) - g(f(x_0))}{x - x_0} &= \lim_{x \rightarrow x_0} \frac{g(f(x)) - g(f(x_0))}{f(x) - f(x_0)} \frac{f(x) - f(x_0)}{x - x_0} \\ &= \lim_{x \rightarrow x_0} \frac{g(f(x)) - g(f(x_0))}{f(x) - f(x_0)} \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = g'(f(x_0))f'(x_0). \end{aligned}$$

But does not quite work as stated since it might be that  $f(x) = f(x_0)$  even if  $x \neq x_0$ . We can fix this by introducing the function  $A$  as we did in the proof, though the overall idea is the same.

**Remark.** If  $f$  is monotone near  $x_0$ , then we can define the *inverse function*  $f^{-1}$  so that  $(f^{-1} \circ f)(x) = x$  near  $x_0$ . If  $f'(x_0)$  exists, then by the chain rule applied to  $x = (f^{-1} \circ f)(x)$  at  $x = x_0$  we have

$$1 = \frac{d}{dx}(f^{-1} \circ f)(x_0) = \frac{d}{dx}f^{-1}(f(x_0)) \cdot f'(x_0) \implies \frac{d}{dx}f^{-1}(f(x_0)) = \frac{1}{f'(x_0)}.$$

**Example 1.1.3.** Let  $f(x) = e^x$  with  $f^{-1}(x) = \ln(x)$ . Since  $f'(x) = f(x) = e^x$ , we have

$$\frac{d}{dx}f^{-1}(f(x_0)) = \frac{1}{f'(x_0)} \implies \frac{d}{dx}\ln(e^{x_0}) = \frac{1}{e^{x_0}}.$$

Letting  $e^{x_0} = h$ , we have  $\frac{d}{dx}\ln(x)|_{x=h} = 1/h$ , which recovers the familiar formula.

# Lecture 2

## Jan. 11 — The Mean Value Theorem

### 2.1 The Mean Value Theorem

**Lemma 2.1.** *Let  $I \subseteq \mathbb{R}$  be open,  $f : I \rightarrow \mathbb{R}$  is differentiable at  $x_0 \in I$  and  $f'(x_0) \neq 0$ . Suppose  $f'(x_0) > 0$ , then there exists  $\delta > 0$  such that for any  $x \in (x_0 - \delta, x_0 + \delta)$ ,*

1. *if  $x > x_0$ , then  $f(x) > f(x_0)$ ,*
2. *if  $x < x_0$ , then  $f(x) < f(x_0)$ .*

*Proof.* Take  $\epsilon = f'(x_0)/2$ . By the definition of the derivative, there exists  $\delta > 0$  such that for any  $|x - x_0| < \delta$ , we have

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| < \epsilon = \frac{1}{2}f'(x_0).$$

By the triangle inequality,

$$\frac{f(x) - f(x_0)}{x - x_0} > \frac{1}{2}f'(x_0) > 0.$$

This quotient being positive immediately implies the desired results.  $\square$

**Theorem 2.1.** *If  $f(x)$  is differentiable in an open interval  $I$  and  $f$  obtains its local maximum (or minimum) at  $x_0 \in I$ , then  $f'(x_0) = 0$ .*

*Proof.* Suppose otherwise that  $f'(x_0) \neq 0$ . Assume without loss of generality that  $f'(x_0) > 0$ . Then by the previous lemma, there exists  $\delta > 0$  such that for  $x \in (x_0 - \delta, x_0 + \delta)$ , if  $x > x_0$  then  $f(x) > f(x_0)$  and if  $x < x_0$  then  $f(x) < f(x_0)$ . So  $x_0$  cannot be a local maximum or minimum, which is a contradiction.  $\square$

**Theorem 2.2** (Rolle's middle value theorem). *Let  $f(x)$  be continuous on  $[a, b]$  and differentiable in  $(a, b)$ . Suppose  $f(a) = f(b)$ , then there exists  $x_0 \in (a, b)$  such that  $f'(x_0) = 0$ .*

*Proof.* Since  $f$  is continuous on a compact set, it obtains both a maximum and minimum on  $[a, b]$ . Let  $M$  be the maximum and  $m$  be the minimum. If  $M = m$ , then  $f(x) \equiv M$  and  $f'(x) = 0$  everywhere. If  $M > m$ , then at least one of the maximum or minimum must be obtained at an interior point  $x_0 \in (a, b)$  since  $f(a) = f(b)$ . By the previous theorem,  $f'(x_0) = 0$  at this point and we are done.  $\square$

**Example 2.0.1.** Show that the equation  $4ax^3 + 3bx^2 + 2cx = a + b + c$  has at least one root in  $(0, 1)$ .

*Proof.* Consider the equation

$$4ax^3 + 3bx^2 + 2cx - (a + b + c) = 0.$$

Notice that the left hand side is the derivative of the function

$$f(x) = ax^4 + bx^3 + cx^2 - (a + b + c)x.$$

So we just need to show that  $f'(x) = 0$  for some  $x$ . For this, we can check that  $f(0) = f(1) = 0$ , and thus by Rolle's theorem there exists  $x_0 \in (0, 1)$  such that  $f'(x_0) = 0$ . So  $x_0$  is a root.  $\square$

**Theorem 2.3** (Lagrange's middle value theorem). *Let  $f(x)$  be continuous on  $[a, b]$  and differentiable in  $(a, b)$ . Then there exists  $x_0 \in (a, b)$  such that*

$$f'(x_0) = \frac{f(b) - f(a)}{b - a}.$$

*Proof.* Subtract the secant line through  $(a, f(a))$  and  $(b, f(b))$  from  $f(x)$  to get

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

Note that  $g(a) = g(b) = f(a)$ . So by Rolle's theorem, there exists  $x_0 \in (a, b)$  such that  $g'(x_0) = 0$ . But

$$0 = g'(x_0) = f'(x_0) - \frac{f(b) - f(a)}{b - a},$$

which is the desired result.  $\square$

**Corollary 2.3.1.** *Suppose  $f \in C([a, b])$ , i.e.  $f$  is continuous on  $[a, b]$ , and that  $f$  is differentiable in  $(a, b)$ . Then the following statements are equivalent:*

1.  $f'(x) \geq 0$  in  $(a, b)$ ,
2.  $f(x)$  is increasing, i.e. if  $x_1 > x_2$ , then  $f(x_1) \geq f(x_2)$ .

*In particular, if  $f'(x) > 0$  in  $(a, b)$ , then  $f(x)$  is strictly increasing, i.e. if  $x_1 > x_2$ , then  $f(x_1) > f(x_2)$ .*

*Proof.*  $(2 \Rightarrow 1)$  For any  $x_0 \in (a, b)$ ,

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} \geq 0$$

since  $f(x_0 + h) - f(x_0) \geq 0$  for  $h > 0$  as  $f$  is increasing.

$(1 \Rightarrow 2)$  Take  $x_1 > x_2$ , then by Lagrange's theorem there exists  $\xi \in (x_2, x_1)$  such that

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) \geq 0.$$

So  $f(x_1) \geq f(x_2)$ . The strict version follows from changing the above inequality to a strict one.  $\square$



## 2.2 Applications

**Example 2.0.2.** Show that

$$\frac{2}{2x+1} < \ln(1 + 1/x)$$

for any  $x > 0$ .

*Proof.* Let  $f(x) = 2/(2x+1) - \ln(1 + 1/x)$ . Taking the derivative yields

$$f'(x) = \frac{1}{(2x+1)^2 x(x+1)} > 0,$$

so  $f$  is strictly increasing in  $(0, \infty)$ . Note that  $f \rightarrow 0$  as  $x \rightarrow \infty$ , so  $f(x) < 0$  for all  $x > 0$ . □

**Example 2.0.3.** Show that  $b/a > b^a/a^b$  when  $b > a > 1$ .

*Proof.* Take log on both sides to get  $\ln b - \ln a > a \ln b - b \ln a$ . This gives

$$(b-1) \ln a > (a-1) \ln b \iff \frac{\ln a}{a-1} > \frac{\ln b}{b-1}.$$

Note that this is a monotonicity property. So let  $f(x) = (\ln x)/(x-1)$  for  $x > 1$ . Then

$$f'(x) = \frac{x-1-x \ln x}{x(x-1)^2} < 0$$

when  $x > 1$  because  $x-1-x \ln x < 0$ . To see the last claim, define  $g(x) = x-1-x \ln x$  and note that  $g'(x) = -\ln x < 0$  for  $x > 1$ . But  $g(0) = 0$ , so  $g(x) < 0$  for  $x > 1$ . So  $f$  is strictly decreasing. □

**Example 2.0.4.** Show that

$$\lim_{x \rightarrow 0} \frac{e^x - e^{\sin x}}{x - \sin x} = 1.$$

*Proof.* Let  $f(x) = e^x$ . Then there exists  $\xi$  between  $x$  and  $\sin x$  such that

$$e^x - e^{\sin x} = (x - \sin x)e^{\xi(x)},$$

where the choice of  $\xi$  may vary for different  $x$ . Then

$$\lim_{x \rightarrow 0} \frac{e^x - e^{\sin x}}{x - \sin x} = \lim_{x \rightarrow 0} e^{\xi(x)}.$$

Now note that  $\xi(x)$  is always between  $x$  and  $\sin x$ , which both tend to 0 as  $x \rightarrow 0$ . So by the squeeze theorem we have  $\xi(x) \rightarrow 0$  as  $x \rightarrow 0$  and thus  $e^{\xi(x)} \rightarrow 1$  as  $x \rightarrow 0$ . □

## 2.3 Cauchy's Mean Value Theorem

**Theorem 2.4** (Cauchy's middle value theorem). *Let  $f, g \in C([a, b])$  and  $f, g$  be differentiable in  $(a, b)$ . Suppose  $g'(x) \neq 0$  for any  $x \in (a, b)$ . Then there exists  $x_0 \in (a, b)$  such that*

$$\frac{f'(x_0)}{g'(x_0)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

*Proof.* Use a similar construction as before and let

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

Note that  $F(b) = F(a) = 0$ , so by Rolle's theorem there exists  $x_0 \in (a, b)$  such that  $F'(x_0) = 0$ . Then

$$0 = F'(x_0) = f'(x_0) - \frac{f(b) - f(a)}{g(b) - g(a)}g'(x_0),$$

which implies the desired result. □

**Remark.** The  $g'(x) \neq 0$  condition guarantees that  $g$  is monotone, even if  $g'$  may fail to be continuous.

**Remark.** If  $g$  is a monotonically increasing function, we can view  $g$  as a mapping  $g : [a, b] \rightarrow [g(a), g(b)]$ , which we can view as a change of variables  $x \mapsto u$ . Since  $g$  is monotone, we have an inverse  $x = g^{-1}(u)$ . Then

$$f(x) = f(g^{-1}(u)) = (f \circ g^{-1})(u) = \tilde{f}(u).$$

By Lagrange's theorem,

$$\frac{\tilde{f}(g(b)) - \tilde{f}(g(a))}{g(b) - g(a)} = \tilde{f}'(u_0)$$

for some  $u_0 \in (g(a), g(b))$ . Now note that

$$\tilde{f}(g(b)) = (f \circ g^{-1})(g(b)) = f(b), \quad \tilde{f}(g(a)) = f(a).$$

So the left-hand side is precisely

$$\text{LHS} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

By the chain rule, we have

$$\text{RHS} = \tilde{f}'(u_0) = (f \circ g^{-1})'(u_0) = f'(g^{-1}(u_0))(g^{-1})'(u_0) = f'(x_0) \frac{1}{g'(x_0)}.$$

This recovers Cauchy's mean value theorem. So they are equivalent even if Cauchy's seems stronger.

# Lecture 3

## Jan. 16 — Taylor's Theorem

### 3.1 Darboux's Lemma

**Lemma 3.1** (Darboux's lemma). *If  $f$  is differentiable in  $(a, b)$ , continuous on  $[a, b]$  and  $f'(a) < f'(b)$ , then for any  $c \in (f'(a), f'(b))$ , there exists  $x_0 \in (a, b)$  such that  $f'(x_0) = c$ .*

*Proof.* See homework. □

**Remark.** There exists an example of a differentiable function  $f(x)$  but  $f'(x)$  is not continuous, e.g.

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

We can compute that

$$f'(x) = \begin{cases} 2x \sin(1/x) - \cos(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0, \end{cases}$$

and we can verify as an exercise that  $f'(x)$  is not continuous at  $x = 0$ .

**Remark.** Darboux's lemma guarantees that  $g'(x) \neq 0$  implies either  $g'(x) > 0$  or  $g'(x) < 0$  everywhere in the conditions for Cauchy's mean value theorem.

### 3.2 L'Hôpital's Rule

**Theorem 3.1** (L'Hôpital's rule,  $0/0$ ). *Let  $f, g$  be differentiable in  $(a, b)$ ,  $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = 0$ , and  $g'(x) \neq 0$  for any  $x \in (a, b)$ . Then if  $\lim_{x \rightarrow a^+} f'(x)/g'(x)$  exists, we have*

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}.$$

*Proof.* By Cauchy's theorem, for any  $x \in (a, b)$ , there exists  $\xi(x) \in (a, x)$  such that

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

If  $x \rightarrow a^+$ , then  $\xi(x) \rightarrow a^+$ , so

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a^+} \frac{f'(\xi(x))}{g'(\xi(x))} = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)},$$

as desired. □

**Corollary 3.1.1.** Let  $f, g$  be differentiable in  $(a, \infty)$ ,  $\lim_{x \rightarrow \infty} f(x) = \lim_{x \rightarrow \infty} g(x) = 0$ , and  $g'(x) \neq 0$  for any  $x \in (a, \infty)$ . Then if  $\lim_{x \rightarrow \infty} f'(x)/g'(x)$  exists, we have

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}.$$

*Proof.* Assume  $a > 0$ . Define  $\tilde{f}(y) = f(1/y)$  and  $\tilde{g}(y) = g(1/y)$  with  $y \in (0, 1/a)$ . By L'Hôpital's rule,

$$\lim_{y \rightarrow 0^+} \frac{\tilde{f}(y)}{\tilde{g}(y)} = \lim_{y \rightarrow 0^+} \frac{\tilde{f}'(y)}{\tilde{g}'(y)} = \lim_{y \rightarrow \infty} \frac{f'(1/y) \cdot (-1/y^2)}{g'(1/y) \cdot (-1/y^2)} = \lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)},$$

as desired. □

**Theorem 3.2** (L'Hôpital,  $\infty/\infty$ ). Let  $f, g$  be differentiable in  $(a, b)$ ,  $\lim_{x \rightarrow a^+} |f(x)| = \lim_{x \rightarrow a^+} |g(x)| = \infty$ , and  $g'(x) \neq 0$  for any  $x \in (a, b)$ . Then if  $\lim_{x \rightarrow a^+} f'(x)/g'(x)$  exists, we have

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}.$$

*Proof.* Left as an exercise. □

**Remark.** Saying that the absolute values of  $f$  and  $g$  go to infinity works, since the existence of the limit rules out oscillatory behavior.

**Remark.** These cases of  $\infty/\infty$  and  $0/0$  are called *indefinite types*. Other indefinite types include  $0 \cdot \infty$ ,  $0^0$ ,  $\infty^0$ ,  $1^\infty$ ,  $\infty - \infty$ , etc. But we can try to reduce them to the cases we know. For example, if  $f(x) \rightarrow 0^+$  and  $g(x) \rightarrow 0^+$  when  $x \rightarrow x_0$ , then  $\lim_{x \rightarrow x_0} f(x)^{g(x)}$  is  $0^0$ . Letting  $y(x) = f(x)^{g(x)}$ , we can take the log to get

$$\ln y(x) = g(x) \ln f(x) = \frac{\ln f(x)}{1/g(x)} = \frac{\infty}{\infty}.$$

**Example 3.0.1.** We can see that (this is a  $\infty - \infty$  case)

$$\lim_{x \rightarrow 0^+} \frac{1}{x^2} - \frac{\cot x}{x} = \lim_{x \rightarrow 0^+} \frac{1 + x \cot x}{x^2} = \lim_{x \rightarrow 0^+} \frac{-(\cot x - x \csc^2 x)}{2x \sin^2 x}.$$

Note that  $x \cot x = x \cos x / \sin x \rightarrow 1$  as  $x \rightarrow 0$ . Now note that  $\sin x / x \rightarrow 1$  as  $x \rightarrow 0$ , so we continue with

$$\lim_{x \rightarrow 0^+} \frac{-(\cot x - x \csc^2 x)}{2x \sin^2 x} = \lim_{x \rightarrow 0^+} \frac{x - \sin x \cos x}{2x^3} \frac{x^2}{\sin^2 x}$$

Since  $x^2 / \sin^2 x \rightarrow 1$  as  $x \rightarrow 0$ , we can look at the remaining part to get

$$\lim_{x \rightarrow 0^+} \frac{x - \sin x \cos x}{2x^3} = \lim_{x \rightarrow 0^+} \frac{1 - \cos 2x}{6x^2} = \lim_{x \rightarrow 0^+} \frac{2 \sin 2x}{12x} = \frac{1}{3}.$$

So  $\lim_{x \rightarrow 0^+} (1/x^2 - \cot x/x) = 1/3$ .

### 3.3 Taylor's Theorem

**Theorem 3.3** (Peano remainder term). *Let  $f : [a, b] \rightarrow \mathbb{R}$  be differentiable at  $x = a$  up to  $n$ th order of derivatives, i.e.  $f'(a), f''(a), \dots, f^{(n)}(a)$  exist. Then as  $x \rightarrow a^+$ , we have*

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + o((x-a)^n).$$

Call the polynomial part of the above  $P_n(x)$ , which is also known as the Taylor polynomial of order  $n$ .

*Proof.* To show that the error term is  $o((x-a)^n)$ , we have

$$\lim_{x \rightarrow a^+} \frac{f(x) - P_n(x)}{(x-a)^n} = \lim_{x \rightarrow a^+} \frac{f'(x) - P'_n(x)}{n(x-a)^{n-1}} = \frac{1}{n!} \lim_{x \rightarrow a^+} \left[ \frac{f^{n-1}(x) - f^{n-1}(a)}{x-a} - f^{(n)}(a) \right] = 0$$

by L'Hôpital's rule, where we used the observation that  $f^{(k)}(a) = P_n^{(k)}(a)$  for  $1 \leq k \leq n$ . The final step is a result of the existence of  $f^{(n)}(a)$ .  $\square$

**Lemma 3.2** (Rolle's theorem for higher order derivatives). *Let  $f \in C^n([a, b])$  and differentiable to  $(n+1)$  order. If  $f'(a) = \dots = f^{(n)}(a) = 0$  and  $f(a) = f(b)$ , then there exists  $x_0 \in (a, b)$  such that  $f^{(n+1)}(x_0) = 0$ .*

*Proof.* Since  $f(a) = f(b)$ , by the usual Rolle's theorem there exists  $x_1 \in (a, b)$  such that  $f'(x_1) = 0$ . Then since  $f'(a) = f'(x_1) = 0$ , by Rolle's theorem again, there exists  $x_2 \in (a, x_1)$  such that  $f''(x_2) = 0$ . Repeat this to get  $x_{n+1} \in (a, x_n) \subseteq (a, b)$  such that  $f^{(n+1)}(x_{n+1}) = 0$ . Take  $x_0 = x_{n+1}$  to finish.  $\square$

**Theorem 3.4** (Lagrange remainder term). *Let  $f \in C^n([a, b])$ , in particular,  $f'(a), \dots, f^{(n)}(a)$  exist. Additionally, assume  $f$  is  $(n+1)$ -th differentiable in  $(a, b)$ .<sup>1</sup> Then*

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + R_n(x) \quad \text{where} \quad R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-a)^{n+1}$$

for some  $\xi \in [a, x]$ .

*Proof.* Define  $P(x) = P_n(x) + \lambda(x-a)^{n+1}$ , where we choose  $\lambda \in \mathbb{R}$  such that  $P(b) = f(b)$ , i.e.

$$\lambda = \frac{f(b) - P_n(b)}{(b-a)^{n+1}}.$$

Consider  $g(x) = f(x) - P(x)$ , which satisfies  $g(a) = g(b) = 0$  and  $g'(a) = \dots = g^{(n)}(a) = 0$ . Then by Rolle's theorem (higher order), there exists  $\xi \in (a, b)$  such that  $g^{(n+1)}(\xi) = 0$ . In other words,

$$f^{(n+1)}(\xi) - P^{(n+1)}(\xi) = 0 \implies f^{(n+1)}(\xi) - \underbrace{(n+1)! \frac{f(b) - P_n(b)}{(b-a)^{n+1}}}_{\lambda} = 0.$$

This implies that

$$f(b) = P_n(b) + \frac{1}{(n+1)!} f^{(n+1)}(\xi) (b-a)^{n+1},$$

and since we picked  $b$  arbitrarily (kind of), we can take  $b = x$  and we are done since  $\xi \in [a, b]$ .  $\square$

<sup>1</sup>Note that the  $(n+1)$ -th derivative need not be continuous here.

**Remark.** The choice of  $\xi$  in Lagrange's remainder term may (and likely does) vary for different  $x$ .

**Remark.** The Taylor polynomial is unique in the sense that if  $f : [a, b] \rightarrow \mathbb{R}$  and  $f'(a), \dots, f^{(n)}(a)$  exist, then if

$$f(x) = p(x) + o((x - a)^n)$$

as  $x \rightarrow a^+$  for some polynomial  $p(x)$  with  $\deg p \leq n$ , then  $p(x) = P_n(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x - a)^k$ . This is because if  $Q(x) = p(x) - P_n(x)$ , then by Taylor's formula (Peano form), we get

$$\lim_{x \rightarrow a^+} \frac{Q(x)}{(x - a)^n} = \lim_{x \rightarrow a^+} \frac{p(x) - \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x - a)^k}{(x - a)^n} = 0.$$

From here this implies that  $Q(x) = 0$  since  $\deg Q \leq n$ . Another way to see this is to plug in  $x = a$ , which deletes everything except the constant, and then ignore the constant and divide by  $(x - a)$  to repeat.

# Lecture 4

## Jan. 18 — Taylor Polynomials

### 4.1 Common Taylor Polynomials

We have

$$\begin{aligned}e^x &= 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + o(x^n), \\ \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + \frac{(-1)^{n-1}x^{2n-1}}{(2n-1)!} + o(x^{2n}), \\ \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + \frac{(-1)^n x^{2n}}{(2n)!} + o(x^{2n+1}), \\ (1+x)^\alpha &= 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n + o(x^n), \\ \ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\frac{x^n}{n} + o(x^n).\end{aligned}$$

### 4.2 Combining Taylor Polynomials

**Remark.** If  $a = 0$  and  $f(x)$  is even in  $(-b, b)$ , then

$$f(x) = \sum_{k=0}^{n/2} a_k x^{2k} + o(x^n).$$

Similarly if  $f(x)$  is odd in  $(-b, b)$ , then

$$f(x) = \sum_{k=0}^{n/2} a_k x^{2k+1} + o(x^{n+1}).$$

**Remark.** To create new Taylor polynomials from known ones, we can observe that if  $f(x) = P_n(x) + o((x-a)^n)$  and  $g(x) = Q_n(x) + o((x-a)^n)$ , then

$$f(x) + g(x) = (P_n(x) + Q_n(x)) + o((x-a)^n) \quad \text{and} \quad f(x)g(x) = \underbrace{(P_n(x)Q_n(x))}_{\text{take first } n \text{ terms}} + o((x-a)^n).$$

If  $P_n(x) = \sum_{k=0}^n a_k (x-a)^k$  and  $Q_n(x) = \sum_{k=0}^n b_k (x-a)^k$ , then  $f(x)g(x)$  has Taylor polynomial  $\sum_{k=0}^n c_k (x-a)^k$  where

$$c_k = \sum_{i=0}^k a_i b_{k-i}.$$

If  $h(x) = f(x)/g(x)$  and  $g(x) \neq 0$  near  $x = a$ , then  $f(x) = h(x)g(x)$ . Let  $h(x) = \sum_{k=0}^n c_k(x-a)^k + o((x-a)^n)$ , then

$$a_k = \sum_{i=0}^k c_i b_{k-i}$$

for  $0 \leq k \leq n$ , after which we can solve for the  $c_k$ .

**Example 4.0.1.** Find the Taylor polynomial for  $\tan x$  up to  $n = 5$ .

*Proof.* Note that  $\tan x$  is odd, so we can write

$$\tan x = x + a_3 x^3 + a_5 x^5 + o(x^5).$$

Now since  $\tan x = \sin x / \cos x$ , we have  $\sin x = \tan x \cos x$ , so

$$x - \frac{x^3}{6} + \frac{x^5}{5!} + o(x^5) = (x + a_3 x^3 + a_5 x^5) \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!}\right)$$

We can solve to get

$$\begin{cases} -\frac{1}{6} = -\frac{1}{2} + a_3 \\ \frac{1}{5!} = \frac{1}{4!} - \frac{a_3}{2!} + a_5 \end{cases} \implies a_3 = -\frac{1}{3}, \quad a_5 = \frac{2}{15}$$

as the coefficients for the Taylor polynomial. □

**Remark.** If

$$f'(x) = \sum_{k=0}^n b_k (x-a)^k + o((x-a)^n),$$

then the anti-derivative of  $f(x)$  has

$$f(x) = f(x_0) + \sum_{k=0}^n a_{k+1} (x-a)^{k+1} + o((x-a)^{n+1}),$$

where  $a_{k+1} = b_k / (k+1)$  for  $0 \leq k \leq n$ . This is because

$$b_k = \frac{(f')^{(k)}(a)}{k!} = \frac{f^{(k+1)}(a)}{k!} \quad \text{and} \quad a_{k+1} = \frac{f^{(k+1)}(a)}{k+1} = \frac{1}{k+1} \frac{f^{(k+1)}(a)}{k!} = \frac{b_k}{k+1}.$$

**Example 4.0.2.** Find the Taylor polynomial for  $f(x) = \arctan x$ .

*Proof.* Recall that

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

Using the above we get

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

as the Taylor polynomial. □



## 4.3 Applications for Taylor Polynomials

### 4.3.1 Finding Limits

**Remark.** Let  $f(x) = ax^n + o(x^n)$  as  $x \rightarrow 0$  and  $g(x) = bx^n + o(x^n)$  where  $b \neq 0$ . Then

$$\lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = \frac{a}{b}.$$

**Remark.** For the polynomial of  $f(g(x))$ , we can do

$$f(u) = \sum_{k=0}^n a_k(u - g(a))^k + o((u - g(a))^n), \quad \text{where} \quad u = g(x) = \sum_{k=0}^n b_k(x - a)^k + o((x - a)^n).$$

Then we can substitute in  $u = g(x)$  to find the overall polynomial.

**Example 4.0.3.** Find

$$\lim_{x \rightarrow 0} \frac{\sqrt{1 + 2 \tan x} - e^x + x^2}{\arcsin x - \sin x}.$$

*Proof.* Note that

$$\begin{aligned} \sqrt{1 + 2 \tan x} - e^x + x^2 &= \frac{2x^3}{3} + o(x^3), \\ \arcsin x - \sin x &= \frac{x^3}{3} + o(x^3). \end{aligned}$$

So the desired limit is 2. □

**Remark.** If  $f(x) = ax^n + o(x^n)$  and  $g(x) = bx^m + o(x^m)$  for  $a, b \neq 0$ , then

$$\lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = \begin{cases} a/b & \text{if } m = n, \\ 0 & \text{if } m < n, \\ \infty & \text{if } m > n. \end{cases}$$

**Example 4.0.4.** Assume  $f(x) = 1 + ax^n + o(x^n)$  where  $a \neq 0$  and

$$g(x) = \frac{1}{bx^n + o(x^n)}, \quad \text{i.e.} \quad \frac{1}{g(x)} = bx^n + o(x^n).$$

for  $b \neq 0$ . Then

$$\lim_{x \rightarrow 0} f(x)^{g(x)} = e^{a/b}.$$

Let  $y(x) = f(x)^{g(x)}$ , then  $\ln y(x) = g(x) \ln f(x)$ . Note that

$$\ln f(x) = \ln(1 + ax^n + o(x^n)) = ax^n + o(x^n),$$

so that

$$\frac{\ln f(x)}{1/g(x)} = \frac{ax^n + o(x^n)}{bx^n + o(x^n)} \rightarrow \frac{a}{b}$$

as  $x \rightarrow 0$ . Thus  $\ln y(x) \rightarrow a/b$  and  $y(x) \rightarrow e^{a/b}$  as  $x \rightarrow 0$ .

**Example 4.0.5.** Find

$$\lim_{x \rightarrow 0} [\cos(xe^x) - \ln(1-x) - x]^{\cot x^3}.$$

*Proof.* Here we have

$$f(x) = \cos(xe^x) - \ln(1-x) - x = 1 - \frac{2}{3}x^3 + o(x^3) \quad \text{and} \quad \frac{1}{g(x)} = \tan x^3 = x^3 + o(x^3).$$

Thus the limit is  $e^{-2/3}$ . □

### 4.3.2 Estimation

**Example 4.0.6.** Let  $f(x)$  be twice differentiable in  $[0, 1]$  and  $f(0) = f(1)$ . Further assume  $|f''(x)| \leq M$  for  $0 \leq x \leq 1$ . Prove that  $|f'(x)| \leq M/2$  for  $0 \leq x \leq 1$ .

*Proof.* Recall that Lagrange's form of Taylor's theorem says

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(\xi)}{2!}(x-a)^2$$

for some  $\xi$  between  $a$  and  $x$ . Thus for any  $x \in (0, 1)$ , we have

$$f(1) = f(x) + f'(x)(1-x) + \frac{f''(\xi_1)}{2}(1-x)^2.$$

Similarly, we have

$$f(0) = f(x) + f'(x)(-x) + \frac{f''(\xi_2)}{2}x^2.$$

Here  $x \leq \xi_1 \leq 1$  and  $0 \leq \xi_2 \leq x$ . Since  $f(1) = f(0)$ , we can solve for  $f'(x)$  to get

$$f'(x) = \frac{f''(\xi_2)x^2 - f''(\xi_1)(1-x)^2}{2}.$$

Then taking absolute values yields

$$|f'(x)| \leq M \left( \frac{x^2 + (1-x)^2}{2} \right) \leq \frac{M}{2} \max_{0 \leq x \leq 1} [x^2 + (1-x)^2] = \frac{M}{2},$$

as desired. □

**Example 4.0.7.** Let  $f(x)$  be twice differentiable in  $[0, 1]$  and  $f'(a) = f'(b) = 0$ . Then there exists  $\xi \in (a, b)$  such that

$$|f''(\xi)| \geq 4 \frac{|f(a) - f(b)|}{(b-a)^2}.$$

*Proof.* Note that this is equivalent to

$$|f(b) - f(a)| \leq f''(\xi) \left( \frac{b-a}{2} \right)^2.$$

Then we have

$$f\left(\frac{b+a}{2}\right) = f(a) + \frac{f''(\xi_1)}{2} \left(\frac{b-a}{2}\right)^2 = f(b) - \frac{f''(\xi_2)}{2} \left(\frac{b-a}{2}\right)^2,$$

so that

$$f(b) - f(a) = \frac{f''(\xi_2) + f''(\xi_1)}{2} \left( \frac{b-a}{2} \right)^2.$$

From here we have

$$|f(b) - f(a)| \leq \underbrace{\frac{|f''(\xi_1)| + |f''(\xi_2)|}{2}}_{=|f''(\xi)|} \left( \frac{b-a}{2} \right)^2$$

for some  $\xi \in (a, b)$  by Darboux's lemma, as desired.

□

# Lecture 5

## Jan. 23 — The Riemann Integral

### 5.1 The Anti-Derivative

Recall the *anti-derivative* from calculus:

**Definition 5.1.** Let  $f : U \rightarrow \mathbb{R}$  where  $U$  is an interval in  $\mathbb{R}$ . If there exists a differentiable function  $F : U \rightarrow \mathbb{R}$  such that  $F'(x) = f(x)$  for all  $x \in U$ , then  $F(x)$  is an *anti-derivative* of  $f$ , denoted

$$F(x) = \int f(x) dx.$$

This is also called the *indefinite integral* of  $f$ .

**Remark.** The anti-derivatives of a function can differ by a constant.

**Example 5.1.1.** Find an anti-derivative of  $f(x) = |x|$  for  $x \in \mathbb{R}$ .

*Proof.* If  $x > 0$ , we have  $f(x) = x$  and so  $F(x) = x^2/2$ . If  $x < 0$ , then  $f(x) = -x$  and so  $F(x) = -x^2/2$ . We can also write this as

$$F(x) = x \cdot \frac{|x|}{2}.$$

Clearly for  $x \neq 0$ , we have  $F'(x) = f(x)$ . At  $x = 0$ , we have

$$\lim_{x \rightarrow 0} \frac{F(x) - f(0)}{x} = \lim_{x \rightarrow 0} \frac{1}{2}|x| = 0,$$

so  $F'(0) = f(0)$  and  $F$  is an anti-derivative of  $f$ . □

**Remark.** The eventual goal is to show that any continuous function  $f : [a, b] \rightarrow \mathbb{R}$  has an anti-derivative.

**Example 5.1.2.** Find an anti-derivative for

$$f(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \\ -1 & \text{if } x < 0. \end{cases}$$

*Proof.* We can try to use  $F(x) = |x|$ , but recall that  $F$  is not differentiable at  $x = 0$ . More generally, suppose that  $f(x)$  has some anti-derivative  $F(x)$ , i.e.  $f(x) = F'(x)$ . By Darboux's theorem,  $f(x)$  must take all values in  $(-1, 1)$ , which is a contradiction with the definition of  $f$ . □

**Remark.** If  $f(x)$  has a jump discontinuity, then it has no anti-derivative.

## 5.2 The Riemann Integral

Recall from calculus that if  $f(x)$  is defined in  $[a, b]$  and  $F'(x) = f(x)$ , then we have<sup>1</sup>

$$\int_a^b f(x) dx = F(x) \Big|_a^b = F(b) - F(a).$$

We called this the *definite integral* of  $f$  in calculus, but we would like a more rigorous definition.

**Definition 5.2.** Let  $a, b \in \mathbb{R}$  and  $a < b$ . A *partition* of the interval  $[a, b]$  is a finite sequence of numbers  $x_0, x_1, \dots, x_n$  such that  $a = x_0 < x_1 < \dots < x_n = b$ .

**Definition 5.3.** The *width* of a partition  $x_0, x_1, \dots, x_n$  is  $\max\{x_i - x_{i-1} : i = 1, 2, \dots, n\}$ .

**Definition 5.4.** For any partition  $x_0, x_1, \dots, x_n$ , define the *Riemann sum* to be

$$S = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}),$$

where  $x'_i$  is any point between  $x_{i-1}$  and  $x_i$ , inclusive.<sup>2</sup>

**Definition 5.5.** Let  $a, b \in \mathbb{R}$  with  $a < b$  and  $f : [a, b] \rightarrow \mathbb{R}$ . We say  $f$  is *Riemann integrable* on  $[a, b]$  if there exists  $A \in \mathbb{R}$  such that for all  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $|S - A| < \epsilon$  whenever  $S$  is any Riemann sum for a partition of  $[a, b]$  with width less than  $\delta$ . We call  $A$  the *Riemann integral* of  $f$  on  $[a, b]$  and denote it by

$$A = \int_a^b f(x) dx.$$

**Remark.** If  $f$  is Riemann integrable, then

$$A = \int_a^b f(x) dx$$

is unique. This is because if  $A$  and  $A'$  are two numbers for the Riemann integral, then for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that

$$|A - S| < \epsilon \quad \text{and} \quad |A' - S| < \epsilon$$

for any Riemann sum  $S$  associated with a partition of width less than  $\delta$ . Then

$$|A - A'| \leq |A - S| + |A' - S| < 2\epsilon,$$

so  $A = A'$  and thus the Riemann integral is unique.

**Example 5.5.1.** Let  $f(x) = c$  on  $[a, b]$ , a constant function. Then for any partition  $x_0, x_1, \dots, x_n$ ,

$$S = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}) = \sum_{i=1}^n c(x_i - x_{i-1}) = c(b - a) \implies \int_a^b c dx = c(b - a).$$

<sup>1</sup>This is the *fundamental theorem of calculus*.

<sup>2</sup>The geometric intuition of the Riemann sum is an approximation for the *area* under the graph of  $f$  by rectangles.

**Example 5.5.2.** Fix  $\xi \in [a, b]$  and let  $f : [a, b] \rightarrow \mathbb{R}$  be defined by

$$f(x) = \begin{cases} 0 & \text{if } x \neq \xi \\ c & \text{if } x = \xi. \end{cases}$$

Check that

$$A = \int_a^b f(x) dx = 0.$$

*Proof.* For any partition  $a = x_0 < x_1 < \cdots < x_n = b$  with width  $\delta$ , we have

$$|S| = \left| \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}) \right| \leq |c|2\delta$$

since  $\xi$  can be in at most two of the intervals of the partition. Then for any  $\epsilon > 0$ , choose  $\delta = \epsilon/(2|c|)$ , so that  $|S| < \epsilon$  for any partition of width less than  $\delta$ . From this we can conclude that  $A = 0$ .  $\square$

**Example 5.5.3.** Consider a step function. Let  $\alpha, \beta \in [a, b]$  with  $\alpha < \beta$ . Define  $f : [a, b] \rightarrow \mathbb{R}$  by

$$f(x) = \begin{cases} 1 & \text{if } x \in (\alpha, \beta) \\ 0 & \text{if } x \notin (\alpha, \beta) \text{ and } x \in [a, b]. \end{cases}$$

Note that  $f$  has no anti-derivative, but it is Riemann integrable. In fact,

$$\int_a^b f(x) dx = \beta - \alpha.$$

To see this, take any partition  $a = x_0 < x_1 < \cdots < x_n = b$  with width less than  $\delta$ . Then

$$S = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}) = \sum_{[x_{i-1}, x_i] \cap [\alpha, \beta] \neq \emptyset} f(x'_i)(x_i - x_{i-1}).$$

Each partition is in two classes: Either (1) it only partially intersects  $[\alpha, \beta]$  or (2) it is contained in  $[\alpha, \beta]$ . So

$$S = \underbrace{1(\text{total length of intervals of class 2})}_{I_1} + \underbrace{|f(x'_i)|(\text{total length of intervals of class 1})}_{I_2}.$$

We have  $|I_1 - (\beta - \alpha)| < 2\delta$  and  $|I_2| < 2\delta$  since there are at most two intervals of class 1. So

$$|S - (\beta - \alpha)| \leq |I_1| + |I_2| < 4\delta.$$

So  $f(x)$  is Riemann integrable and

$$\int_a^b f(x) dx = \beta - \alpha,$$

as desired.

**Example 5.5.4.** Define  $f : [a, b] \rightarrow \mathbb{R}$  by

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Then  $f(x)$  is not Riemann integrable. For any partition  $a = x_0 < x_1 < \cdots < x_n = b$ ,

$$S = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}) = \begin{cases} b - a & \text{if } x'_i \text{ are all rational} \\ 0 & \text{if } x'_i \text{ are all irrational.} \end{cases}$$

We can always choose  $x'_i$  to be in either case since the rationals and irrationals are both dense in  $\mathbb{R}$ . So there is no  $A \in \mathbb{R}$  such that  $|A - S| < \epsilon$ , no matter how small we take  $\delta$  to be.

**Remark.** The function  $f$  from the previous example is not Riemann integrable, but it is Lebesgue integrable. In fact,

$$L = \int_a^b f(x) dx = 0$$

with respect to the Lebesgue measure. This is because the set of rational numbers  $\mathbb{Q}$  has measure zero.

## 5.3 Properties of the Riemann Integral

**Proposition 5.1.** *We have the following linearity properties of the Riemann integral:*

1. *If  $f, g : [a, b] \rightarrow \mathbb{R}$  are Riemann integrable, then  $f \pm g$  are also integrable and*

$$\int_a^b (f \pm g) dx = \int_a^b f(x) dx \pm \int_a^b g(x) dx$$

2. *For any  $c \in \mathbb{R}$ ,  $cf$  is integrable and*

$$\int_a^b cf dx = c \int_a^b f(x) dx.$$

*Proof.* See textbook, fairly straightforward. □

**Remark.** Since we only discuss Riemann integration in this class, we will sometimes simply say “integrable” instead of “Riemann integrable.”

**Proposition 5.2.** *If  $f : [a, b] \rightarrow \mathbb{R}$  is integrable and  $f(x) \geq 0$ , then*

$$\int_a^b f(x) dx \geq 0.$$

*Proof.* Let

$$A = \int_a^b f(x) dx.$$

Then for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that for any partition of width  $< \delta$ , we have  $|A - S| < \epsilon$ . But

$$S = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}) \geq 0,$$

Then we have  $A > S - \epsilon \geq -\epsilon$ , so taking  $\epsilon \rightarrow 0$  gives  $A \geq 0$ . □

**Corollary 5.0.1.** *If  $f, g : [a, b] \rightarrow \mathbb{R}$  are integrable and  $f(x) \geq g(x)$  for all  $x \in [a, b]$ , then*

$$\int_a^b f(x) \, dx \geq \int_a^b g(x) \, dx.$$

*Proof.* By linearity,

$$\int_a^b f(x) \, dx - \int_a^b g(x) \, dx = \int_a^b (f(x) - g(x)) \, dx \geq 0$$

since  $f(x) - g(x) \geq 0$  by assumption. □

**Corollary 5.0.2.** *If  $f : [a, b] \rightarrow \mathbb{R}$  is integrable and  $m \leq f(x) \leq M$  for all  $x \in [a, b]$ , then*

$$m(b - a) \leq \int_a^b f(x) \, dx \leq M(b - a).$$



# Lecture 6

## Jan. 25 — Riemann Integrability

### 6.1 Conditions for Integrability

**Lemma 6.1.** *A function  $f : [a, b] \rightarrow \mathbb{R}$  is integrable if and only if for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $|S_1 - S_2| < \epsilon$  whenever  $S_1$  and  $S_2$  are Riemann sums for partitions of width less than  $\delta$ .*

*Proof.* ( $\Rightarrow$ ) If  $f$  is integrable, then for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that

$$\left| S - \int_a^b f(x) dx \right| < \frac{\epsilon}{2}$$

for any Riemann sum  $S$  of a partition with width less than  $\delta$ . Then

$$|S_1 - S_2| \leq \left| S_1 - \int_a^b f(x) dx \right| + \left| S_2 - \int_a^b f(x) dx \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

as desired.

( $\Leftarrow$ ) Take the special partition into intervals of equal length, with width  $(b-a)/n$ . Pick the middle point in each interval, and let

$$S_n = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1})$$

be the corresponding Riemann sum. Now we check that  $\{S_n\}_{n=1}^\infty$  is a Cauchy sequence. This is because for any  $\epsilon > 0$ , if  $N$  is large enough, then for any  $n, m \geq N$ , we have  $|S_n - S_m| < \epsilon$  if  $1/N < \delta$ . Then  $\{S_n\}_{n=1}^\infty$  converges, so let  $\lim_{n \rightarrow \infty} S_n = A$ . Now for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that for any Riemann sum  $S$  with width  $< \delta$ , if  $1/n < \delta$ , then  $|S_n - S| < \epsilon/2$ . So

$$|S - A| \leq |S_n - S| + |S_n - A| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

if  $n$  is large enough. Thus

$$A = \int_a^b f(x) dx$$

exists and is the Riemann integral of  $f$ . □

**Remark.** Recall the step function  $f : [a, b] \rightarrow \mathbb{R}$  given by

$$f(x) = \begin{cases} 1 & \text{if } x \in (\alpha, \beta) \subseteq [a, b] \\ 0 & \text{if } x \notin (\alpha, \beta). \end{cases}$$

Last time we saw that  $f$  is integrable and that

$$\int_a^b f(x) dx = \beta - \alpha.$$

Now let us consider a more general step function. We call  $f$  a *step function* on  $[a, b]$  if there exists a partition  $x_0 < x_1 < \cdots < x_n$  of  $[a, b]$  such that  $f(x)$  is constant on each subinterval  $(x_{i-1}, x_i)$ .

**Lemma 6.2.** *If  $f : [a, b] \rightarrow \mathbb{R}$  is a step function for a partition  $x_0 < x_1 < \cdots < x_n$  and  $f(x) = c_i$  when  $x \in (x_{i-1}, x_i)$ , then  $f$  is integrable and*

$$\int_a^b f(x) dx = \sum_{i=1}^n c_i(x_i - x_{i-1}).$$

*Proof.* Define

$$\varphi_i(x) = \begin{cases} 1 & \text{if } x \in (x_{i-1}, x_i) \\ 0 & \text{otherwise.} \end{cases}$$

Now let

$$h = f - \sum_{i=1}^n c_i \varphi_i.$$

Then  $h(x)$  is nonzero only at  $\{x_i\}_{i=0}^n$ . Each  $\varphi_i$  is integrable and  $h$  is integrable with

$$\int_a^b h(x) dx = 0,$$

so  $f$  is also integrable and

$$\int_a^b f(x) dx = \sum_{i=1}^n c_i \int_a^b \varphi_i(x) dx = \sum_{i=1}^n c_i(x_i - x_{i-1})$$

by linearity and the integral of a simple step function that we calculated before. □

**Proposition 6.1.** *A function  $f : [a, b] \rightarrow \mathbb{R}$  is integrable if and only if for any  $\epsilon > 0$ , there exist step functions  $f_1, f_2$  such that  $f_1(x) \leq f(x) \leq f_2(x)$  for all  $x \in [a, b]$  and*

$$\int_a^b (f_2 - f_1) dx < \epsilon.$$

*Proof.* ( $\Leftarrow$ ) For any  $\epsilon > 0$ , choose step functions  $f_1, f_2$  such that

$$\int_a^b (f_2 - f_1) dx < \frac{\epsilon}{3}.$$

Then there exists  $\delta > 0$  such that for any partition with width  $< \delta$ , the Riemann sums  $S_1, S_2$  for  $f_1, f_2$  satisfy

$$|S_1 - \int_a^b f_1(x) dx| < \frac{\epsilon}{3} \quad \text{and} \quad |S_2 - \int_a^b f_2(x) dx| < \frac{\epsilon}{3}.$$

So for any partition width  $< \delta$ , the Riemann sum of  $f$  is

$$S = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}),$$

and  $S_1 \leq S \leq S_2$  since

$$S_1 = \sum_{i=1}^n f_1(x'_i)(x_i - x_{i-1}) \quad \text{and} \quad S_2 = \sum_{i=1}^n f_2(x'_i)(x_i - x_{i-1}).$$

So  $S$  is in the interval  $(S_1, S_2)$ , which has length  $< \epsilon$  by the triangle inequality on the previous results. For any two Riemann sums of  $f$  with partitions of width  $< \delta$ , we have  $|S' - S''| < \epsilon$ . Thus  $f$  is integrable.

( $\Rightarrow$ ) First we show that  $f$  is bounded in  $[a, b]$ . This is because for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that any two Riemann sums  $S_1, S_2$  corresponding to partitions of width  $< \delta$  satisfy  $|S_1 - S_2| < \epsilon$ . Let

$$S_1 = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}),$$

and replace  $x'_{i_0} \in (x_{i_0-1}, x_{i_0})$  with  $x''_{i_0} \in (x_{i_0-1}, x_{i_0})$ . Keep  $x'_i$  for  $i \neq i_0$ . Define this new Riemann sum to be  $S_2$ . Then

$$|S_2 - S_1| \leq |f(x''_{i_0}) - f(x'_{i_0})|(x_{i_0} - x_{i_0-1}) < \epsilon,$$

so that

$$|f(x''_{i_0})| \leq |f(x'_{i_0})| + \frac{\epsilon}{x_{i_0} - x_{i_0-1}},$$

i.e.  $f$  is bounded in  $(x_{i_0-1}, x_{i_0})$  since  $x''_{i_0}$  was arbitrary. Since we also picked  $i_0$  arbitrarily, we can repeat this for any interval to conclude that  $f$  is bounded in  $[a, b]$ .

Now for any partition  $x_0 < x_1 < \dots < x_n$  with width  $< \delta$ , define

$$m_i = \inf\{f(x) : x \in (x_{i-1}, x_i)\} \quad \text{and} \quad M_i = \sup\{f(x) : x \in (x_{i-1}, x_i)\}.$$

Define the step function

$$f_1(x) = \begin{cases} m_i & \text{if } x \in (x_{i-1}, x_i) \\ \min\{m_1, \dots, m_n\} & \text{if } x = x_i \text{ for } i = 0, \dots, n. \end{cases}$$

Similarly define

$$f_2(x) = \begin{cases} M_i & \text{if } x \in (x_{i-1}, x_i) \\ \max\{M_1, \dots, M_n\} & \text{if } x = x_i \text{ for } i = 0, \dots, n. \end{cases}$$

Observe that  $f_1(x) \leq f(x) \leq f_2(x)$  for any  $x \in [a, b]$  by construction. Now we verify that

$$\int_a^b (f_2 - f_1) dx < \epsilon$$

if  $\delta > 0$  is small enough. This is because for any  $\eta > 0$ , there exists  $x'_i, x''_i \in [x_{i-1}, x_i]$  such that  $f(x'_i) < m_i + \eta$  and  $f(x''_i) > M_i - \eta$ . Then

$$\sum_{i=1}^n (f(x''_i) - f(x'_i))(x_i - x_{i-1}) > \sum_{i=1}^n (M_i - m_i - 2\eta)(x_i - x_{i-1}) = \int_a^b (f_2 - f_1) dx - 2\eta(b - a).$$

If  $\delta > 0$  is small enough, then

$$\sum_{i=1}^n (f(x'_i) - f(x''_i))(x_i - x_{i-1}) < \epsilon$$

since this is a difference of two Riemann sums with partitions of width  $< \delta$ . Thus

$$\int_a^b (f_2 - f_1) dx < \epsilon + 2\eta(b - a).$$

But  $\eta$  was arbitrary, so taking  $\eta \rightarrow 0$  gives the desired result.  $\square$

**Corollary 6.0.1.** *If  $f : [a, b] \rightarrow \mathbb{R}$  is integrable, then it is bounded.*

*Proof.* This was shown in the proof of the previous proposition.  $\square$

**Theorem 6.1.** *If  $f : [a, b] \rightarrow \mathbb{R}$  is continuous, then  $f$  is integrable.*

*Proof.* Since  $f$  is continuous on the compact set  $[a, b]$ , it is uniformly continuous. So for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that for any  $x', x'' \in [a, b]$ , we have  $|f(x') - f(x'')| < \epsilon$  whenever  $|x' - x''| < \delta$ . Now let  $S_1, S_2$  be two Riemann sums with partitions of width  $< \delta$ . Assume without loss of generality that  $S_1, S_2$  are defined over the same partition (we can always combine two partitions to give a finer partition, if necessary). Let

$$S_1 = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}) \quad \text{and} \quad S_2 = \sum_{i=1}^n f(x''_i)(x_i - x_{i-1}).$$

Then

$$|S_1 - S_2| \leq \sum_{i=1}^n |f(x'_i) - f(x''_i)|(x_i - x_{i-1}) < \epsilon \sum_{i=1}^n (x_i - x_{i-1}) = \epsilon(b - a).$$

Since  $\epsilon > 0$  was arbitrary, we conclude that  $f$  is integrable by Lemma 6.1.  $\square$

## 6.2 The Fundamental Theorem of Calculus

**Theorem 6.2** (Fundamental theorem of calculus). *If  $f : [a, b] \rightarrow \mathbb{R}$  has anti-derivative  $F : [a, b] \rightarrow \mathbb{R}$  and  $f \in \mathcal{R}([a, b])$ ,<sup>1</sup> then*

$$\int_a^b f(x) dx = F(b) - F(a).$$

*Proof.* Since  $f$  is integrable, let

$$A = \int_a^b f(x) dx.$$

For any  $\epsilon > 0$ , there exists  $\delta > 0$  such that for any Riemann sum  $S$  with partition of width  $< \delta$ , we have  $|S - A| < \epsilon$ . Let  $x_0 < x_1 < \dots < x_n$  be a partition of width  $< \delta$ . Then by telescoping,

$$F(b) - F(a) = \sum_{i=1}^n (F(x_i) - F(x_{i-1})) = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1})$$

---

<sup>1</sup>Here  $\mathcal{R}([a, b])$  is the class of Riemann integrable functions on  $[a, b]$ .

by Lagrange's mean value theorem, where  $x'_i \in (x_{i-1}, x_i)$ . Then

$$|F(b) - F(a) - A| = |S - A| < \epsilon,$$

so letting  $\epsilon \rightarrow 0$  gives  $F(b) - F(a) = A$ . □

**Remark.** The fundamental theorem of calculus requires both being Riemann integrable and having an anti-derivative, which do not always overlap. In fact, neither is a subset of the other.

**Example 6.0.1.** The step function

$$f(x) = \begin{cases} -1 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } 1 < x \leq 2 \end{cases}$$

is integrable but has no anti-derivative.

**Example 6.0.2.** Define

$$F(x) = \begin{cases} 0 & \text{if } x = 0 \\ x^2 \sin(1/x) & \text{if } x \neq 0. \end{cases}$$

Then we have

$$F'(x) = f(x) = \begin{cases} 0 & \text{if } x = 0 \\ (-2/x)(\cos(1/x^2)) + 2x \sin(1/x^2) & \text{if } x \neq 0. \end{cases}$$

We can check that  $F'(0) = 0$  via the definition of the derivative. Note that  $f$  has an anti-derivative, namely  $F$ . However,  $f$  is not integrable since it is not bounded near  $x = 0$ .

# Lecture 7

## Jan. 30 — Riemann Integrability, Part 2

### 7.1 Conditions for an Anti-Derivative

**Lemma 7.1.** *Let  $c \in (a, b)$ . Then  $f \in \mathcal{R}([a, b])$  if and only if  $f \in \mathcal{R}([a, c])$  and  $f \in \mathcal{R}([c, b])$ . Moreover,*

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx. \quad (*)$$

*Proof.* ( $\Rightarrow$ ) If  $f \in \mathcal{R}([a, b])$ , then for any  $\epsilon > 0$ , there exist two step functions  $f_1, f_2$  such that  $f_1 \leq f \leq f_2$  and

$$\int_a^b (f_2 - f_1) dx < \epsilon.$$

Let  $f_1, f_2$  be the restrictions to  $[a, c]$ . Then still  $f_1 \leq f \leq f_2$  on  $[a, c]$  and

$$\int_a^c (f_2 - f_1) \leq \int_a^b (f_2 - f_1) < \epsilon$$

since  $f_2 - f_1$  is a nonnegative step function. (Note that the desired result is easy to verify for step functions.) So  $f \in \mathcal{R}([a, c])$ , and the same argument works to show that  $f \in \mathcal{R}([c, b])$ .

( $\Leftarrow$ ) If  $f \in \mathcal{R}([a, c])$  and  $f \in \mathcal{R}([c, b])$ , then for any  $\epsilon > 0$ , there exist step functions  $g_1, g_2, h_1, h_2$  such that  $g_1 \leq f \leq g_2$  on  $[a, c]$ ,  $h_1 \leq f \leq h_2$  on  $[c, b]$ , and

$$\int_a^c (g_2 - g_1) dx < \epsilon, \quad \int_c^b (h_2 - h_1) dx < \epsilon.$$

Now define

$$f_i = \begin{cases} g_i & \text{if } x \in [a, c) \\ h_i & \text{if } x \in [c, b] \end{cases}$$

for  $i = 1, 2$ . Then  $f_1 \leq f \leq f_2$  on  $[a, b]$ , and

$$\int_a^b (f_2 - f_1) dx = \int_a^c (g_2 - g_1) dx + \int_c^b (h_2 - h_1) dx < 2\epsilon,$$

so  $f \in \mathcal{R}([a, b])$ . Now to prove (\*), note that  $f \in \mathcal{R}([a, c])$ , so for any  $\epsilon > 0$  there exist Riemann sums  $S_1$  on  $[a, c]$  and  $S_2$  on  $[c, b]$  such that

$$|S_1 - \int_a^c f(x) dx| < \frac{\epsilon}{3}, \quad |S_2 - \int_c^b f(x) dx| < \frac{\epsilon}{3}.$$

Now choose  $\delta > 0$  such that if the Riemann sum  $S$  has partition with width  $< \delta$ , then

$$|S - \int_a^c f(x) dx| < \frac{\epsilon}{3}, \quad |S - \int_c^b f(x) dx| < \frac{\epsilon}{3}, \quad |S - \int_a^b f(x) dx| < \frac{\epsilon}{3}.$$

Now combine  $S_1, S_2$  on  $[a, b]$  to be a Riemann sum  $S = S_1 + S_2$ , so that

$$|S - \int_a^b f(x) dx| < \frac{\epsilon}{3}.$$

By the triangle inequality on the previous results,

$$\left| \int_a^b f(x) dx - \left( \int_a^c f(x) dx + \int_c^b f(x) dx \right) \right| < \epsilon.$$

Since  $\epsilon$  is arbitrarily small, we conclude that

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$$

as desired. □

**Remark.** The formula  $(*)$  is true for any three numbers  $a, b, c$ , as long as  $f$  is integrable. This is because by convention, if  $a > b$ , then

$$\int_a^b f(x) dx = - \int_b^a f(x) dx.$$

**Theorem 7.1.** If  $f : [a, b] \rightarrow \mathbb{R}$  is continuous, then<sup>1</sup>

$$F(x) = \int_a^x f(\xi) d\xi$$

is an anti-derivative of  $f$ .

*Proof.* For any  $x_0 \in (a, b)$ , we check that  $F'(x_0) = f(x_0)$ . We can compute using Lemma 7.1 that

$$\begin{aligned} \left| \frac{F(x_0 + h) - F(x_0)}{h} - f(x_0) \right| &= \left| \frac{1}{h} \left( \int_a^{x_0+h} f(x) dx - \int_a^{x_0} f(x) dx \right) - f(x_0) \right| \\ &= \left| \frac{1}{h} \int_{x_0}^{x_0+h} f(x) dx - f(x_0) \right| = \left| \frac{1}{h} \int_{x_0}^{x_0+h} (f(x) - f(x_0)) dx \right|. \end{aligned}$$

The last step is from observing

$$f(x_0) = \frac{1}{h} \int_{x_0}^{x_0+h} f(x_0) dx.$$

Since  $f$  is continuous, for any  $\epsilon > 0$ , there exists  $\delta$  such that if  $|x_0 - x| < \delta$ , then  $|f(x) - f(x_0)| < \epsilon$ . This gives

$$\left| \frac{1}{h} \int_{x_0}^{x_0+h} (f(x) - f(x_0)) dx \right| \leq \frac{1}{h} \int_{x_0}^{x_0+h} |f(x) - f(x_0)| dx \leq \frac{\epsilon h}{h} = \epsilon$$

if  $|h| < \delta$ . Thus,

$$\lim_{h \rightarrow 0} \frac{F(x_0 + h) - F(x_0)}{h} - f(x_0) = f(x_0),$$

so we indeed have  $F'(x_0) = f(x_0)$ . □

---

<sup>1</sup>Note that this integral is well-defined since any continuous function is integrable, and a continuous function restricted to a subset of its domain, i.e.  $[a, x] \subseteq [a, b]$ , remains continuous.

## 7.2 More Conditions for Integrability

**Definition 7.1.** Let  $f : [a, b] \rightarrow \mathbb{R}$  be bounded and  $x_0 < x_1 < \cdots < x_n$  be a partition of  $[a, b]$ . Define

$$\omega_i = \sup\{|f(x) - f(y)| : x, y \in [x_{i-1}, x_i]\}$$

for  $i = 1, 2, \dots, n$ . We will call this the *oscillation amplitude* of  $f$ .

**Theorem 7.2.** A function  $f : [a, b] \rightarrow \mathbb{R}$  is integrable if and only if for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that for any partition with width  $< \delta$ , we have

$$\sum_{i=1}^n \omega_i \Delta x_i < \epsilon,$$

where  $\Delta x_i = x_i - x_{i-1}$ .

*Proof.* ( $\Leftarrow$ ) For any  $\epsilon > 0$ , choose any two Riemann sums  $S_1, S_2$  over partitions with width  $< \delta$ . Assume without loss of generality that  $S_1$  and  $S_2$  are defined over the same (maybe refined) partition. Let

$$S_1 = \sum_{i=1}^n f(x'_i)(x_i - x_{i-1}), \quad S_2 = \sum_{i=1}^n f(x''_i)(x_i - x_{i-1}).$$

Then we have

$$|S_1 - S_2| \leq \sum_{i=1}^n |f(x'_i) - f(x''_i)| \Delta x_i \leq \sum_{i=1}^n \omega_i \Delta x_i < \epsilon.$$

Then by Lemma 6.1, we conclude that  $f$  is integrable.

( $\Rightarrow$ ) Since  $f$  is integrable, by Lemma 6.1 we have that for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that for any two Riemann sums  $S_1, S_2$  over partitions of width  $< \delta$ , we have  $|S_1 - S_2| < \epsilon$ . In the interval  $[x_{i-1}, x_i]$ , let

$$M_i = \sup_{x \in [x_{i-1}, x_i]} |f(x)|, \quad m_i = \inf_{x \in [x_{i-1}, x_i]} |f(x)|.$$

In particular note that  $\omega_i = M_i - m_i$ . Now for any  $\eta > 0$ , there exist  $x'_i, x''_i \in [x_{i-1}, x_i]$  such that

$$f(x'_i) > M_i - \eta, \quad f(x''_i) < m_i + \eta.$$

Let

$$S_1 = \sum_{i=1}^n f(x'_i) \Delta x_i, \quad S_2 = \sum_{i=1}^n f(x''_i) \Delta x_i.$$

Then we have

$$|S_1 - S_2| \leq \left| \sum_{i=1}^n (f(x'_i) - f(x''_i)) \Delta x_i \right|$$

Note that  $f(x'_i) - f(x''_i) \geq M_i - m_i - 2\eta$  for  $\eta$  sufficiently small. Thus

$$|S_1 - S_2| \geq \sum_{i=1}^n \omega_i \Delta x_i - 2\eta \sum_{i=1}^n \Delta x_i,$$

so that

$$\sum_{i=1}^n \omega_i \Delta x_i \leq |S_1 - S_2| + 2\eta(b - a) < \epsilon + 2\eta(b - a).$$

From here letting  $\eta \rightarrow 0$  gives the desired result. □



**Theorem 7.3.** *A function  $f : [a, b] \rightarrow \mathbb{R}$  is integrable if and only if for any  $\epsilon > 0$ , there exists a partition such that*

$$\sum_{i=1}^n \omega_i(f) \Delta x_i < \epsilon.$$

*Proof.* ( $\Rightarrow$ ) This is immediate from the previous theorem.

( $\Leftarrow$ ) We show that for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that for any partition with width  $< \delta$ , we have

$$\sum_{i=1}^n \omega_i(f) \Delta x_i < \epsilon.$$

This will imply that  $f$  is integrable by the previous theorem. Let  $y_0 < y_1 < \cdots < y_m$  be the partition satisfying

$$\sum_{j=1}^m \omega_j(f) \Delta y_j < \epsilon,$$

and choose

$$\delta < \frac{1}{4} \min_{j=1, \dots, m} \Delta y_j.$$

For any partition  $x_0 < \cdots < x_n$  with width  $< \delta$ . We divide the intervals  $[x_{i-1}, x_i]$  into two classes. The first case (1) is where  $[x_{i-1}, x_i]$  is contained in one of the  $[y_{j-1}, y_j]$ , and the second case (2) is where  $[x_{i-1}, x_i]$  contains an interior point  $y_j$ . In the first case, we have

$$\sum_{(1)} \omega_i(f) \Delta x_i \leq \sum_{j=1}^m \omega_j(f) \Delta y_j.$$

For the second case, since  $[x_{i-1}, x_i]$  contains an interior point  $y_j$  but  $\delta < \Delta y_j, \Delta y_{j+1}$ , we must have

$$y_{j-1} < x_{i-1} < y_j < x_i < y_{j+1},$$

so that

$$\omega_i(f) \Delta x_i \leq \frac{1}{2} (\omega_j(f) \Delta y_j + \omega_{j+1}(f) \Delta y_{j+1}).$$

This implies

$$\sum_{(2)} \omega_i(f) \Delta x_i \leq \frac{1}{2} \sum_{j=1}^m \omega_j(f) \Delta y_j.$$

Thus

$$\sum_{i=1}^n \omega_i(f) \Delta x_i = \sum_{(1)} \omega_i(f) \Delta x_i + \sum_{(2)} \omega_i(f) \Delta x_i \leq 2 \sum_{j=1}^m \omega_j(f) \Delta y_j < 2\epsilon,$$

so that  $f$  is integrable. □

# Lecture 8

## Feb. 1 — Riemann Integrability, Part 3

### 8.1 Even More Conditions for Integrability

**Example 8.0.1.** If  $f(x)$  is monotone on  $[a, b]$ , then  $f \in \mathcal{R}([a, b])$ .

*Proof.* Suppose  $f(x)$  is monotone increasing on  $[a, b]$  and  $f(x)$  is not constant (since the result is trivial if  $f$  is constant). Then  $f(a) \leq f(x) \leq f(b)$ . For any  $\epsilon > 0$ , for any partition  $x_0 < \cdots < x_n$  with width

$$\delta < \frac{\epsilon}{f(b) - f(a)},$$

we have on  $[x_{i-1}, x_i]$  that  $M_i = f(x_i)$  and  $f(x_{i-1}) = m_i$  since  $f$  is monotone. Then

$$\omega_i(f) = f(x_i) - f(x_{i-1}) = M_i - m_i.$$

Thus

$$\sum_{i=1}^n \omega_i(f) \Delta x_i = \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \Delta x_i < \frac{\epsilon}{f(b) - f(a)} \sum_{i=1}^n (f(x_i) - f(x_{i-1})) = \epsilon$$

since the sum telescopes and comes out to  $f(b) - f(a)$ . Thus  $f$  is integrable.  $\square$

**Theorem 8.1** (Du Bois-Reymond). *Let  $f$  be bounded on  $[a, b]$ . Then  $f \in \mathcal{R}([a, b])$  if and only if for any  $\epsilon, a > 0$ , there exists a partition such that the total length of subintervals with  $\omega_i(f) \geq \epsilon$  is  $< a$ .*

*Proof.* For any partition  $x_0 < \cdots < x_n$ , split

$$\sum_{i=1}^n \omega_i(f) \Delta x_i = \sum_{(A)} \omega_i(f) \Delta x_i + \sum_{(B)} \omega_i(f) \Delta x_i$$

where (A) is over subintervals with width  $\omega_i(f) < \epsilon$  and (B) is over subintervals with width  $\omega_i(f) \geq \epsilon$ .

( $\Rightarrow$ ) Let

$$\Omega = \sup_{x, y \in [a, b]} |f(x) - f(y)|.$$

For any  $\epsilon > 0$ , for

$$\epsilon_1 = \frac{\epsilon}{2(b-a)} \quad \text{and} \quad a = \frac{\epsilon}{2\Omega},$$

by assumption there exists a partition  $x_0 < \cdots < x_n$  such that

$$\begin{aligned} \sum_{i=1}^n \omega_i(f) \Delta x_i &= \sum_{(A)} \omega_i(f) \Delta x_i + \sum_{(B)} \omega_i(f) \Delta x_i \\ &< \frac{\epsilon}{2(b-a)} \sum_{(a)} \Delta x_i + \Omega \sum_{(B)} \Delta x_i < \frac{\epsilon}{2(b-a)}(b-a) + \Omega \frac{\epsilon}{2\Omega} = \epsilon. \end{aligned}$$

So we see that  $f \in \mathcal{R}([a, b])$  as desired.

( $\Rightarrow$ ) If  $f \in \mathcal{R}([a, b])$ , then for any  $\epsilon, a > 0$ , there exists a partition  $x_0 < \cdots < x_n$  such that

$$\sum_{i=1}^n \omega_i(f) \Delta x_i < a\epsilon.$$

Then we have

$$\epsilon \sum_{(B)} \Delta x_i \leq \sum_{(B)} \omega_i(f) \Delta x_i < a\epsilon \implies \sum_{(B)} \Delta x_i < a,$$

which shows the desired result.  $\square$

**Corollary 8.1.1.** *If  $f : [a, b] \rightarrow \mathbb{R}$  is bounded and has only finitely many discontinuity points, then  $f \in \mathcal{R}([a, b])$ .*

*Proof.* Suppose  $f(x)$  has  $p$  discontinuity points on  $[a, b]$  and  $m \leq f(x) \leq M$  for all  $x \in [a, b]$ . Then for any  $\epsilon > 0$ , first (1) we construct  $p$  small open intervals on  $[a, b]$  containing the  $p$  discontinuity points with

$$\text{total length} < \frac{\epsilon}{2(M-m)}.$$

Next (2) for any subintervals in  $[a, b]$  excluding the above  $p$  subintervals,  $f$  is continuous on them, so there exists a partition such that

$$\sum_{(2)} \omega_i(f) \Delta x_i < \frac{\epsilon}{2}.$$

Now combine (1) and (2) to get

$$\sum_{i=1}^n \omega_i(f) \Delta x_i = \sum_{(1)} \omega_i(f) \Delta x_i + \sum_{(2)} \omega_i(f) \Delta x_i < (M-m) \frac{\epsilon}{2(M-m)} + \frac{\epsilon}{2} = \epsilon.$$

Thus  $f \in \mathcal{R}([a, b])$ , as desired.  $\square$

**Example 8.0.2.** Consider

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0 \\ A & \text{if } x = 0 \end{cases}$$

for any constant  $A \in \mathbb{R}$ . Then by the previous corollary,  $f \in \mathcal{R}([0, 1])$ .

**Theorem 8.2.** *If  $f, g \in \mathcal{R}([a, b])$ , then  $fg \in \mathcal{R}([a, b])$ .*

*Proof.* Since  $f, g$  are integrable, they are bounded. So assume  $|f|, |g| \leq M$ . Then for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that for any partition of width  $< \delta$ , we have

$$\sum_{i=1}^n \omega_i(f) \Delta x_i < \frac{\epsilon}{2M}, \quad \sum_{i=1}^n \omega_i(g) \Delta x_i < \frac{\epsilon}{2M}.$$

Notice

$$\omega_i(fg) \leq M(\omega_i(f) + \omega_i(g))$$

because

$$\begin{aligned} |f(x)g(x) - f(y)g(y)| &\leq |g(x)||f(x) - f(y)| + |f(y)||g(x) - g(y)| \\ &\leq M(|f(x) - f(y)| + |g(x) - g(y)|). \end{aligned}$$

Taking suprememes over  $x, y \in [x_{i-1}, x_i]$  from here gives  $\omega_i(fg) \leq M(\omega_i(f) + \omega_i(g))$ . Then

$$\sum_{i=1}^n \omega_i(fg) \Delta x_i \leq M \left( \sum_{i=1}^n \omega_i(f) \Delta x_i + \sum_{i=1}^n \omega_i(g) \Delta x_i \right) < M \left( \frac{\epsilon}{2M} + \frac{\epsilon}{2M} \right) = \epsilon.$$

Thus  $fg \in \mathcal{R}([a, b])$  as desired. □

**Theorem 8.3.** *If  $f \in \mathcal{R}([a, b])$ , then  $|f| \in \mathcal{R}([a, b])$  and*

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx.$$

*Proof.* Since  $f \in \mathcal{R}([a, b])$ , for any  $\epsilon > 0$  there exists a partition  $x_0 < \cdots < x_n$  such that

$$\sum_{i=1}^n \omega_i(f) \Delta x_i < \epsilon.$$

Since

$$||f(x)| - |f(y)|| \leq |f(x) - f(y)|,$$

taking supremums over  $x, y \in [x_{i-1}, x_i]$  gives  $\omega_i(|f|) \leq \omega_i(f)$ . Then

$$\sum_{i=1}^n \omega_i(|f|) \Delta x_i \leq \sum_{i=1}^n \omega_i(f) \Delta x_i < \epsilon.$$

So we indeed have  $|f| \in \mathcal{R}([a, b])$ . Now observe that  $-|f| \leq f \leq |f|$ . After integrating, we get

$$-\int_a^b |f(x)| dx \leq \int_a^b f(x) dx \leq \int_a^b |f(x)| dx.$$

This immediately implies the desired result. □

**Example 8.0.3** (Cauchy-Schwarz). If  $f, g \in \mathcal{R}([a, b])$ , then

$$\left| \int_a^b f(x)g(x) dx \right| \leq \left( \int_a^b f(x)^2 dx \right)^{1/2} \left( \int_a^b g(x)^2 dx \right)^{1/2}. \quad (*)$$

*Proof.* Let

$$A = \int_a^b f^2 dx, \quad B = \int_a^b |fg| dx, \quad C = \int_a^b g^2 dx.$$

Note that it suffices to show that  $B^2 \leq AC$ , which will imply  $(*)$  by the previous theorem. Then

$$0 \leq \int_a^b (t|f| - |g|)^2 dx = At^2 - 2Bt + C$$

for any  $t \in \mathbb{R}$ . So the discriminant must satisfy  $(2B)^2 - 4AC \leq 0$ , which gives  $B^2 \leq AC$  as desired.  $\square$

**Example 8.0.4** (Riemann-Lebesgue lemma). If  $f \in \mathcal{R}([a, b])$ , then

$$\lim_{\lambda \rightarrow \infty} \int_a^b f(x) \sin(\lambda x) dx = 0.$$

*Proof.* Since  $f \in \mathcal{R}([a, b])$ , for any  $\epsilon > 0$  there exists a partition  $x_0 < \cdots < x_n$  of  $[a, b]$  such that

$$\sum_{i=1}^n \omega_i(f) \Delta x_i < \frac{\epsilon}{2}.$$

Also assume  $|f| \leq M$  on  $[a, b]$  since  $f$  is integrable. Then we choose

$$\lambda > \frac{4nM}{\epsilon}.$$

We can estimate

$$\begin{aligned} \left| \int_a^b f(x) \sin(\lambda x) dx \right| &= \left| \sum_{i=1}^n \int_{x_{i-1}}^{x_i} (f(x) - f(x_i) + f(x_i)) \sin(\lambda x) dx \right| \\ &\leq \sum_{i=1}^n |f(x_i)| \left| \int_{x_{i-1}}^{x_i} \sin(\lambda x) dx \right| + \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \underbrace{|f(x) - f(x_i)|}_{\leq \omega_i(f)} \underbrace{|\sin(\lambda x)|}_{\leq 1} dx \\ &\leq M \sum_{i=1}^n \frac{\overbrace{|\cos(\lambda x_i) - \cos(\lambda x_{i-1})|}^{\leq 2}}{\lambda} + \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \omega_i(f) dx \\ &\leq M \frac{2n}{\lambda} + \sum_{i=1}^n \omega_i(f) \Delta x_i < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

So as  $\lambda \rightarrow \infty$ , the integral goes to 0.  $\square$

**Remark.** Recall that

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational} \\ 1 & \text{if } x \text{ is rational} \end{cases}$$

is not Riemann integrable, but we might expect that this should integrate to 0. The Lebesgue integral will fix this, which was discovered much later.

# Lecture 9

## Feb. 6 — Exchange of Limit Operations

### 9.1 Motivation

If we have a sequence of functions  $\{f_n\}$  where  $f_n \rightarrow f$  pointwise, then does

$$\int_a^b f_n dx \rightarrow \int_a^b f dx$$

if each  $f_n$  is integrable? Does  $f'_n \rightarrow f'$  if  $f_n$  is differentiable?

**Example 9.0.1.** Define

$$f_n(x) = \begin{cases} 4n^2x & \text{if } x \in [0, 1/2n] \\ 4n - 4n^2x & \text{if } x \in (1/2n, 1/n) \\ 0 & \text{if } x \in [1/n, 1], \end{cases}$$

where the graph of  $f_n$  looks like a triangle with peak at  $x = 1/2n$  and height  $2n$ . When we let  $n \rightarrow \infty$ , we see that for any  $x \in [0, 1]$ , we have  $f_n(x) \rightarrow 0$ . But

$$\int_0^1 f_n(x) dx = \text{area of triangle} = \frac{1}{2}(2n) \cdot \frac{1}{n} = 1.$$

So we see that in this case,

$$\lim_{n \rightarrow \infty} \int_0^1 f_n dx \neq \int_0^1 \lim_{n \rightarrow \infty} f_n dx.$$

### 9.2 Exchange of the Limit and Integral

**Theorem 9.1.** Let  $f_1, \dots, f_n, \dots$  be a uniformly convergent sequence of continuous functions on  $[a, b]$ . Then

$$\int_a^b \lim_{n \rightarrow \infty} f_n(x) dx = \lim_{n \rightarrow \infty} \int_a^b f_n(x) dx.$$

*Proof.* Suppose that  $f_n \rightarrow f$  uniformly. By definition of uniform convergence, for any  $\epsilon > 0$  there exists  $N$  such that if  $n \geq N$ , then

$$\max_{x \in [a, b]} |f_n(x) - f(x)| < \frac{\epsilon}{b - a}$$

Each  $f_n \rightarrow f$  uniformly and each  $f_n$  is continuous,  $f$  is also continuous. In particular,  $f$  is integrable and

$$-\frac{\epsilon}{b - a} < f_n(x) - f(x) < \frac{\epsilon}{b - a},$$

so integrating on both sides gives

$$-\epsilon < \int_a^b f_n(x) dx - \int_a^b f(x) dx < \epsilon \implies \left| \int_a^b f_n(x) dx - \int_a^b f(x) dx \right| < \epsilon.$$

Then this implies

$$\lim_{n \rightarrow \infty} \int_a^b f_n(x) dx = \int_a^b f(x) dx,$$

as desired.  $\square$

**Remark.** The previous theorem still holds even if each  $f_n$  is only Riemann integrable. The only thing we need to check is that the limit function  $f$  is also Riemann integrable. This is because for any  $\epsilon > 0$ , if  $n$  is large enough,

$$-\frac{\epsilon}{3(b-a)} + f_n(x) \leq f(x) \leq f_n(x) + \frac{\epsilon}{3(b-a)}.$$

Since  $f_n \in \mathcal{R}([a, b])$ , there exist two step functions  $g_1, g_2$  satisfying  $g_1 \leq f_n \leq g_2$ , and

$$\int_a^b (g_2 - g_1) < \frac{\epsilon}{3}.$$

Now note that

$$g_1(x) - \frac{\epsilon}{3(b-a)} \leq f(x) \leq g_2(x) + \frac{\epsilon}{3(b-a)},$$

so we see

$$\int_a^b \left[ \left( g_2(x) + \frac{\epsilon}{3(b-a)} \right) - \left( g_1(x) - \frac{\epsilon}{3(b-a)} \right) \right] dx = \frac{\epsilon}{3} + \frac{2\epsilon}{3} = \epsilon.$$

This gives  $f \in \mathcal{R}([a, b])$ , so we can carry through the rest of the previous proof.

## 9.3 Exchange of the Limit and Derivative

**Theorem 9.2.** Let  $f_1, \dots, f_n, \dots$  be a sequence of functions on an open interval  $U$  in  $\mathbb{R}$  and that each  $f_n$  has a continuous derivative. Suppose  $\{f'_n\}$  converges uniformly on  $U$  and for some  $a \in U$ ,  $\{f'_n(a)\}$  converges. Then

$$\lim_{n \rightarrow \infty} f_n(x) = f(x)$$

exists and  $f(x)$  is differentiable. Furthermore, we have

$$f' = \lim_{n \rightarrow \infty} f'_n.$$

*Proof.* By the fundamental theorem of calculus, we have

$$\int_a^x f'_n(t) dt = f_n(x) - f_n(a). \quad (*)$$

Let  $\lim_{n \rightarrow \infty} f'_n = g$ , where  $g$  is continuous since  $f'_n \rightarrow g$  uniformly and each  $f'_n$  is continuous. Then take  $n \rightarrow \infty$  in  $(*)$ , where

$$\text{LHS} \rightarrow \int_a^x g(t) dt.$$

Let  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ , which exists by  $(*)$ . Then  $\text{RHS} \rightarrow f(x) - f(a)$ , so we see that

$$f(x) - f(a) = \int_a^x g(t) dt.$$

Then  $f$  is an anti-derivative of  $g$ , or in other words,  $f' = g$  as desired.  $\square$

## 9.4 Infinite Series

**Definition 9.1.** Suppose we have a sequence of numbers  $a_1, a_2, a_3, \dots, a_n, \dots$ . Then

$$a_1 + a_2 + a_3 + \dots + a_n + \dots = \sum_{n=1}^{\infty} a_n$$

is called an *infinite series*. We say the infinite series *converges* to  $A$  if the *partial sums*

$$S_m = \sum_{n=1}^m a_n$$

converge to  $A$  as  $m \rightarrow \infty$ .

**Example 9.1.1** (Geometric series). For a fixed  $a$ , the series

$$\sum_{n=0}^{\infty} a^n = 1 + a + a^2 + \dots + a^n + \dots$$

converges if and only if  $|a| < 1$ , and the limit is  $1/(1 - a)$ . This is because

$$S_m = 1 + a + \dots + a^m = \frac{1 - a^{m+1}}{1 - a}.$$

If  $|a| < 1$ , then  $a^{m+1} \rightarrow 0$  as  $m \rightarrow \infty$ , so  $S_m \rightarrow 1/(1 - a)$ . On the other hand, if  $|a| > 1$ , then  $|a^{m+1}| \rightarrow \infty$  as  $m \rightarrow \infty$ . If  $a = 1$ , then

$$S_m = 1 + 1 + \dots + 1 = m,$$

so  $S_m \rightarrow \infty$ . If  $a = -1$ , then

$$\sum_{n=0}^{\infty} (-1)^n = 1 - 1 + 1 - 1 + \dots,$$

which diverges since its partial sums oscillate. So the condition is indeed necessary and sufficient.

**Proposition 9.1.** A series  $\sum_{n=1}^{\infty} a_n$  converges if and only if for every  $\epsilon > 0$ , there exists integer  $N$  such that if  $n > m \geq N$ , then

$$|a_{m+1} + a_{m+2} + \dots + a_n| < \epsilon.$$

*Proof.* Let  $S_m = \sum_{n=1}^m a_n$  be the partial sums. Then  $\sum_{n=1}^{\infty} a_n$  converges if and only if  $\{S_m\}$  is Cauchy. This is equivalent to say that for all  $\epsilon > 0$ , there exists  $N$  such that if  $n > m \geq N$ , then

$$|a_{m+1} + a_{m+2} + \dots + a_n| = |S_n - S_m| < \epsilon.$$

This is precisely the desired result. □

**Corollary 9.2.1.** If  $\sum_{n=1}^{\infty} a_n$  converges, then  $a_n \rightarrow 0$  as  $n \rightarrow \infty$ .

*Proof.* Take  $m = n - 1$  in the previous proposition, which gives  $|a_n| < \epsilon$  for  $n \geq N + 1$ . □

**Corollary 9.2.2.** If  $\sum_{n=1}^{\infty} a_n$  and  $\sum_{n=1}^{\infty} b_n$  differs in only finitely many terms, then the two series have the same convergence properties.



*Proof.* Simply take  $N$  larger than the last spot where the two series differ. Then the difference of partial sums in the previous proposition are the same for both series.  $\square$

**Example 9.1.2** (Harmonic series). The series

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} + \cdots$$

diverges. To see this, choose  $n = 2m$  in the previous proposition and

$$a_{m+1} + a_{m+2} + \cdots + a_{2m} = \frac{1}{m+1} + \frac{1}{m+2} + \cdots + \frac{1}{2m} \geq \frac{1}{2m}m = \frac{1}{2}.$$

So the series must diverge.

**Proposition 9.2.** *If  $a_n \geq 0$ , then  $\sum_{n=1}^{\infty} a_n$  either converges or has arbitrarily large partial sums, i.e. diverges to  $\infty$ .*

*Proof.* Let  $S_m = \sum_{n=1}^m a_n$ . Since  $a_n \geq 0$ , we see that  $S_m$  is an increasing nonnegative sequence. Then by the monotone convergence theorem,  $\{S_m\}$  converges if and only if it is bounded above.  $\square$

**Proposition 9.3** (Comparison test). *If  $\sum_{n=1}^{\infty} a_n$  and  $\sum_{n=1}^{\infty} b_n$  are two infinite series such that  $|a_n| \leq b_n$  and  $\sum_{n=1}^{\infty} b_n$  converges, then  $\sum_{n=1}^{\infty} a_n$  converges and*

$$\left| \sum_{n=1}^{\infty} a_n \right| \leq \sum_{n=1}^{\infty} b_n.$$

*Proof.* If  $\sum_{n=1}^{\infty} b_n$  converges, then for any  $\epsilon > 0$ , there exists  $N$  such that if  $n > m \geq N$ , we have

$$b_{m+1} + b_{m+2} + \cdots + b_n < \epsilon.$$

Then by the triangle inequality, we have

$$|a_{m+1} + a_{m+2} + \cdots + a_n| \leq |a_{m+1}| + |a_{m+2}| + \cdots + |a_n| \leq b_{m+1} + b_{m+2} + \cdots + b_n < \epsilon.$$

Thus  $\sum_{n=1}^{\infty} a_n$  also converges. The last part is left as an exercise.  $\square$

**Example 9.1.3** ( $p$ -series). The series

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges if and only if  $p > 1$ . We will show this with the integral test later.

**Proposition 9.4** (Ratio test). *If  $\sum_{n=1}^{\infty} a_n$  is a nonzero infinite series and there exists  $\rho < 1$  such that*

$$\left| \frac{a_{n+1}}{a_n} \right| \leq \rho$$

*for all  $n$  sufficiently large, then  $\sum_{n=1}^{\infty} a_n$  converges. If*

$$\left| \frac{a_{n+1}}{a_n} \right| \geq 1$$

*for all  $n$  large enough, then the series diverges.*

*Proof.* First we show the second part. If  $|a_{n+1}| \geq |a_n|$  for  $n \geq N$ , then

$$|a_n| \geq |a_{n-1}| \geq \cdots \geq |a_N|.$$

Then  $\{a_n\}$  does not converge to 0, so  $\sum_{n=1}^{\infty} a_n$  diverges. Look up the first part, but it should just be a comparison to the geometric series with common ratio  $\rho$ .  $\square$

# Lecture 10

## Feb. 8 — Infinite Series

### 10.1 Lots of Convergence Tests

#### 10.1.1 The Comparison Test

**Theorem 10.1** (Comparison test, second version). *Let  $\sum_{n=1}^{\infty} a_n$  and  $\sum_{n=1}^{\infty} b_n$  be two infinite series satisfying  $0 \leq a_n \leq b_n$ . Then*

1. *If  $\sum_{n=1}^{\infty} b_n$  converges, then  $\sum_{n=1}^{\infty} a_n$  converges.*
2. *If  $\sum_{n=1}^{\infty} a_n$  diverges, then  $\sum_{n=1}^{\infty} b_n$  diverges.*

*Proof.* (1) Let  $A_n$  and  $B_n$  be the partial sums of  $\sum_{n=1}^{\infty} a_n$  and  $\sum_{n=1}^{\infty} b_n$ , respectively. Then  $B_n$  is bounded above since  $\sum_{n=1}^{\infty} b_n$  converges. But  $A_n \leq B_n$  since  $0 \leq a_n \leq b_n$ , so  $A_n$  is also bounded above. Now note that  $A_n$  is increasing since  $a_n \geq 0$ , so by the monotone convergence theorem,  $A_n$  must converge.

(2) Since  $A_n$  is increasing,  $\sum_{n=1}^{\infty} a_n$  must diverge to  $\infty$ , i.e.  $A_n$  is unbounded. But  $A_n \leq B_n$ , so  $B_n$  is also unbounded and thus we see that  $\sum_{n=1}^{\infty} b_n$  diverges.  $\square$

**Remark.** In the above theorem, (1) remains true if

- $0 \leq a_n \leq b_n$  when  $n \geq n_0$
- $0 \leq a_n \leq Mb_n$  for some  $M > 0$ ,
- or there exists  $0 < d_n < M$  such that  $0 \leq a_n \leq d_nb_n$ .

**Corollary 10.1.1.** *If  $a_n, b_n > 0$  and*

$$\frac{a_{n+1}}{a_n} \leq \frac{b_{n+1}}{b_n},$$

*then  $\sum_{n=1}^{\infty} b_n$  converges implies that  $\sum_{n=1}^{\infty} a_n$  converges.*

*Proof.* Let  $d_n = a_n/b_n > 0$ . Then

$$d_{n+1} = \frac{a_{n+1}}{b_{n+1}} \leq \frac{a_n}{b_n} = d_n,$$

which we can extend to  $d_n \leq \dots \leq d_1$ . Then  $\{d_n\}$  is a bounded sequence, so  $a_n = b_nd_n$ , which implies the desired conclusion by the above remark.  $\square$

**Remark.** If  $n$  large,  $e^{an} \gg n^b \gg (\ln n)^c$  for any  $a, b, c > 0$ . In particular,  $e^{an}/n^b \rightarrow \infty$  when  $n \rightarrow \infty$ .

**Example 10.0.1.** Determine the convergence of

1.  $\sum_{n=2}^{\infty} \frac{1}{(\ln n)^p}$  for  $p > 0$ ,
2.  $\sum_{n=2}^{\infty} \frac{\ln(n!)}{n^p}$  for  $p > 0$ ,
3. and  $\sum_{n=1}^{\infty} \frac{n^{n-2}}{e^n n!}$ .

*Proof.* (1) We have

$$\frac{1}{(\ln n)^p} > \frac{1}{n}$$

for  $n$  large, so the sum diverges by comparison to the harmonic series.

(2) Note that

$$\ln(n!) = \sum_{k=1}^n \ln k > \frac{n \ln 2}{2},$$

so we have

$$\frac{\ln(n!)}{n^p} > \frac{\ln 2}{2} \frac{1}{n^{p-1}}.$$

By comparing to the  $p$ -series, we see that the series diverges when  $p \leq 2$ . Also we have

$$\frac{\ln(n!)}{n^p} < \frac{n \ln n}{n^p} = \frac{\ln n}{n^{p-1}},$$

so when  $p > 2$ , we get convergence.

(3) Let  $a_n = n^{n-2}/(e^n n!)$ . Recall that  $(1 + 1/n)^n \rightarrow e$  as  $n \rightarrow \infty$  and also  $(1 + 1/n)^n$  is increasing. Then

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)^{n-1} e^n n!}{e^{n+1} (n+1)! n^{n-2}} = \frac{\left(1 + \frac{1}{n}\right)^{n-2}}{e} = \underbrace{\frac{\left(1 + \frac{1}{n}\right)^n}{e}}_{<1} \left(1 + \frac{1}{n}\right)^{-2} < \left(\frac{n}{n+1}\right)^2 = \frac{\frac{1}{(n+1)^2}}{\frac{1}{n^2}}.$$

Let  $b_n = 1/n^2$ , then  $a_{n+1}/a_n \leq b_{n+1}/b_n$ , so by the previous corollary, we see that  $\sum_{n=1}^{\infty} a_n$  converges.  $\square$

**Theorem 10.2** (Comparison test, third version). *Let  $(A) \sim \sum_{n=1}^{\infty} a_n$  and  $(B) \sim \sum_{n=1}^{\infty} b_n$  be two positive series and suppose that*

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \ell > 0.$$

*Then  $(A)$  converges if and only if  $(B)$  converges.*

*Proof.* Let  $\epsilon = \ell/2 > 0$ . Then there exists  $N$  such that when  $n \geq N$ , we have

$$\frac{\ell}{2} < \frac{a_n}{b_n} < \frac{3\ell}{2} \implies \frac{\ell}{2} b_n < a_n < \frac{3\ell}{2} b_n.$$

Thus  $(A)$  converges if and only if  $(B)$  converges.  $\square$

**Example 10.0.2.** Determine the convergence of

1.  $\sum_{n=1}^{\infty} \frac{2n^2 + 5n + 1}{\sqrt{n^6 - 3n^2 + 1}},$
2.  $\sum_{n=1}^{\infty} \frac{1}{n^{1+\frac{1}{n}}},$
3. and  $\sum_{n=1}^{\infty} \left[ 1 - \sqrt[3]{\frac{n-1}{n+1}} \right]^p$  for  $p > 0$ .

*Proof.* (1) Let

$$a_n = \frac{2n^2 + 5n + 1}{\sqrt{n^6 - 3n^2 + 1}} \sim \frac{2n^2}{\sqrt{n^6}} = \frac{2}{n}.$$

Then  $a_n/(2/n) \rightarrow 1$  as  $n \rightarrow \infty$ , so  $\sum a_n$  diverges since the harmonic series diverges.

(2) Let

$$a_n = \frac{1}{n^{1+\frac{1}{n}}} \quad \text{and} \quad b_n = \frac{1}{n}.$$

Then  $a_n/b_n = 1/(n^{1/n}) \rightarrow 1$  as  $n \rightarrow \infty$ , so  $\sum a_n$  diverges since the harmonic series diverges.

(3) Write

$$\begin{aligned} \sqrt[3]{\frac{n-1}{n+1}} &= \sqrt[3]{\frac{1-\frac{1}{n}}{1+\frac{1}{n}}} = \left(1 - \frac{1}{n}\right)^{1/3} \left(1 + \frac{1}{n}\right)^{-1/3} \\ &= \left(1 - \frac{1}{3n} + o(1/n)\right) \left(1 - \frac{1}{3n} + o(1/n)\right) = 1 - \frac{2}{3n} + o(1/n), \end{aligned}$$

where we made a Taylor expansion. Then we see that

$$a_n \sim \left(\frac{2}{3n}\right)^p,$$

so  $\sum a_n$  converges if and only if  $p > 1$ . □

### 10.1.2 The Root Test

**Theorem 10.3** (Root test). *Let  $\sum_{n=1}^{\infty} a_n$  be a positive series and suppose that*

$$\limsup_{n \rightarrow \infty} \sqrt[n]{a_n} = \ell.$$

*Then*

1. *if  $\ell < 1$ , then  $\sum_{n=1}^{\infty} a_n$  converges,*
2. *and if  $\ell > 1$ , then  $\sum_{n=1}^{\infty} a_n$  diverges.*

*Proof.* (1) When  $\ell < 1$ , then there exists  $q$  with  $\ell < q < 1$  and  $N$  such that when  $n \geq N$ ,  $\sqrt[n]{a_n} < q$ . Then  $a_n < q^n$ , so  $\sum a_n$  converges by comparing to the geometric series.

(2) When  $\ell > 1$ , there exists  $\ell > q > 1$  and a subsequence  $\{n_k\}$  such that  $(a_{n_k})^{1/n_k} > q$ . This implies that  $a_{n_k} > q^{n_k}$ , so  $a_{n_k} \rightarrow \infty$  as  $n_k \rightarrow \infty$ . Thus  $\sum a_n$  diverges since we do not have  $a_n \rightarrow 0$ . □

**Example 10.0.3.** Determine the convergence of

1.  $\sum_{n=1}^{\infty} \left[ 1 + \frac{1}{\sqrt{n}} \right]^{-n^{3/2}},$
2. and  $\sum_{n=1}^{\infty} \left( \frac{3n}{n+5} \right)^n \left( \frac{n+2}{n+3} \right)^{n^2}.$

*Proof.* (1) Let  $a_n$  be the  $n$ th term in the sum and we see that

$$\sqrt[n]{a_n} = \left[ 1 + \frac{1}{\sqrt{n}} \right]^{-\sqrt{n}}.$$

Since  $\sqrt{n} \rightarrow \infty$  as  $n \rightarrow \infty$ , we may replace  $\sqrt{n}$  with  $n$  to see that  $\sqrt[n]{a_n} \rightarrow 1/e < 1$ , so  $\sum a_n$  converges.

(2) Let

$$\sqrt[n]{a_n} = \frac{3n}{n+5} \left( \frac{n+2}{n+3} \right)^n.$$

For the second term, we see that

$$\left( \frac{n+2}{n+3} \right)^n = \left( \frac{1+2/n}{1+3/n} \right)^n \longrightarrow \frac{e^2}{e^3} = \frac{1}{e}$$

as  $n \rightarrow \infty$ . Then  $\sqrt[n]{a_n} \rightarrow 3/e > 1$ , so  $\sum a_n$  diverges. □

**Remark.** In the above, we used the fact that

$$\lim_{n \rightarrow \infty} \left( 1 + \frac{a}{n} \right)^{bn} = e^{ab}.$$

# Lecture 11

## Feb. 15 — Absolute Convergence

### 11.1 Example of the Root Test

**Example 11.0.1.** Determine the converge of the series

$$\sum_{n=1}^{\infty} \frac{n!}{n^{\sqrt{n}}}.$$

*Proof.* By Stirling's formula,

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n},$$

in the sense that their ratio tends to 1 as  $n \rightarrow \infty$ . Then

$$\sqrt[n]{a_n} \sim \frac{n}{e} (2\pi)^{1/2n} n^{1/2n-1/\sqrt{n}},$$

Note that  $(2\pi)^{1/2n} \rightarrow 1$  as  $n \rightarrow \infty$  since  $2\pi$  is a constant, and

$$\ln(n^{1/2n-1/\sqrt{n}}) = \left(\frac{1}{2n} - \frac{1}{\sqrt{n}}\right) \ln n = \frac{\ln n}{\frac{1}{1/2n-1/\sqrt{n}}} = \frac{\ln n}{2n/\sqrt{1-2\sqrt{n}}} \sim \frac{\ln n}{2n} = \frac{1/n}{2} \rightarrow 0$$

by L'Hôpital's rule, so  $n^{1/2n-1/\sqrt{n}} \rightarrow 1$  as  $n \rightarrow \infty$ . Then  $\sqrt[n]{a_n} \sim n/e \gg 1$ , so this series diverges.  $\square$

### 11.2 The Integral Test

**Theorem 11.1** (Integral test). *Let  $\{a_n\}$  be a positive decreasing sequence. If there exists a continuous decreasing  $f(x)$  on  $[1, \infty)$  such that  $a_n = f(n)$ , then*

$$\sum_{n=1}^{\infty} a_n \text{ converges} \quad \text{if and only if} \quad \int_1^{\infty} f(x) dx \text{ converges.}$$

*Proof.* ( $\Leftarrow$ ) Suppose that

$$\int_1^{\infty} f(x) dx.$$

converges, i.e. the limit

$$\lim_{A \rightarrow \infty} \int_1^A f(x) dx$$

exists. Then

$$a_k = f(k) = f(k)[k - (k-1)] \leq \int_{k-1}^k f(x) dx$$

since  $f$  is decreasing. So

$$\sum_{k=2}^n a_k \leq \sum_{k=2}^n \int_{k-1}^k f(x) dx = \int_1^n f(x) dx.$$

Since the integral converges as  $n \rightarrow \infty$ , the partial sums of  $\sum_{n=1}^{\infty} a_n$  are bounded, so  $\sum_{n=1}^{\infty} a_n$  converges by the monotone convergence theorem since  $a_n \geq 0$ .

( $\Rightarrow$ ) Suppose

$$\sum_{n=1}^{\infty} a_n$$

converges. Then

$$a_k = f(k)[(k+1) - k] \geq \int_k^{k+1} f(x) dx$$

since  $f$  is decreasing. So

$$\sum_{k=1}^n a_k \geq \sum_{k=1}^n \int_k^{k+1} f(x) dx = \int_1^{n+1} f(x) dx.$$

Since the sum converges as  $n \rightarrow \infty$ , the integral is bounded and thus converges as  $f(x) \geq 0$ .  $\square$

**Example 11.0.2.** For  $p > 1$ , show that

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges if and only if  $p > 1$ .

*Proof.* Let  $a_n = 1/n^p$  and choose

$$f(x) = \frac{1}{x^p}$$

for  $x > 0$  and note that  $f(n) = a_n$ . Then look at the integral

$$\int_1^{\infty} \frac{1}{x^p} dx,$$

where we can appeal to integration rules to see that this integral converges if and only if  $p > 1$ .  $\square$

**Example 11.0.3.** Show that

$$\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^p}$$

converges if and only if  $p > 1$ .

*Proof.* Look at

$$\int_2^{\infty} \frac{1}{x(\ln x)^p} dx = \left[ \frac{1}{-p+1} (\ln x)^{-p+1} \right]_{x=2}^{x=\infty},$$

which converges if and only if  $p > 1$ .  $\square$



**Example 11.0.4.** Suppose  $a_n > 0$  and let  $S_n = \sum_{k=1}^n a_k$ . Then

1.  $\sum_{n=1}^{\infty} \frac{a_{n+1}}{S_n \ln S_n}$  diverges if  $\sum_{n=1}^{\infty} a_n$  diverges,
2. and  $\sum_{n=1}^{\infty} \frac{a_{n+1}}{S_n (\ln S_n)^2}$  converges if  $\sum_{n=1}^{\infty} a_n$  converges.

*Proof.* (1) Notice  $a_n = S_n - S_{n-1} > 0$ , so

$$\frac{a_{n+1}}{S_n \ln S_n} = \frac{S_{n+1} - S_n}{S_n \ln S_n} \geq \int_{S_n}^{S_{n+1}} \frac{1}{x \ln x} dx = \ln(\ln S_{n+1}) - \ln(\ln S_n).$$

So we see that

$$\sum_{k=1}^n \frac{a_{k+1}}{S_k \ln S_k} \geq \ln(\ln S_{n+1}) - \ln(\ln a_1)$$

since the sum telescopes. But  $\ln(\ln S_{n+1}) \rightarrow \infty$  as  $n \rightarrow \infty$ , so this sum diverges.

The proof for (2) is left as an exercise, but the idea is similar. □

**Remark.** If  $a_n = 1$ , then  $S_n = n$  for all  $n$ , which implies that

$$\sum_{n=1}^{\infty} \frac{1}{n \ln n} \text{ diverges} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{n (\ln n)^2} \text{ converges,}$$

which matches the previous example.

**Example 11.0.5.** Let  $f(x_0)$  be positive and decreasing on  $[0, \infty)$ . Then

$$(A) = \int_a^{\infty} f(x) dx \text{ converges} \quad \text{if and only if} \quad (B) = \int_a^{\infty} f(x) \sin^2 x dx \text{ converges.}$$

*Proof.* ( $\Rightarrow$ ) This direction is obvious since  $0 \leq f(x) \sin^2 x \leq f(x)$  at every point.

( $\Leftarrow$ ) Suppose otherwise that (A) diverges. Then

$$\infty = \int_a^{\infty} f(x) dx = \sum_{n=0}^{\infty} \int_{a+n\pi}^{a+(n+1)\pi} f(x) dx \leq \pi \sum_{n=0}^{\infty} f(a+n\pi)$$

since  $f$  is decreasing. This implies that  $\sum_{n=0}^{\infty} f(a+n\pi)$  diverges. But then

$$\int_a^{\infty} f(x) \sin^2 x dx \geq \sum_{n=0}^{\infty} f(a+(n+1)\pi) \int_{a+n\pi}^{a+(n+1)\pi} \sin^2 x dx = \frac{\pi}{2} \sum_{n=0}^{\infty} f(a+(n+1)\pi),$$

so we see that

$$\int_a^{\infty} f(x) \sin^2 x dx$$

diverges since  $\sum_{n=1}^{\infty} f(a+n\pi)$  diverges. Contradiction. □

## 11.3 Absolute and Conditional Convergence

**Definition 11.1.** For a series  $(A) = \sum_{n=1}^{\infty} a_n$ , if  $\sum_{n=1}^{\infty} |a_n|$  converges, then we say that  $(A)$  *converges absolutely*. If  $(A)$  converges but  $\sum_{n=1}^{\infty} |a_n|$  diverges, then we say that  $(A)$  *converges conditionally*.

**Example 11.1.1.** The series

$$\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

converges conditionally. This is because the series itself converges by the alternating series test (which we will see later), but taking absolute values gives the harmonic series, which diverges.

**Example 11.1.2.** For  $p > 1$ , the series

$$\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n^p}$$

converges absolutely. Unlike before, taking absolute values gives a  $p$ -series, which converges for  $p > 1$ .

**Definition 11.2.** For any  $a_n \in \mathbb{R}$ , define the *positive* and *negative parts* of  $a_n$  by

$$a_n^+ = \begin{cases} a_n & \text{if } a_n \geq 0, \\ 0 & \text{if } a_n < 0, \end{cases} \quad \text{and} \quad a_n^- = \begin{cases} -a_n & \text{if } a_n \leq 0, \\ 0 & \text{if } a_n > 0. \end{cases}$$

Note that  $a_n^+, a_n^- \geq 0$  and in particular,  $a_n = a_n^+ - a_n^-$  and  $|a_n| = a_n^+ + a_n^-$ .

**Theorem 11.2.** Suppose  $\sum_{n=1}^{\infty} a_n$  converges absolutely. Then

1.  $\sum_{n=1}^{\infty} a_n^+$  and  $\sum_{n=1}^{\infty} a_n^-$  converge absolutely,
2. and  $|\sum_{n=1}^{\infty} a_n| \leq \sum_{n=1}^{\infty} |a_n|$ .

*Proof.* (1) Write

$$\sum_{n=1}^{\infty} |a_n| = \sum_{n=1}^{\infty} (a_n^+ + a_n^-).$$

Since  $a_n^+, a_n^- \geq 0$ , this implies that  $\sum_{n=1}^{\infty} a_n^+$  and  $\sum_{n=1}^{\infty} a_n^-$  converge since they are bounded.

(2) Note that

$$\left| \sum_{k=1}^n a_k \right| \leq \sum_{k=1}^n |a_k|$$

in the finite case. Then let  $n \rightarrow \infty$  to get

$$\left| \sum_{k=1}^{\infty} a_k \right| \leq \sum_{k=1}^{\infty} |a_k|,$$

which is the desired result. □

**Example 11.2.1.** Does the series

$$\sum_{n=1}^{\infty} \left[ \frac{\cos n}{\sqrt[3]{n^2}} - \sin \left( \frac{\cos n}{\sqrt[3]{n^2}} \right) \right]$$

converge absolutely?

*Proof.* Let

$$a_n = \frac{\cos n}{\sqrt[3]{n^2}} - \sin\left(\frac{\cos n}{\sqrt[3]{n^2}}\right)$$

and Taylor expand to get

$$a_n = \frac{\cos n}{\sqrt[3]{n^2}} - \left(\frac{\cos n}{\sqrt[3]{n^2}} + O(1/n^2)\right) = O(1/n^2).$$

Since  $\sum_{n=1}^{\infty} 1/n^2$  converges, we see that  $\sum_{n=1}^{\infty} a_n$  converges absolutely.  $\square$

### 11.3.1 The Alternating Series Test

**Definition 11.3.** A series of the form

$$\sum_{n=1}^{\infty} (-1)^n a_n = a_1 - a_2 + a_3 - a_4 + \dots$$

where  $a_n > 0$  is called an *alternating series*.

**Theorem 11.3** (Leibniz test). *Suppose  $\{a_n\}_{n=1}^{\infty}$  is a decreasing sequence and  $\lim_{n \rightarrow \infty} a_n = 0$ . Then the infinite series  $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$  converges.*

*Proof.* Let

$$S_n = \sum_{k=1}^n (-1)^{k+1} a_k.$$

Then we have

$$S_{2n} = (a_1 - a_2) + (a_3 - a_4) + \dots + (a_{2n-1} - a_{2n}) \geq S_{2n-2} \geq 0$$

and also that

$$S_{2n} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \dots - (a_{2n-2} - a_{2n-1}) - a_{2n} \leq a_1$$

since  $\{a_n\}$  is decreasing. In particular,  $S_{2n}$  is bounded and increasing, so

$$S = \lim_{n \rightarrow \infty} S_{2n}$$

exists by the monotone convergence theorem. Now observe that

$$S_{2n+1} = S_{2n} + a_{2n+1},$$

so taking  $n \rightarrow \infty$  gives  $S_{2n+1} \rightarrow S$  since  $S_{2n} \rightarrow S$  and  $a_{2n+1} \rightarrow 0$ . Thus  $\lim_{n \rightarrow \infty} S_n = S$ .  $\square$

**Remark.** This shows that the series

$$\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n}$$

from earlier indeed converges.

**Remark.** The proof also gives an error estimate for an alternating series. We have

$$|S_n - S| \leq a_{n+1}$$

since  $|S_n - S| = a_{n+1} - (a_{n+2} - a_{n+3}) + \dots \leq a_{n+1}$ . We get this since the tail is still an alternating series, and an alternating series is bounded by its first term, as shown in the proof.

# Lecture 12

## Feb. 20 — Series of Functions

### 12.1 Rearrangements of Series

**Definition 12.1.** Let  $\sigma : \mathbb{N} \rightarrow \mathbb{N}$  be a bijection, i.e. it is one-to-one and onto. Given an infinite series  $\sum_{n=1}^{\infty} a_n$ , we define the *rearranged series* to be  $\sum_{n=1}^{\infty} a_{\sigma(n)}$ .

**Theorem 12.1** (Dirichlet). *If  $(A) = \sum_{n=1}^{\infty} a_n$  converges absolutely and  $(B) = \sum_{n=1}^{\infty} a_{\sigma(n)}$  is any rearrangement of  $(A)$ , then  $(B)$  also converges and*

$$\sum_{n=1}^{\infty} a_{\sigma(n)} = \sum_{n=1}^{\infty} a_n.$$

*Proof.* First assume  $a_n \geq 0$  and let  $S = \sum_{n=1}^{\infty} a_n$  with partial sums  $S_n = \sum_{k=1}^n a_k$ . Let  $\sigma_n$  be the partial sums of  $(B)$ . Then

$$\sigma_m \leq \sum_{n=1}^{\infty} a_n = S \quad (*)$$

since  $a_n \geq 0$ . Then  $\sigma_m$  is bounded, so  $(B)$  converges by the monotone convergence theorem. Also by  $(*)$ , we can take  $m \rightarrow \infty$  to get

$$\lim_{m \rightarrow \infty} \sigma_m \leq S \implies \sum_{n=1}^{\infty} a_{\sigma(n)} \leq S.$$

We also get the reverse inequality by thinking of  $(A)$  as a rearrangement of  $(B)$ , so the two are equal.

Now for the general case, suppose that  $\sum_{n=1}^{\infty} a_n$  is a series with differing signs. Let  $a_n^+, a_n^- \geq 0$  be the positive and negative parts of  $a_n$ , respectively, so that  $|a_n| = a_n^+ + a_n^-$ . Now  $b_n = a_{\sigma(n)}^+$ , and  $|b_n| = b_n^+ + b_n^-$ , so  $b_n^{\pm} = a_{\sigma(n)}^{\pm}$ . By the previous part,  $\sum_{n=1}^{\infty} b_n^{\pm}$  converges and

$$\sum_{n=1}^{\infty} b_n^{\pm} = \sum_{n=1}^{\infty} a_n^{\pm}.$$

Then we have

$$\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} (b_n^+ - b_n^-) = \sum_{n=1}^{\infty} (a_n^+ - a_n^-) = \sum_{n=1}^{\infty} a_n,$$

which is the desired conclusion. □

**Theorem 12.2** (Riemann). *If  $(A) = \sum_{n=1}^{\infty} a_n$  converges conditionally, then there exists a rearrangement  $(B) = \sum_{n=1}^{\infty} a_{\sigma(n)}$  such that any one of the following occurs:*

1.  $(B)$  converges to any number  $\sigma$ ,
2.  $(B)$  diverges to  $\infty$ ,
3.  $(B)$  diverges to  $-\infty$ ,
4.  $(B)$  oscillates in an unbounded manner,
5. or  $(B)$  oscillates in a bounded manner.

*Proof.* Look this up. □

**Remark.** This shows that when a series only converges conditionally, then pretty much anything can happen when rearranging, which is not the case for absolutely convergent series.

## 12.2 Series of Functions

**Definition 12.2.** A *series of functions* is a series of the form  $\sum_{n=1}^{\infty} f_n(x)$ , where each  $f_n(x)$  is defined on an interval  $I$ . Define its *partial sums* to be  $S_n(x) = \sum_{k=1}^n f_k(x)$ . We say that

$$\sum_{n=1}^{\infty} f_n(x) = S(x),$$

i.e. the series *converges* to  $S(x)$ , if  $\lim_{n \rightarrow \infty} S_n(x) = S(x)$ , which can be *pointwise* or *uniform*.

**Theorem 12.3** (Cauchy criterion for uniform convergence). *A series  $\sum_{n=1}^{\infty} f_n(x)$  converges uniformly on  $I$  if and only if for every  $\epsilon > 0$ , there exists  $N$  such that whenever  $n \geq N$ , for any  $x \in I$  and integer  $p$  we have*

$$\left| \sum_{k=n+1}^{n+p} f_k(x) \right| < \epsilon.$$

*Proof.* This is because  $\{S_n(x)\}$  is a Cauchy sequence, so for any  $\epsilon > 0$ , there exists  $N$  such that when  $n \geq N$ ,

$$\sup_{x \in I} |S_{n+p}(x) - S_n(x)| < \epsilon,$$

which is precisely the given condition. □

**Corollary 12.3.1.** *If  $\sum_{n=1}^{\infty} f_n(x)$  converges uniformly on  $I$ , then  $f_n(x)$  converges uniformly to 0 on  $I$ .*

**Corollary 12.3.2.** *Let  $f_n(x) \in C([a, b])$ . If  $\sum_{n=1}^{\infty} f_n(x)$  converges uniformly on  $(a, b)$ , then it converges uniformly on  $[a, b]$ .*

*Proof.* Since  $\sum_{n=1}^{\infty} f_n(x)$  converges uniformly, for any  $\epsilon > 0$  there exists  $N$  such that when  $n \geq N$ , we have

$$\left| \sum_{k=n+1}^{n+p} f_k(x) \right| < \epsilon$$

for all  $x \in (a, b)$ . Let  $x \rightarrow a^+$  and  $x \rightarrow b^-$  to get

$$\left| \sum_{k=n+1}^{n+p} f_k(a) \right| \leq \epsilon \quad \text{and} \quad \left| \sum_{k=n+1}^{n+p} f_k(b) \right| \leq \epsilon.$$

So for all  $x \in [a, b]$ , we have

$$\left| \sum_{k=n+1}^{n+p} f_k(x) \right| \leq \epsilon,$$

which gives uniform convergence on  $[a, b]$ . □

**Example 12.2.1.** The series

$$\sum_{n=1}^{\infty} \frac{1}{n^x}$$

does not converge uniformly on  $(1, \infty)$ .

*Proof.* Note that each  $1/n^x$  is continuous but at  $x = 1$ , the series becomes the harmonic series  $\sum_{n=1}^{\infty} 1/n$ , which diverges. So by the previous corollary, this series cannot converge uniformly on  $(1, \infty)$ . □

**Theorem 12.4** (Weierstrass  $M$ -test). *If there exists a nonnegative and convergent series  $\sum_{n=1}^{\infty} M_n$  such that for all  $x \in I$ , we have*

$$|f_n(x)| \leq M_n,$$

*then  $\sum_{n=1}^{\infty} f_n(x)$  converges uniformly on  $I$ .*

*Proof.* Since  $\sum_{n=1}^{\infty} M_n$  converges uniformly, for any  $\epsilon > 0$  there exists  $N$  such that

$$\left| \sum_{k=n+1}^{n+p} M_k \right| < \epsilon.$$

This directly implies

$$\left| \sum_{k=n+1}^{n+p} f_k(x) \right| \leq \sum_{k=n+1}^{n+p} |f_k(x)| < \epsilon$$

since each  $M_k$  is nonnegative. Then  $\sum_{n=1}^{\infty} f_n(x)$  converges uniformly on  $I$  by the Cauchy criterion. □

**Example 12.2.2.** If  $\sum_{n=1}^{\infty} a_n$  converges absolutely, then

$$\sum_{n=1}^{\infty} \frac{a_n x^n}{1 + x^{2n}}$$

converges uniformly on  $(-\infty, \infty)$ .

*Proof.* This is because

$$\left| \frac{a_n x^n}{1 + x^{2n}} \right| \leq |a_n| \underbrace{\frac{|x|^n}{1 + |x|^{2n}}}_{\leq 1} \leq |a_n|,$$

so by the Weierstrass  $M$ -test, we get uniform convergence on  $(-\infty, \infty)$ . □

## 12.3 Exchange of the Limit and Infinite Series

**Theorem 12.5.** Let  $f_n \in \mathcal{R}([a, b])$  for each  $n = 1, 2, \dots$ . If  $\sum_{n=1}^{\infty} f_n(x)$  converges uniformly to  $S(x)$  on  $[a, b]$ , then  $S(x) \in \mathcal{R}([a, b])$  and

$$\sum_{n=1}^{\infty} \int_a^b f_n(x) dx = \int_a^b S(x) dx = \int_a^b \sum_{n=1}^{\infty} f_n(x) dx.$$

*Proof.* Let  $S_n(x) = \sum_{k=1}^n f_k(x)$ . Then  $S_n \rightarrow S$  uniformly on  $[a, b]$ , so we can apply Theorem 9.1 to get  $S(x) \in \mathcal{R}([a, b])$  and

$$\sum_{n=1}^{\infty} \int_a^b f_n(x) dx = \lim_{n \rightarrow \infty} \int_a^b S_n(x) dx = \int_a^b S(x) dx,$$

which is the desired conclusion.<sup>1</sup> □

**Theorem 12.6.** Let  $f_n(x)$  be continuously differentiable on  $[a, b]$ . Suppose that

$$\sum_{n=1}^{\infty} f_n(x) = S(x)$$

pointwise on  $[a, b]$ , and  $\sum_{n=1}^{\infty} f'_n(x)$  converges uniformly to  $G(x)$  on  $[a, b]$ . Then  $\sum_{n=1}^{\infty} f_n(x)$  converges uniformly to  $S(x)$  and  $S'(x) = G(x)$ . In other words, in this case we have

$$\left( \sum_{n=1}^{\infty} f_n(x) \right)' = \sum_{n=1}^{\infty} f'_n(x).$$

*Proof.* Similarly apply Theorem 9.2, our previous theorem on exchanging the limit and the derivative. □

**Example 12.2.3.** Show that

$$S(x) = \sum_{n=1}^{\infty} \frac{\sin nx}{n^3}$$

is continuously differentiable on  $(-\infty, \infty)$ .

*Proof.* For each fixed  $n$ , we see that  $(\sin nx)/n^3$  is continuously differentiable, and we at least have pointwise convergence since

$$\left| \frac{\sin nx}{n^3} \right| \leq \frac{1}{n^3}.$$

Now the series of derivatives

$$\sum_{n=1}^{\infty} \frac{n \cos nx}{n^3} = \sum_{n=1}^{\infty} \frac{\cos nx}{n^2}$$

converges uniformly on  $(-\infty, \infty)$  by the Weierstrass  $M$ -test since

$$\left| \frac{\cos nx}{n^2} \right| \leq \frac{1}{n^2}.$$

So we can apply the previous theorem to see that  $S$  is continuously differentiable on  $(-\infty, \infty)$ . □

---

<sup>1</sup>This is just applying the previous theorem on exchanging the limit and the integral.

**Remark.** Changing the  $n^3$  to  $n^2$  in the previous example will cause this argument to fail, since we get

$$\left| \frac{\cos nx}{n} \right| \leq \frac{1}{n},$$

which does not converge. So we cannot use the Weierstrass  $M$ -test.

**Example 12.2.4.** Let

$$S(x) = \sum_{n=1}^{\infty} \frac{|x|}{n^2 + x^2}.$$

Study its differentiability on  $(-\infty, \infty)$ .

*Proof.* Observe that

$$\frac{|x|}{n^2 + x^2} \leq \frac{|x|}{n^2},$$

so the series converges pointwise on  $(-\infty, \infty)$ . Now let  $f_n(x) = |x|/(n^2 + x^2)$ , and we see that

$$\sum_{n=1}^{\infty} f'_n(x) = \sum_{n=1}^{\infty} \frac{\frac{(n^2+x^2)|x|}{x} - 2x|x|}{(n^2 + x^2)^2} = \sum_{n=1}^{\infty} \frac{\frac{n^2|x|}{x} - x|x|}{(n^2 + x^2)^2}$$

for any  $0 < |x| < A$ . Now

$$\left| \frac{n^2|x|}{x} - x|x| \right| \leq n^2 + A^2,$$

so  $\sum_{n=1}^{\infty} f'_n(x)$  converges uniformly for any  $x$  away from 0. So  $S(x)$  is differentiable when  $x \neq 0$ . At  $x = 0$ , we use the definition of the derivative to see that

$$\lim_{\Delta x \rightarrow 0} \frac{S(\Delta x) - S(0)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{|\Delta x|}{\Delta x} \sum_{n=1}^{\infty} \frac{1}{n^2 + (\Delta x)^2}.$$

From this we get

$$S'_-(0) = -\sum_{n=1}^{\infty} \frac{1}{n^2} \neq \sum_{n=1}^{\infty} \frac{1}{n^2} = S'_+(0),$$

so  $S(x)$  is not differentiable at  $x = 0$ . □



# Lecture 13

## Feb. 22 — Power Series

### 13.1 Power Series

**Definition 13.1.** Given a point  $x_0$ , a *power series* around  $x_0$  is a series of the form

$$\sum_{n=0}^{\infty} a_n(x - x_0)^n = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \dots$$

**Remark.** The question is: For what  $x$  does the series converge?

**Example 13.1.1.** Consider the series<sup>1</sup>

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

In fact this series converges for all  $x \in \mathbb{R}$ . Let  $a_n = x^n/n!$ . By the ratio test, for  $x \neq 0$ , we have

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = \lim_{n \rightarrow \infty} \frac{|x|}{n+1} = 0,$$

so the series converges at  $x \neq 0$ . Of course, if  $x = 0$ , then every term past  $n = 0$  is zero, so the series also converges there. Thus the series converges everywhere on  $\mathbb{R}$ .

**Example 13.1.2.** The series

$$\sum_{n=0}^{\infty} n!x^n$$

converges only at  $x = 0$ . For  $x \neq 0$ , we can similarly apply the ratio test to find that

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = \lim_{n \rightarrow \infty} |n+1||x| = \infty,$$

so the series diverges for  $x \neq 0$ .

**Example 13.1.3.** Recall that the geometric series

$$\sum_{n=0}^{\infty} x^n$$

converges if and only if  $|x| < 1$ .

---

<sup>1</sup>Later we will see that this is the Taylor series for the exponential function  $e^x$ .

## 13.2 Radius of Convergence

**Lemma 13.1.** *We have the following:*

1. If  $(A) = \sum_{n=0}^{\infty} a_n x^n$  converges at  $x = x_1 \neq 0$ , then it converges absolutely for all  $x$  with  $|x| < |x_1|$ .
2. If  $(A)$  diverges at  $x = x_2 \neq 0$ , then it diverges for all  $x$  with  $|x| > |x_2|$ .

*Proof.* (1) If  $\sum_{n=0}^{\infty} a_n x_1^n$  converges, then  $a_n x_1^n \rightarrow 0$  as  $n \rightarrow \infty$ . In particular, the terms are bounded, so there exists  $M > 0$  such that  $|a_n x_1^n| \leq M$  for every  $n \in \mathbb{N}$ . For any  $|x| < |x_1|$ , let

$$q = \left| \frac{x}{x_1} \right| < 1.$$

Then we have

$$|a_n x^n| = \left| a_n x_1^n \left( \frac{x}{x_1} \right)^n \right| \leq M q^n.$$

Comparing with the geometric series, we get that  $\sum_{n=0}^{\infty} a_n x^n$  converges absolutely when  $|x| < |x_1|$ .

(2) Suppose otherwise that there exists  $x_3$  such that  $|x_3| > |x_2|$  and  $\sum_{n=0}^{\infty} a_n x_3^n$  converges. Then by (1), we see that the power series converges at  $x = x_2$ . Contradiction.  $\square$

**Corollary 13.0.1.** *If  $(A) = \sum_{n=0}^{\infty} a_n x^n$  converges at some  $x_1 \neq 0$  and diverges at  $x_2 \neq 0$ , then there exists  $R > 0$  such that  $(A)$  converges for  $|x| < R$  and diverges for  $|x| > R$ .*

*Proof.* Let  $E$  be the set of all convergence points of  $(A)$ . By Lemma 13.1, we have  $E \subseteq \{|x| \leq |x_1|\}$  since  $(A)$  diverges for all  $|x| > |x_2|$ . Let

$$R = \sup\{|x| : x \in E\},$$

which exists since  $E$  is nonempty and bounded above by  $|x_2|$ . Also  $R > 0$  since  $x_1 \in E$  and  $x_1 \neq 0$ . Now if  $|x| < R$ , then there exists  $x_1 \in E$  such that  $|x| < |x_1| < R$  and  $x = x_1$  is a convergence point.<sup>2</sup> By Lemma 13.1, we get that  $(A)$  converges at  $x$ . If  $|x| > R$ , then there exists  $x_2$  such that  $|x| > |x_2| > R$  and  $(A)$  diverges at  $x = x_2$ . Then by Lemma 13.1,  $(A)$  diverges at  $x$ .  $\square$

**Remark.** The  $R$  in Corollary 13.0.1 is called the *radius of convergence* of the power series. If  $(A) = \sum_{n=0}^{\infty} a_n x^n$  converges for all  $x \in \mathbb{R}$ , then by convention we use  $R = \infty$ . If  $(A)$  converges only at  $x = 0$ , we use  $R = 0$ . At  $x = \pm R$ , the convergence or divergence of  $(A)$  needs to be checked separately.

**Theorem 13.1.** *For a power series  $\sum_{n=0}^{\infty} a_n x^n$ , let*

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} = \rho.$$

*Then*

1. if  $\rho = 0$ , then  $R = \infty$ ,
2. if  $\rho = \infty$ , then  $R = 0$ ,
3. and if  $\rho \in (0, \infty)$ , then  $R = 1/\rho$ .

---

<sup>2</sup>This is by definition of the supremum.

*Proof.* We use the root test for  $\sum_{n=0}^{\infty} |a_n x^n|$ . Let  $A_n = |a_n x^n|$ . Then

$$\limsup_{n \rightarrow \infty} \sqrt[n]{A_n} = \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} \cdot |x| = \rho |x|.$$

First suppose  $0 < \rho < \infty$ . Then  $\sum_{n=0}^{\infty} |a_n x^n|$  converges if  $|x|\rho < 1$  and diverges if  $|x|\rho > 1$ . This gives  $R = 1/\rho$ . Now if  $\rho = 0$ , then

$$\limsup_{n \rightarrow \infty} \sqrt[n]{A_n} = 0 < 1,$$

so  $\sum_{n=0}^{\infty} |a_n x^n|$  regardless of the choice of  $x$ , i.e.  $R = \infty$ . Finally if  $\rho = \infty$ , then for  $x_0 \neq 0$ ,

$$\limsup_{n \rightarrow \infty} \sqrt[n]{A_n} = \infty > 1.$$

By the root test, this implies that  $\sum_{n=0}^{\infty} |a_n x^n|$  diverges for all  $x \neq 0$ , i.e.  $R = 0$ . □

**Corollary 13.1.1.** For  $\sum_{n=0}^{\infty} a_n x^n$  with  $a_n \neq 0$ , if

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = \rho,$$

then the radius of convergence is  $R = 1/\rho$ .

*Proof.* Left as an exercise. □

**Example 13.1.4.** Find the convergence intervals for

1.  $\sum_{n=1}^{\infty} \frac{2^n (x+1)^n}{n \ln^2(n+1)},$
2. and  $\sum_{n=1}^{\infty} n^n x^{n^2}.$

*Proof.* (1) Let  $a_n$  be the summand. Then

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = \lim_{n \rightarrow \infty} \frac{2n \ln^2(n+1)}{(n+1) \ln^2(n+2)} = 2,$$

so  $R = 1/2$ . So we get convergence for  $|x+1| < 1/2$ , i.e.  $x \in (-3/2, -1/2)$ . At  $x = -3/2, -1/2$ , we get

$$\sum_{n=1}^{\infty} \frac{1}{n \ln^2(n+1)}$$

after taking absolute values, which converges by the integral test. So the interval is  $[-3/2, -1/2]$ .

(2) Let the general term be

$$a_k = \begin{cases} n^n & \text{if } k = n^2 \text{ for some } n \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}$$

Then we see that

$$\limsup_{k \rightarrow \infty} \sqrt[k]{|a_k|} = \lim_{n \rightarrow \infty} \sqrt[n^2]{n^n} = \lim_{n \rightarrow \infty} \sqrt[n]{n} = 1,$$

so  $R = 1$ . At  $x = \pm 1$ , the series diverges since the general term  $n^n x^{n^2}$  does not go to 0 when  $|x| = 1$ . So we conclude that the interval of convergence is  $(-1, 1)$ . □

**Theorem 13.2.** If  $(A) = \sum_{n=0}^{\infty} a_n x^n$  has radius of convergence  $R > 0$  (including  $R = \infty$ ), then for any  $0 < r < R$ , the power series  $(A)$  converges uniformly on  $[-r, r]$ . Moreover, if  $(A)$  converges at  $x = R < \infty$  (or  $x = -R$ ), then  $(A)$  converges uniformly on  $[0, R]$  (or  $[-R, 0]$ ).

*Proof.* For  $x \in [-r, r]$ , we have  $|a_n x^n| \leq |a_n| r^n$ , and  $\sum_{n=0}^{\infty} |a_n| r^n$  converges since  $r < R$ . So by the Weierstrass  $M$ -test, we get uniform convergence on  $[-r, r]$ . Second part left as an exercise.  $\square$

**Corollary 13.2.1.** For a power series  $(A) = \sum_{n=0}^{\infty} a_n x^n$ , we have the following:

1. if  $(A)$  has radius of convergence  $R > 0$ , then  $f(x) = \sum_{n=0}^{\infty} a_n x^n$  is continuous on  $(-R, R)$ ,
2.  $f(x)$  is differentiable on  $(-R, R)$ ,
3. and  $f(x) \in C^\infty(-R, R)$ , i.e. it is infinitely differentiable.

*Proof.* (1) We get

$$\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} \lim_{x \rightarrow x_0} a_n x^n = f(x_0)$$

since we have uniform convergence.

(2) One can verify that  $\sum_{n=0}^{\infty} a_n n x^{n-1}$  also has radius of convergence  $R$ . In particular, the derivative series also converges uniformly, so by Theorem 9.2, we can differentiate term by term.

(3) We can repeat (2) as many times as we want.  $\square$

**Theorem 13.3.** Suppose  $f(x) = \sum_{n=0}^{\infty} a_n x^n$  has radius of convergence  $R > 0$ . Then for any  $x \in (-R, R)$ , we have  $f \in \mathcal{R}([0, x])$  and

$$\int_0^x f(t) dt = \int_0^x \sum_{n=0}^{\infty} a_n t^n dt = \sum_{n=0}^{\infty} a_n \int_0^x t^n dt = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}.$$

*Proof.* As  $\sum_{n=0}^{\infty} a_n x^n$  converges uniformly in  $[-r, r]$ , by Theorem 9.1 we can integrate term by term.  $\square$

**Example 13.1.5.** Show that

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = -\ln(1-x)$$

for  $-1 < x < 1$ .

*Proof.* By the previous theorem, we can write

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = \int_0^x \sum_{n=1}^{\infty} t^{n-1} dt = \int_0^x \frac{1}{1-t} dt = -\ln(1-x),$$

as desired. Note that we used the geometric series in the second step.  $\square$

**Example 13.1.6.** Find the sum

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{3n+1}.$$

*Proof.* Let

$$S(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{3n+1} x^{3n+1}.$$

In particular,  $S(1)$  is the desired sum, so  $S$  converges at  $x = 1$  by the alternating series test. Then

$$\begin{aligned} S(1) &= \sum_{n=0}^{\infty} (-1)^n \int_0^1 x^{3n} dx = \int_0^1 \sum_{n=0}^{\infty} (-x^3)^n dt = \int_0^1 \frac{1}{1+x^3} dx \\ &= \frac{1}{3} \left[ \ln \frac{1+x}{\sqrt{1-x+x^2}} + \sqrt{3} \arctan \frac{2x-1}{\sqrt{3}} \right]_{x=0}^{x=1} = \frac{1}{3} \left( \ln 2 + \frac{\pi}{\sqrt{3}} \right) \end{aligned}$$

since we have uniform convergence on  $[0, 1]$ .

□

# Lecture 14

## Feb. 27 — Taylor Series

### 14.1 Taylor Series

Recall from Theorem 13.2.1 that  $f(x) = \sum_{n=0}^{\infty} a_n x^n$  is  $C^\infty$  on  $(-R, R)$ , where  $R$  is its radius of convergence. Here we can differentiate term by term, i.e. we have

$$\begin{aligned}f(x) &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 \cdots, \\f'(x) &= a_1 + 2a_2 x + 3a_3 x^2 + \cdots, \\f''(x) &= 2a_2 + 6a_3 x + \cdots.\end{aligned}$$

In particular, if we let  $x = 0$ , then

$$a_0 = f(0), \quad a_1 = f'(0), \quad a_2 = \frac{f''(0)}{2!}, \quad \dots, \quad a_n = \frac{f^{(n)}(0)}{n!}.$$

Here we call

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$$

the *Taylor series* of  $f$  at 0.

**Corollary 14.0.1.** *If  $f(x)$  has the power series expansion  $\sum_{n=0}^{\infty} c_n (x - a)^n$  on an open interval  $I$  containing  $a$ , then  $f \in C^\infty(I)$  and*

$$c_n = \frac{f^{(n)}(a)}{n!}.$$

*for all  $n \geq 0$ . In particular, if  $f$  has a power series expansion*

$$f = \sum_{n=0}^{\infty} c_n (x - a)^n,$$

*then  $c_n$  is unique.*

**Remark.** Since Taylor series are unique, we may use any way we want to find its coefficients.

**Example 14.0.1.** Suppose we would like to find the Taylor series of  $\ln(1 - x)$ . It is easier to observe

$$(\ln(1 - x))' = \frac{1}{1 - x}$$

and find the Taylor series of its derivative to be

$$\frac{1}{1 - x} = 1 + x + x^2 + \cdots + x^n + \dots$$

Then we can integrate term by term to find the Taylor series of  $\ln(1 - x)$ .

## 14.2 Convergence of Taylor Series

Given a function  $f \in C^\infty(I)$ , do we always have

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$$

in the interval  $I$ ?

**Example 14.0.2.** Consider

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Away from  $x = 0$ , we have

$$f' = \frac{2}{x^3} e^{-1/x^2}, \quad f'' = e^{-1/x^2} \left[ -\frac{6}{x^4} + \frac{4}{x^6} \right], \quad f''' = e^{-1/x^2} P_7(1/x), \quad \dots$$

So  $f$  is  $C^\infty$  away from  $x = 0$ . In particular,  $f^{(n)}(x) \rightarrow 0$  as  $x \rightarrow 0$  for all  $n \geq 0$  since

$$\lim_{x \rightarrow 0} \frac{1}{x^m} e^{-1/x^2} = 0$$

for all  $m \geq 0$ . Now also

$$f'(0) = \lim_{x \rightarrow 0} \frac{e^{-1/x^2}}{x} = 0,$$

so  $f'$  is continuous on  $\mathbb{R}$ . We can continue this to see that each  $f^{(n)}$  is continuous on  $\mathbb{R}$ . So  $f \in C^\infty(\mathbb{R})$  and  $f^{(n)}(0) = 0$  for all  $n \geq 0$ . Thus its Taylor series is identically zero, but  $f(x)$  is not the zero function.

**Remark.** Recall Lagrange's remainder term for the Taylor polynomial, which says

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

where

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

and  $\xi$  is between  $x$  and  $x_0$ . We can use this to justify the convergence of Taylor series.

**Theorem 14.1.** *Let  $R \in (0, \infty)$  and  $f \in C^\infty(x_0 - R, x_0 + R)$ . If there exists  $M > 0$  such that for all  $x \in (x_0 - R, x_0 + R)$ , we have*

$$|f^{(n)}(x)| \leq M^n$$

*for each  $n = 1, 2, \dots$ , then*

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

*for all  $x \in (x_0 - R, x_0 + R)$ .*

*Proof.* By Lagrange's remainder term formula, we have

$$|R_n(x)| = \frac{1}{(n+1)!} |f^{(n+1)}(\xi)| |x - x_0|^{n+1} \leq \frac{1}{(n+1)!} M^{n+1} R^{n+1} = \frac{(MR)^{n+1}}{(n+1)!} \rightarrow 0$$

as  $n \rightarrow \infty$  since  $M, R$  are fixed. Thus we get the desired equality, since the error term goes to zero.  $\square$

**Example 14.0.3.** For  $f(x) = e^x$ , we have  $f^{(n)}(x) = e^x$ . Then for  $x \in (-R, R)$ , we have  $|f^{(n)}(x)| \leq e^R$ . So by the previous theorem, we get

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$$

for all  $x \in \mathbb{R}$ , since we can take  $R$  as large as we would like.

## 14.3 Metric Spaces

**Definition 14.1.** We call a pair  $(X, \rho)$  a *metric space* if  $X$  is nonempty and  $\rho : X \times X \rightarrow \mathbb{R}^+$  satisfies:

1. positive-definiteness:  $\rho(x, y) \geq 0$  and  $\rho(x, y) = 0$  if and only if  $x = y$ ,
2. symmetry:  $\rho(x, y) = \rho(y, x)$ ,
3. and the triangle inequality:  $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$  for all  $x, y, z \in X$ .

We say  $\rho$  is a *distance function* if it satisfies the above properties.

**Example 14.1.1.** For  $X = \mathbb{R}^3$ , we may take  $\rho(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|$ .

**Example 14.1.2.** For  $X = C([a, b])$ , the set of all continuous functions on  $[a, b]$ , we can define

$$\rho(x, y) = \max_{t \in [a, b]} |x(t) - y(t)|$$

for any two  $x(t), y(t) \in C([a, b])$ . This is called the *maximum norm* or the  $\ell^\infty$  norm.

**Definition 14.2** (Convergence). We say that  $x_n \rightarrow x_0$  in  $(X, \rho)$  if  $\rho(x_n, x_0) \rightarrow 0$  as  $n \rightarrow \infty$ . We write

$$\lim_{n \rightarrow \infty} x_n = x_0.$$

**Example 14.2.1.** If  $x_n(t) \in C([a, b])$ , then  $x_n \rightarrow x_0$  means  $x_n(t) \rightarrow x_0(t)$  uniformly in  $[a, b]$ .<sup>1</sup>

**Definition 14.3.** We say that  $\{x_n\}$  is a *Cauchy sequence* in  $(X, \rho)$  if  $\rho(x_n, x_m) \rightarrow 0$  when  $n, m \rightarrow \infty$ , i.e. for any  $\epsilon > 0$ , there exists  $N$  such that if  $n, m \geq N$ , then  $\rho(x_n, x_m) < \epsilon$ .

**Definition 14.4.** If every Cauchy sequence in  $(X, \rho)$  has a limit  $x_n \rightarrow x^*$ , then we say that  $(X, \rho)$  is a *complete* metric space.

**Example 14.4.1.** The continuous functions  $C([a, b])$  with the maximum norm is complete.

**Definition 14.5.** Let  $(X, \rho), (Y, r)$  be two metric spaces. Then we say  $T : X \rightarrow Y$  is *continuous* if for any  $\{x_n\}$  and  $x_0 \in X$ , we have  $\rho(x_0, x_n) \rightarrow 0$  implies  $r(T(x_0), T(x_n)) \rightarrow 0$ .

<sup>1</sup>This is if we use the maximum norm from earlier.



**Theorem 14.2.** *A function  $T : (X, \rho) \rightarrow (Y, r)$  is continuous if and only if for all  $\epsilon > 0$  and  $x_0 \in X$ , there exists  $\delta = \delta(x_0, \epsilon) > 0$  such that  $\rho(x, x_0) < \delta$  implies  $r(T(x), T(x_0)) < \epsilon$ .*

*Proof.* Check this as an exercise □

## 14.4 Existence and Uniqueness Problem for ODEs

Consider the ordinary differential equation (ODE) problem

$$\begin{cases} \frac{dx}{dt} = F(t, x), \\ x(0) = \xi. \end{cases} \quad (1)$$

We would like to show that this ODE has a local solution.<sup>2</sup> To do this, we can transform the ODE into<sup>3</sup>

$$x(t) = \xi + \int_0^t f(\tau, x(\tau)) d\tau.$$

Now we would like to find a fixed point of this map. Let  $h > 0$  and consider  $X = C([-h, h])$  with the maximum norm. Define the mapping  $T : X \rightarrow X$  by

$$(Tx)(t) = \xi + \int_0^t f(\tau, x(\tau)) d\tau$$

for any  $x \in X$ . Then solving (1) is equivalent to finding a point  $x \in X$  such that  $x = Tx$ , i.e.  $x$  is a fixed point of  $T$ . We can do this via contraction mapping.

## 14.5 The Contraction Mapping Principle

**Definition 14.6.** We say  $T : (X, \rho) \rightarrow (X, \rho)$  is a *contraction mapping* if there exists  $a \in (0, 1)$  such that

$$\rho(Tx, Ty) \leq a\rho(x, y)$$

for any  $x, y \in X$ .

**Example 14.6.1.** Let  $X = [0, 1]$  and  $T(x)$  be differentiable on  $[0, 1]$  with  $T(x) \in [0, 1]$  and  $|T'(x)| \leq a < 1$ . Then  $T : X \rightarrow X$  is a contraction mapping since

$$\rho(Tx, Ty) = |T(x) - T(y)| = |T'(\xi)(x - y)| \leq a|x - y|$$

by the mean value theorem.

**Theorem 14.3** (Contraction mapping principle<sup>4</sup>). *Let  $(X, \rho)$  be a complete metric space and  $T : X \rightarrow X$  be a contraction mapping. Then  $T$  has a unique fixed point on  $X$ .*

<sup>2</sup>The solution may not exist for longer time periods, e.g. it might blow up at some point.

<sup>3</sup>This process is called *Picard iteration*.

<sup>4</sup>This is also known as the *Banach fixed-point theorem*.

*Proof.* For any  $x_0 \in X$ , define a sequence recursively by  $x_{n+1} = Tx_n$ , i.e.  $x_n = T^n x_0$ . Now we show that  $\{x_n\}$  is a Cauchy sequence. To do this, observe that

$$\rho(x_{n+1}, x_n) = \rho(Tx_n, Tx_{n-1}) \leq a\rho(x_n, x_{n-1}) \leq a^2\rho(x_{n-1}, x_{n-2}) \leq \cdots \leq a^n\rho(x_1, x_0)$$

for some  $0 < a < 1$  since  $\rho$  is a contraction mapping. So for any integer  $p > 0$ , by the triangle inequality we have

$$\rho(x_{n+p}, x_n) \leq \sum_{i=1}^p \rho(x_{n+i}, x_{n+i-1}) \leq \sum_{i=0}^{p-1} a^{n+i} \rho(x_0, x_1) \leq \sum_{i=0}^{\infty} a^{n+i} \rho(x_0, x_1) = \frac{a^n}{1-a} \rho(x_0, x_1).$$

Since  $\rho(x_0, x_1)$  is fixed and  $0 < a < 1$ , we see that  $\{x_n\}$  is a Cauchy sequence. Since  $X$  is complete, we have  $x_n \rightarrow x^* \in X$ . Now we show that  $x^*$  is a fixed point of  $T$ . This is because  $x_{n+1} = Tx_n$ , and letting  $n \rightarrow \infty$  gives  $x^* = Tx^*$ , since  $T$  is continuous.<sup>5</sup> Next, we show that  $x^*$  is the only fixed point. To do this, suppose otherwise that  $x^{**}$  is also a fixed point. Then

$$\rho(x^*, x^{**}) = \rho(Tx^*, Tx^{**}) \leq a\rho(x^*, x^{**}).$$

Since  $0 < a < 1$ , we must have  $\rho(x^*, x^{**}) = 0$ , i.e.  $x^* = x^{**}$ . □

---

<sup>5</sup>One can show that any contraction mapping is continuous.

# Lecture 15

## Feb. 29 — Contraction Mapping

### 15.1 Newton's Method

Suppose we want to find a solution of  $f(x) = 0$ , for some differentiable function  $f$ . We can use *Newton's method*, which starts with some initial guess  $x_0$ , and iteratively computes better guesses. More precisely, we look at  $x_i$ , find the tangent line to  $f$  at  $x_i$ , and then take  $x_{i+1}$  to be the  $x$ -intercept of this line. The tangent line is given by

$$y = f(x_n) + f'(x_n)(x - x_n).$$

Setting  $y = 0$  gives the recurrence

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

There are some nuances as to when Newton's method converges, since the  $x_i$  may end up oscillating or have some other problem if  $x_0$  is badly chosen (e.g. if it is far away from the actual zero  $\hat{x}$ ).

**Example 15.0.1.** Let  $f \in C^2([a, b])$  and  $\hat{x} \in (a, b)$  such that  $f(\hat{x}) = 0$  and  $f'(\hat{x}) \neq 0$ . Then there exists a neighborhood of  $\hat{x}$ , denoted by  $U(\hat{x})$ , such that for all  $x_0 \in U(\hat{x})$ , the sequence

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

converges to  $\hat{x}$  as  $n \rightarrow \infty$ .

*Proof.* Define the mapping

$$Tx = x - \frac{f(x)}{f'(x)}$$

in an interval containing  $\hat{x}$ , e.g.  $I_\delta = [\hat{x} - \delta, \hat{x} + \delta]$ . First we check that if  $\delta$  is small, then  $f : I_\delta \rightarrow I_\delta$ . For this, let  $x_1 = Tx$ . If  $|x - \hat{x}| < \delta$ , then we have

$$|x_1 - \hat{x}| = |Tx - T\hat{x}| \leq a|x - \hat{x}| < \delta,$$

for some  $0 < a < 1$ . This is because for all  $x_1, x_2 \in I_\delta$ , letting  $g(x) = Tx$ , we have

$$|Tx_1 - Tx_2| = |g(x_1) - g(x_2)| = |g'(\xi)||x_1 - x_2|$$

for some  $\xi \in I_\delta$  by the mean value theorem. Now notice that

$$|g'(\xi)| = \left| 1 - \frac{f'(\xi)}{f'(\xi)} + \frac{f(\xi)f''(\xi)}{f'(\xi)^2} \right| = \left| \frac{f(\xi)f''(\xi)}{f'(\xi)^2} \right| < a < 1$$

if  $\delta$  is small enough, since  $f(x)$  is small in  $I_\delta$ , and  $f''(x)$  is bounded and  $f'(x)$  is bounded away from zero. Then by the contraction mapping principle, we get the unique fixed point  $\hat{x}$ .  $\square$

**Remark.** The rough idea is that  $f'(x)$  is bounded away from zero in some open neighborhood  $U$  of  $\hat{x}$  since  $f'(\hat{x}) \neq 0$ , and  $f'$  is continuous since  $f \in C^2$ . Now pick some compact interval  $I \subseteq U$ , and we get that  $f'$  must be bounded on  $I$ . This gives that  $f$  is Lipschitz, i.e.  $|f(x) - f(y)| \leq L|x - y|$  for all  $x, y \in I$  and some constant  $L \in \mathbb{R}$ . Then we just need to show that  $0 < L < 1$ .

## 15.2 Existence and Uniqueness for ODEs

Recall the ODE

$$\begin{cases} \frac{dx}{dt} = F(t, x). \\ x(0) = \xi. \end{cases}$$

We rewrite this into the integral form

$$x(t) = \xi + \int_0^t F(\tau, x(\tau)) d\tau. \quad (*)$$

Now let  $h > 0$  be some fixed constant that we choose later. Define  $X = C([-h, h])$  and  $T : X \rightarrow X$  by

$$(Tx)(t) = \xi + \int_0^t F(\tau, x(\tau)) d\tau.$$

Then  $(*)$  is equivalent to finding a fixed point of  $T$ . Assume  $F(t, x)$  satisfies a *local Lipschitz condition*, i.e. there exist  $\delta > 0$  and  $L > 0$  such that when  $|t| \leq h$ ,  $|x_1 - \xi| \leq \delta$ , and  $|x_2 - \xi| \leq \delta$ , we have

$$|F(t, x_1) - F(t, x_2)| \leq L|x_1 - x_2|.$$

Define

$$B(\xi, \delta) := \{x(t) \in C([-h, h]) \mid \max_{|t| \leq h} |x(t) - \xi| \leq \delta\}.$$

Think of this as the ball of radius  $\delta$  around the constant function  $\xi$  in  $C([-h, h])$ . We want to check

- (i) if  $h, \delta$  are small enough, then  $T : B(\xi, \delta) \rightarrow B(\xi, \delta)$ ,
- (ii) and  $T$  is a contraction.

Once we do, then  $T$  has a unique fixed point  $\hat{x}(t)$  by the contraction mapping principle, which is the solution of the ODE for  $t \in [-h, h]$ . For the first property, let

$$M = \max\{|F(t, x)| \mid t \in [-h, h] \times [\xi - \delta, \xi + \delta]\},$$

which exists since  $F$  is continuous and  $[-h, h] \times [\xi - \delta, \xi + \delta]$  is compact. If  $h$  is small, then

$$\max_{t \in [-h, h]} |(Tx)(t) - \xi| = \max_{t \in [-h, h]} \left| \int_0^t F(\tau, x(\tau)) d\tau \right| \leq Mh \leq \delta$$

since  $t \leq h$ . So  $T : B(\xi, \delta) \rightarrow B(\xi, \delta)$ . Now we show that  $T$  is a contraction. For any  $x(t), y(t) \in B(\xi, \delta)$ ,

$$\begin{aligned} \rho(Tx, Ty) &= \max_{t \in [-h, h]} \left| \int_0^t F(\tau, x(\tau)) - F(\tau, y(\tau)) d\tau \right| \leq h \max_{|t| \leq h} |F(t, x(t)) - F(t, y(t))| \\ &= Lh \max_{|t| \leq h} |x(t) - y(t)| = Lh\rho(x, y) \end{aligned}$$

by the Lipschitz condition. Now we can choose  $h$  such that  $Lh < 1$ , which ensures that  $T$  is a contraction.

**Remark.** This is called the *Picard-Lindelöf theorem*, or the *existence and uniqueness theorem for ODEs*.

## 15.3 Implicit Function Theorem

**Example 15.0.2** (Implicit function theorem). Let  $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  and  $U \times V \subseteq \mathbb{R} \times \mathbb{R}$  a neighborhood of  $(x_0, y_0)$ . Assume  $f$  and  $\partial f / \partial y$  are continuous on  $U \times V$ , and  $f(x_0, y_0) = 0$  and

$$\frac{\partial f}{\partial y}(x_0, y_0) \neq 0.$$

Then there exists a neighborhood  $U_0 \times V_0 \subseteq U \times V$  and a unique continuous function  $\varphi : U_0 \rightarrow V_0$  satisfying

$$f(x, \varphi(x)) = 0, \quad \varphi(x_0) = y_0.$$

*Proof.* We want to solve  $f(x, y) = 0$  in a neighborhood of  $(x_0, y_0)$ . Define the mapping  $\varphi \mapsto T\varphi$ , where  $\varphi \in C([x_0 - r, x_0 + r])$  for some  $r > 0$ . We take the definition

$$(T\varphi)(x) = \varphi(x) - \left( \frac{\partial f}{\partial y}(x_0, y_0) \right)^{-1} f(x, \varphi(x)).$$

Observe that if  $\varphi$  is a fixed point of  $T$ , then  $f(x, \varphi(x)) = 0$ . Let  $X = C([x_0 - r, x_0 + r])$ , and we will choose  $r$  later. For any  $\varphi, \psi \in X$ , define the distance

$$\rho(\varphi, \psi) = \max_{x \in [x_0 - r, x_0 + r]} |\varphi(x) - \psi(x)|.$$

Now we see that

$$\begin{aligned} \rho(T\varphi, T\psi) &= \max_{x \in [x_0 - r, x_0 + r]} \left| \varphi(x) - \psi(x) - \left( \frac{\partial f}{\partial y}(x_0, y_0) \right)^{-1} (f(x, \varphi(x)) - f(x, \psi(x))) \right| \\ &= \max_{x \in [x_0 - r, x_0 + r]} \left| \left( 1 - \frac{\partial f}{\partial y}(x_0, y_0)^{-1} \frac{\partial f}{\partial y}(x, \hat{y}) \right) (\varphi(x) - \psi(x)) \right| \end{aligned}$$

for some  $\hat{y}$  between  $\varphi(x)$  and  $\psi(x)$  by the mean value theorem. Now define the ball

$$B(y_0, \delta) = \{\varphi \in C(x_0 - r, x_0 + r) \mid \rho(\varphi, y_0) \leq \delta\}.$$

If  $\varphi, \psi \in B(y_0, \delta)$  and  $r, \delta$  are small enough, then

$$\left| 1 - \frac{\partial f}{\partial y}(x_0, y_0)^{-1} \frac{\partial f}{\partial y}(x, \hat{y}) \right| < \frac{1}{2},$$

which ensures a contraction. So it only remains to check that  $T : B(y_0, \delta) \rightarrow B(y_0, \delta)$ . For  $\varphi \in B(y_0, \delta)$ ,

$$\begin{aligned} \rho(T\varphi, y_0) &\leq \rho(T\varphi, T y_0) + \rho(T y_0, y_0) \\ &\leq \frac{1}{2} \rho(\varphi, y_0) + \max_{x \in [x_0 - r, x_0 + r]} \left| \frac{\partial f}{\partial y}(x_0, y_0)^{-1} f(x, y_0) \right| \leq \frac{1}{2} \delta + \frac{1}{2} \delta = \delta \end{aligned}$$

if  $r$  is small enough since  $f(x_0, y_0) = 0$ . So  $T\varphi \in B(y_0, \delta)$ . Thus  $T$  is a contraction mapping on  $B(y_0, \delta)$ , so it has a unique fixed point by the contraction mapping principle.  $\square$

**Remark.** The idea is that if  $(\partial f / \partial y)(x_0, y_0) \neq 0$ , then the solution set to  $f(x, y) = 0$  in a neighborhood of  $(x_0, y_0)$  is given by the curve  $\varphi$ .

# Lecture 16

## Mar. 5 — The Derivative in $\mathbb{R}^n$

### 16.1 Partial Derivatives

**Definition 16.1.** Let  $f$  be a real-valued function defined on an open set  $U \subseteq \mathbb{R}^n$ . For a fixed point  $a = (a_1, \dots, a_n) \in U$ , the *partial derivative* of  $f$  at  $a$  with respect to  $x_i$  is

$$\frac{\partial f}{\partial x_i}(a) = \lim_{x_i \rightarrow a_i} \frac{f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n) - f(a_1, \dots, a_n)}{x_i - a_i} = \lim_{h \rightarrow 0} \frac{f(a + h\vec{e}_i) - f(a)}{h},$$

when this limit exists. Here  $\vec{e}_i$  is the  $i$ th standard basis vector in  $\mathbb{R}^n$ .

**Remark.** The following are equivalent notations for partial derivatives:

$$\frac{\partial f}{\partial x_i}(a), \quad f'_i(a), \quad f_{x_i}, \quad D_i f.$$

**Definition 16.2.** The *gradient* of  $f$  at  $a$  is a vector of the partial derivatives, i.e.

$$\nabla f(a) = \left( \frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right).$$

**Remark.** In one dimension, if  $f(x)$  is differentiable at  $x_0$ , then  $f(x)$  is continuous at  $x_0$ , because

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0),$$

which implies

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + o(x - x_0).$$

As  $x \rightarrow x_0$ , the RHS goes to zero, so  $f(x) \rightarrow f(x_0)$ . However, this need not hold in higher dimensions. For  $n \geq 2$ , even if  $\partial f / \partial x_i(a)$  exists for all  $i = 1, 2, \dots, n$ ,  $f(x_0)$  might not be continuous at  $x = a$ .

**Example 16.2.1.** Consider the function

$$f(x, y) = \begin{cases} xy/(x^2 + y^2) & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

At  $(0, 0)$ , we have

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{x \rightarrow 0} \frac{f(x, 0) - 0}{x} = \lim_{x \rightarrow 0} \frac{0 - 0}{x} = 0.$$

Similarly,  $\partial f / \partial y(0, 0) = 0$ . But  $f(x, y)$  is not continuous at  $(0, 0)$ . For continuity, we need

$$\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0$$

when  $(x, y) \rightarrow (0, 0)$  along any path  $\Gamma$  from  $(x, y)$  to  $(0, 0)$ . So it suffices to find two paths  $\Gamma_1$  and  $\Gamma_2$  with different limits to show that the limit does not exist. So choose  $\Gamma : y = mx$ . Then

$$\lim_{\substack{(x,y) \rightarrow (0,0) \\ \text{along } \Gamma}} \frac{x(mx)}{x^2 + m^2 x^2} = \frac{m}{1 + m^2}.$$

This limit clearly depends on  $m$ , so we can simply choose two different values of  $m$  to get differing limits. Hence we see that  $f(x, y)$  is not continuous at  $(0, 0)$ .

## 16.2 Differentiability

**Definition 16.3.** Let  $f : U \rightarrow \mathbb{R}$  where  $U \subseteq \mathbb{R}^n$  and  $a = (a_1, \dots, a_n) \in U$ . Then  $f$  is *differentiable* at  $a$  if there exist constants  $c_1, \dots, c_n \in \mathbb{R}$  such that

$$\lim_{h \rightarrow 0} \frac{f(a+h) - (f(a) + c_1 h_1 + \dots + c_n h_n)}{\|h\|} = 0.$$

Here the distance is

$$d(x, a) = \sqrt{(x_1 - a_1)^2 + \dots + (x_n - a_n)^2}.$$

**Remark.** If we let  $\vec{c} = (c_1, \dots, c_n)$ , then the definition of differentiability becomes

$$f(x) = f(a) + \vec{c} \cdot (x - a) + o(d(x, a)).$$

We will soon show that in fact  $\vec{c} = \nabla f(a)$ .

**Remark.** In the one-dimensional case, for  $f(x)$  where  $x \in I \subseteq \mathbb{R}$ , we defined the derivative as

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

This is equivalent to

$$\lim_{x \rightarrow a} \frac{f(x) - f(a) - f'(a)(x - a)}{|x - a|} = 0,$$

which is the same as our new definition for differentiability.

**Proposition 16.1.** *If  $f(x)$  is differentiable at  $a$ , then  $\partial f / \partial x_i(a)$  exists and  $c_i = \partial f / \partial x_i(a)$ .*

*Proof.* We can compute that

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \rightarrow 0} \frac{f(a + h\vec{e}_i) - f(a)}{h} = \lim_{h \rightarrow 0} \frac{f(a) + \vec{c} \cdot h\vec{e}_i + o(|h|) - f(a)}{h} = \lim_{h \rightarrow 0} \frac{\vec{c} \cdot h\vec{e}_i + o(|h|)}{h} = c_i,$$

which is the desired result. □

**Proposition 16.2.** *If  $f(x)$  is differentiable at  $a$ , then  $f(x)$  is continuous at  $a$ .*

*Proof.* From the definition, we have

$$f(x) = f(a) + \vec{c} \cdot (x - a) + o(d(x, a)).$$

As we take  $x \rightarrow a$ , we get  $f(x) \rightarrow f(a)$  just as before.  $\square$

**Remark.** If  $f(x)$  is differentiable, then  $f(x)$  has all partial derivatives  $\partial f / \partial x_i$  at  $x = a$ . But the converse is not true in general. So differentiability is a strictly stronger condition in  $\mathbb{R}^n$ .

**Lemma 16.1.** *Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , where  $U$  is an open set. Then  $f$  is differentiable at  $x = a$  if and only if there exist functions  $A_1, \dots, A_n$  on  $U$ , continuous at  $x = a$ , such that*

$$f(x) - f(a) = A_1(x)(x_1 - a_1) + A_2(x)(x_2 - a_2) + \dots + A_n(x)(x_n - a_n)$$

for all  $x \in U$ . In this case,  $\partial f / \partial x_i(a) = A_i(a)$ .

*Proof.* ( $\Rightarrow$ ) Assume  $f(x)$  is differentiable at  $x = a$ . Then

$$\lim_{x \rightarrow a} \frac{f(x) - (f(a) + f'_1(a)(x_1 - a_1) + \dots + f'_n(a)(x_n - a_n))}{d(x, a)} = 0.$$

Define the function  $e : U \rightarrow \mathbb{R}$  via

$$e(x) = \frac{f(x) - (f(a) + f'_1(a)(x_1 - a_1) + \dots + f'_n(a)(x_n - a_n))}{|x_1 - a_1| + \dots + |x_n - a_n|}.$$

Observe that the denominator is a distance equivalent to  $d(x, a)$ .<sup>1</sup> Then  $\lim_{x \rightarrow a} e(x) = 0$ , since the distances are equivalent. Then we get

$$\begin{aligned} f(x) &= f(a) + f'_1(a)(x_1 - a_1) + \dots + f'_n(a)(x_n - a_n) + e(x)(|x_1 - a_1| + \dots + |x_n - a_n|) \\ &= f(a) + A_1(x)(x_1 - a_1) + \dots + A_n(x)(x_n - a_n), \end{aligned}$$

where we define

$$A_i(x) = \begin{cases} f'_i(a) + e(x) & \text{if } x_i - a_i \geq 0 \\ f'_i(a) - e(x) & \text{if } x_i - a_i < 0. \end{cases}$$

Each  $A_i$  is continuous and satisfies  $A_i(a) = f'_i(a)$ , as desired.

( $\Leftarrow$ ) Suppose  $A_1, \dots, A_n$  exist such that

$$f(x) - f(a) = A_1(x)(x_1 - a_1) + \dots + A_n(x)(x_n - a_n).$$

We check that  $f$  is differentiable at  $x = a$ . Choose  $c_i = A_i(a)$ . Then

$$\left| \frac{f(x) - (f(a) + \sum_{i=1}^n c_i(x_i - a_i))}{d(x, a)} \right| = \left| \frac{\sum_{i=1}^n (A_i(x) - A_i(a))(x_i - a_i)}{d(x, a)} \right| \leq \sum_{i=1}^n |A_i(x) - A_i(a)| \rightarrow 0$$

as  $x \rightarrow a$ , since each  $A_i$  is continuous at  $a$ . Thus  $f$  is differentiable at  $x = a$ .  $\square$

<sup>1</sup>Two metrics (distances)  $d_1, d_2 : X \times X \rightarrow \mathbb{R}$  on a set  $X$  are *equivalent* if there exist constants  $c, C \in \mathbb{R}$  such that  $cd_1(x, y) \leq d_2(x, y) \leq Cd_1(x, y)$  for all  $x, y \in X$ .



**Theorem 16.1.** Let  $U$  be an open set in  $\mathbb{R}^n$  and suppose that  $f : U \rightarrow \mathbb{R}$  has partial derivatives  $f'_1, \dots, f'_n$  on  $U$  which are continuous at  $x = a$ . Then  $f$  is differentiable at  $x = a$ .

*Proof.* We change  $x_i$  to  $a_i$  one by one to get

$$\begin{aligned} f(x) - f(a) &= (f(x_1, \dots, x_n) - f(a_1, x_2, \dots, x_n)) \\ &\quad + (f(a_1, x_2, \dots, x_n) - f(a_1, a_2, x_3, \dots, x_n)) \\ &\quad + (f(a_1, a_2, x_3, \dots, x_n) - f(a_1, a_2, a_3, x_4, \dots, x_n)) \\ &\quad \vdots \\ &\quad + (f(a_1, \dots, a_{n-1}, x_n) - f(a_1, \dots, a_n)). \end{aligned}$$

Each of these terms differ in only one variable, so we can apply the mean value theorem to get

$$\begin{aligned} f(x) - f(a) &= f'_1(\xi_1, a_2, \dots, a_n)(x_1 - a_1) \\ &\quad + f'_2(a_1, \xi_2, a_3, \dots, a_n)(x_2 - a_2) \\ &\quad \vdots \\ &\quad + f'_n(a_1, \dots, a_{n-1}, \xi_n)(x_n - a_n). \end{aligned}$$

So we can set

$$\begin{aligned} A_1(x) &= f'_1(\xi_1, a_2, \dots, a_n)(x_n - a_n) \\ A_2(x) &= f'_2(a_1, \xi_2, a_3, \dots, a_n)(x_n - a_n) \\ &\quad \vdots \\ A_n(x) &= f'_n(a_1, \dots, a_{n-1}, \xi_n)(x_n - a_n). \end{aligned}$$

where  $\xi_i$  is between  $a_i$  and  $x_i$  for each  $1 \leq i \leq n$ . Each  $A_i$  is continuous since each  $f'_i$  is continuous, so we get  $A_i(x) \rightarrow A_i(a)$  when  $x \rightarrow a$ . So  $f(x)$  is differentiable at  $x = a$ .  $\square$

**Remark.** In functional analysis, there is an analogous theorem that relates the Gâteaux derivative (similar to a partial derivative) and the Fréchet derivative (similar to differentiability).

**Remark.** The existence of continuous partial derivatives at  $x = a$  is a sufficient condition for differentiability at  $x = a$ , but it is not necessary.

**Example 16.3.1.** Consider the function

$$f(x, y) = \begin{cases} (x^2 + y^2) \sin(1/(x^2 + y^2)) & \text{if } x^2 + y^2 \neq 0, \\ 0 & \text{if } x^2 + y^2 = 0. \end{cases}$$

First we verify that  $f$  is differentiable at  $(x, y) = (0, 0)$ . We can compute

$$|f(x, y) - f(0, 0)| = (x^2 + y^2) \left| \sin \frac{1}{x^2 + y^2} \right| = O((\sqrt{x^2 + y^2})^2),$$

so  $f$  is differentiable at  $(0, 0)$ , with the zero linear approximation  $f(x, y) \approx 0x + 0y$ . This also gives  $f'_1(0, 0) = f'_2(0, 0) = 0$ . However, if  $(x, y) \neq (0, 0)$ , then

$$f_x(x, y) = 2x \sin \frac{1}{x^2 + y^2} - \frac{2x}{x^2 + y^2} \cos \frac{1}{x^2 + y^2}$$

by the product rule. Letting  $(x, y) \rightarrow 0$ , we see that the second term on the RHS has no limit. For example, along  $y = 0$ , we get

$$\left. \frac{2x}{x^2 + y^2} \cos \frac{1}{x^2 + y^2} \right|_{y=0} = \frac{2}{x} \cos \frac{1}{x^2},$$

which has no limit as  $x \rightarrow 0$ . So  $f_x(x, y)$  is not continuous at  $(0, 0)$ . By symmetry,  $f_y(x, y)$  is also not continuous at  $(0, 0)$ . So a function  $f$  can be differentiable at  $x = a$  while its partial derivatives  $f'_i(x)$  are not continuous at  $x = a$ .

# Lecture 17

## Mar. 7 — Vector-Valued Functions

### 17.1 The Derivative for Vector-Valued Functions

**Definition 17.1.** Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ , where  $U$  is an open set. We say  $f = (f_1, \dots, f_m)$  is *differentiable* if each  $f_i$  is differentiable. In other words, for  $x, a \in U$ ,

$$f(x) - f(a) = \begin{pmatrix} \partial f_1 / \partial x_1(a) & \dots & \partial f_1 / \partial x_n(a) \\ \vdots & \ddots & \vdots \\ \partial f_m / \partial x_1(a) & \dots & \partial f_m / \partial x_n(a) \end{pmatrix} (x - a) + o(|x - a|).$$

We may call the matrix of partial derivatives  $f'(a)$  or  $\partial f / \partial x$ .

**Theorem 17.1** (Chain rule). *Let  $U, V$  be open sets in  $\mathbb{R}^n$  and  $\mathbb{R}^m$ , respectively, and  $f : U \rightarrow V$ ,  $g : V \rightarrow \mathbb{R}$  be functions. Let  $a \in U$  such that  $f$  is differentiable at  $a$  and  $g$  is differentiable at  $f(a)$ . Then  $g \circ f(x) = g(f(x))$  is differentiable at  $a$ , and*

$$(g \circ f)'_j(a) = \sum_{i=1}^m g'_i(f(a)) (f_i)'_j(a),$$

where  $(f_i)'_j$  denotes  $\partial f_i / \partial x_j$ .

*Proof.* See textbook (Rosenlicht). □

**Remark.** In matrix notation, we can write this via

$$\nabla(g \circ f) = \nabla g \frac{\partial f}{\partial x},$$

where  $\nabla g$  is a vector in  $\mathbb{R}^m$  and  $\partial f / \partial x$  is an  $m \times n$  matrix.

**Remark.** When  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ , we get

$$\frac{\partial f}{\partial x}(a) = \begin{pmatrix} \partial f_1 / \partial x_1(a) & \dots & \partial f_1 / \partial x_n(a) \\ \vdots & \ddots & \vdots \\ \partial f_n / \partial x_1(a) & \dots & \partial f_n / \partial x_n(a) \end{pmatrix},$$

an  $n \times n$  matrix. So we can take its determinant, and we call  $\det(\partial f / \partial x)$  the *Jacobian* of  $f$ .

## 17.2 Implicit Function Theorem in $\mathbb{R}^n$

**Remark.** Recall Theorem 15.0.2, the implicit function theorem on  $\mathbb{R}$ . Note that all solutions to  $f(x, y) = 0$  are in fact given by the continuous curve  $y = \varphi(x)$  in a neighborhood of  $(x_0, y_0)$ . To see this, suppose  $f(x, y) - f(x, \varphi(x)) = 0$  for some  $y$ . Fixing  $x$ , by the mean value theorem we get

$$\frac{\partial f}{\partial y}(x, \xi)(y - \varphi(x)) = 0.$$

Since  $\partial f / \partial y(x, \xi) \neq 0$ , this implies  $y = \varphi(x)$ . So there are no solutions off of the curve.

**Theorem 17.2.** Let  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$  and  $U \times V \subseteq \mathbb{R}^n \times \mathbb{R}^m$  be a neighborhood of  $(x_0, y_0)$ . Suppose  $f$  and  $\partial f / \partial y$  are continuous on  $U \times V$ , and  $f(x_0, y_0) = 0$  and

$$\det \left( \frac{\partial f}{\partial y}(x_0, y_0) \right) \neq 0.$$

Then there exists a neighborhood  $U_0 \times V_0 \subseteq U \times V$  of  $(x_0, y_0)$  and a unique continuous function  $\varphi : U_0 \rightarrow V_0$  satisfying

$$\begin{cases} f(x, \varphi(x)) = 0, \\ \varphi(x_0) = y_0. \end{cases}$$

*Proof.* Take  $T : \varphi \mapsto T\varphi$ , where  $\varphi \in C(\overline{B(x_0, r)}, \mathbb{R}^m)$  and<sup>1</sup>

$$\rho(\varphi, \psi) = \max_{\substack{x \in \overline{B(x_0, r)} \\ 1 \leq i \leq m}} |\varphi_i(x) - \psi_i(x)|$$

is the metric. Define

$$(T\varphi)(x) = \varphi(x) - \left( \frac{\partial f}{\partial y}(x_0, y_0) \right)^{-1} f(x, \varphi(x)).$$

Note that this is a matrix inverse. Define

$$X = \{\varphi \in C(\overline{B(x_0, r)}, \mathbb{R}^m) : \rho(\varphi, y_0) \leq \delta\}$$

Now it suffices to show that

1.  $\rho(T\varphi, T\psi) \leq \frac{1}{2}\rho(\varphi, \psi)$  for  $r, \delta$  small enough,
2. and  $T : X \rightarrow X$

in order to apply the contraction mapping theorem and finish.<sup>2</sup> Check these two details as an exercise.  $\square$

**Corollary 17.2.1** (Inverse function theorem). If  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that  $f, \partial f / \partial y$  are continuous in a neighborhood of  $y_0 \in U$  and  $\det(\partial f / \partial y(y_0)) \neq 0$ , then there exists  $g : V \rightarrow \mathbb{R}^n$ , where  $V$  is a neighborhood of  $f(y_0)$ , such that  $g \circ f = \text{id}$ , i.e.  $g(f(x)) = x$ .

*Proof.* Let  $F(x, y) = x - f(y)$ , and we want to solve  $F(x, y) = 0$  near the point  $(f(y_0), y_0)$ . Apply the implicit function theorem to  $F$  to finish. See textbook (Rosenlicht) for details.  $\square$

<sup>1</sup>Recall that  $\overline{A}$  is the *closure* of a set  $A$ , i.e. the smallest closed set containing  $A$ .

<sup>2</sup>The continuity of  $\varphi$  is implied since it is a fixed point in a function space of continuous functions.

## 17.3 Higher Order Derivatives

**Definition 17.2.** If  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $\partial f / \partial x_i$  is differentiable with respect to  $x_j$ , then we can define<sup>3</sup>

$$\frac{\partial}{\partial x_j} \left( \frac{\partial f}{\partial x_i} \right) = \frac{\partial^2 f}{\partial x_j \partial x_i} = f_{x_j x_i}.$$

These are the *second order partial derivatives*. Similarly, we can define higher order partial derivatives.

**Remark.** We want to know: When is  $f_{x_j x_i} = f_{x_i x_j}$  (i.e.  $(f'_i)'_j = (f'_j)'_i$ )?

**Theorem 17.3.** Let  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $a \in U$ . If  $(f'_i)'_j$  and  $(f'_j)'_i$  exist and are continuous in a neighborhood of  $a$ , then

$$(f'_i)'_j(a) = (f'_j)'_i(a).$$

*Proof.* Take a point  $x = (x_1, x_2) \neq a$  in a small neighborhood of  $a = (a_1, a_2)$ . Define

$$\Delta(x) = \frac{f(x_1, x_2) - f(x_1, a_2) - f(a_1, x_2) + f(a_1, a_2)}{(x_1 - a_1)(x_2 - a_2)}.$$

Let  $\varphi(x_1, x_2) = f(x_1, x_2) - f(x_1, a_2)$ , so that

$$\Delta x = \frac{\varphi(x_1, x_2) - \varphi(a_1, x_2)}{(x_1 - a_1)(x_2 - a_2)}.$$

by the mean value theorem,

$$\varphi(x_1, x_2) - \varphi(a_1, x_2) = (x - a_1)\varphi'_1(\xi_1, x_2)$$

for some  $\xi_1$  between  $x_1$  and  $a_1$ . Then

$$\Delta x = \frac{\varphi'_1(\xi_1, x_2)}{x_2 - a_2} = \frac{f'_1(\xi_1, x_2) - f'_1(\xi_1, a_2)}{x_2 - a_2} = (f'_1)'_2(\xi_1, \xi_2)$$

by the mean value theorem again, for some  $\xi_2$  between  $a_2$  and  $x_2$ . In the same manner, we get

$$\Delta x = (f'_2)'_1(\xi'_1, \xi'_2),$$

where  $(\xi'_1, \xi'_2)$  is in a neighborhood of  $a$ . Now let  $(x_1, x_2) \rightarrow (a_1, a_2)$ , then  $(\xi_1, \xi_1), (\xi'_1, \xi'_2) \rightarrow (a_1, a_2)$ . By assumption,  $(f'_2)'_1(\xi'_1, \xi'_2) \rightarrow (f'_2)'_1(a)$  and  $(f'_1)'_2(\xi_1, \xi_2) \rightarrow (f'_1)'_2(a)$  since  $(f'_2)'_1$  and  $(f'_1)'_2$  are continuous, so

$$(f'_1)'_2(a) = (f'_2)'_1(a)$$

since  $(f'_2)'_1(\xi'_1, \xi'_2) = (f'_1)'_2(\xi_1, \xi_2) = \Delta x$ . □

---

<sup>3</sup>Note the indices, we have  $f_{x_j x_i} = (f'_i)'_j$ .

# Lecture 18

## Mar. 12 — More on the Implicit Function Theorem

### 18.1 Example of Unequal Mixed Partial Derivatives

**Example 18.0.1.** Consider the function.

$$f(x, y) = \begin{cases} xy(x^2 - y^2)/(x^2 + y^2) & \text{if } x^2 + y^2 \neq 0 \\ 0 & \text{if } x^2 + y^2 = 0. \end{cases}$$

We find

$$f'_x(x, y) = \begin{cases} y((x^2 - y^2)(x^2 + y^2) + 4x^2y^2)/(x^2 + y^2)^2 & \text{if } x^2 + y^2 \neq 0 \\ 0 & \text{if } x^2 + y^2 = 0. \end{cases}$$

and

$$f'_y(x, y) = \begin{cases} x((x^2 - y^2)(x^2 + y^2) + 4x^2y^2)/(x^2 + y^2)^2 & \text{if } x^2 + y^2 \neq 0 \\ 0 & \text{if } x^2 + y^2 = 0. \end{cases}$$

Now  $f'_x(0, y) = -y$  and so  $f''_{xy}(0, 0) = -1$ , but  $f'_y(x, 0) = x$  and so  $f''_{yx}(0, 0) = 1$ . Thus  $f''_{xy}(0, 0) \neq f''_{yx}(0, 0)$ .

### 18.2 Revisiting the Implicit Function Theorem

**Theorem 18.1** (Mean value theorem in  $\mathbb{R}^n$ ). *Suppose  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  where  $U$  is open, and that  $f$  is differentiable. Then for any  $a, b \in U$ , there exists  $\xi$  on the line segment connecting  $a, b$  such that*

$$f(a) - f(b) = \nabla f(\xi) \cdot (a - b).$$

*Proof.* Define  $h(t) = f(ta + (1 - t)b)$  for  $t \in [0, 1]$ . Then

$$f(a) - f(b) = h(1) - h(0) = h'(c) = \nabla f(ca + (1 - c)b) \cdot (a - b)$$

for some  $c \in (0, 1)$  by the usual mean value theorem. So we can pick  $\xi = ca + (1 - c)b$ . □

The following is a corollary of Theorem 17.2:

**Corollary 18.1.1.** *If  $f_1, \dots, f_n$  are continuously differentiable in a neighborhood of  $(a, b)$ , then  $U, V$  can be chosen such that  $\varphi : U \rightarrow V$  is continuously differentiable, i.e.  $\varphi \in C^1(U, V)$ .*<sup>1</sup>

---

<sup>1</sup>In the notation of Theorem 17.2, we had  $f = (f_1, \dots, f_n)$ ,  $(x_0, y_0)$  instead of  $(a, b)$ , and  $U_0, V_0$  instead of  $U, V$ .

*Proof.* Recall that

$$\begin{cases} f_i(x, \varphi(x)) = 0 \\ \varphi(a) = b \end{cases}$$

for all  $1 \leq i \leq n$ . Then we have

$$\begin{aligned} 0 = f_i(x, \varphi(x)) - f_i(a, \varphi(a)) &= \frac{\partial f_i}{\partial x_1}(z^i)(x_1 - a_1) + \cdots + \frac{\partial f_i}{\partial x_m}(z^i)(x_m - a_m) \\ &+ \frac{\partial f_i}{\partial y_1}(z^i)(\varphi_1(x) - b_1) + \cdots + \frac{\partial f_i}{\partial y_n}(z^i)(\varphi_n(x) - b_n) \end{aligned}$$

where  $z_i$  is between  $(x, \varphi(x))$  and  $(a, b)$ . Now since  $\det(\partial f / \partial y_j(z^i)) \neq 0$ , we can solve this linear system to get

$$\varphi_i(x) - b_i = A_{i1}(x)(x_1 - a_1) + A_{i2}(x)(x_2 - a_2) + \cdots + A_{im}(x)(x_m - a_m),$$

where the coefficient functions  $A_{i1}, \dots, A_{im}$  are continuous in a neighborhood of  $a$  by assumption. So  $\varphi(x) = (\varphi_1(x), \dots, \varphi_n(x))$  is continuously differentiable in a neighborhood of  $a$ , as desired.  $\square$

**Remark.** We can improve this if we know  $f$  is more smooth. In general, if  $f \in C^k$ , then  $\varphi \in C^k$  also.

**Example 18.0.2** (Implicit differentiation). Suppose  $F(x, y) = 0$  and  $y = \varphi(x)$ . If  $F \in C^1$  and  $F'_y \neq 0$ ,

$$\frac{d}{dx} F(x, \varphi(x)) = 0 \implies F'_x + F'_y \varphi_x = 0 \implies \varphi_x = -\frac{F'_x}{F'_y}.$$

We can justify this rigorously via the implicit function theorem.

## 18.3 Lagrange Multipliers

Recall from multivariable calculus that we can use the method of Lagrange multipliers to solve the minimum (or maximum) of a function  $f(x, y)$  subject to a constraint  $g(x, y) = 0$ . The *Lagrange multiplier equation* is

$$\begin{cases} \nabla f = \lambda \nabla g \\ g = 0, \end{cases}$$

which is a system of three equations in  $(x, y, \lambda)$ . Here  $\lambda$  is called the *Lagrange multiplier*.

**Proposition 18.1.** Assume that at a local minimum (or maximum) point  $(x_0, y_0)$  of  $f$ , we have

$$\nabla g(x_0, y_0) \neq \vec{0},$$

i.e.  $g_x^2(x_0, y_0) + g_y^2(x_0, y_0) \neq 0$ . Then there exists  $\lambda$  such that  $\nabla f = \lambda \nabla g$  at  $(x_0, y_0)$ .

*Proof.* Assume  $g'_y(x_0, y_0) \neq 0$ . Then by the implicit function theorem, there exists some  $y = \varphi(x)$  in a neighborhood of  $y_0$  such that  $g(x, \varphi(x)) = 0$ . Then  $h(x) = f(x, \varphi(x))$  obtains a minimum (or maximum) at  $x_0$ . Then  $h'(x_0) = 0$ , so

$$f'_x(x_0, y_0) + f'_y(x_0, y_0)\varphi'_x(x_0) = 0.$$

Also, by implicit differentiation,

$$\varphi'_x(x_0) = -\frac{g'_x(x_0, y_0)}{g'_y(x_0, y_0)}.$$

So we can set

$$\lambda = -\frac{f'_y(x_0, y_0)}{g'_y(x_0, y_0)} \implies \begin{cases} f'_x(x_0, y_0) + \lambda g'_x(x_0, y_0) = 0 \\ f'_y(x_0, y_0) + \lambda g'_y(x_0, y_0) = 0. \end{cases}$$

Thus we have  $(\nabla f + \lambda \nabla g)(x_0, y_0) = 0$ , as desired.  $\square$

**Remark.** The condition  $\nabla g(x_0, y_0) \neq \vec{0}$  is necessary to derive the Lagrange multiplier equation.

**Example 18.0.3.** Minimize  $f(x, y) = x^2 + y^2$  subject to  $g = (x - 1)^3 - y^2 = 0$ . The distance of  $g$  to the origin is clearly minimized at  $(1, 0)$ , but we have

$$\nabla g(1, 0) = (3(x - 1)^2, -2y)|_{(1,0)} = (0, 0)$$

and  $\nabla f(1, 0) = (2, 0)$ . So  $\nabla f \neq \lambda \nabla g$  in this case since  $\nabla g(1, 0) = \vec{0}$ .

**Remark.** So in general, we should also check the points where  $\nabla g = \vec{0}$  separately.

**Theorem 18.2.** Suppose  $f(x_1, \dots, x_n)$  obtains a local minimum (or maximum) at  $p^0 = (x_1^{(0)}, \dots, x_n^{(0)})$  subject to

$$\begin{cases} g_1(x_1, \dots, x_n) = 0 \\ \vdots \\ g_m(x_1, \dots, x_n) = 0, \end{cases}$$

where  $m < n$ . Assume that  $f, g_i \in C^1$  in a neighborhood of  $p^0$  and the matrix

$$\frac{\partial g}{\partial x}(p^0) = \begin{pmatrix} \partial g_1/\partial x_1 & \dots & \partial g_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial g_m/\partial x_1 & \dots & \partial g_m/\partial x_n \end{pmatrix}(p^0)$$

has rank  $m$ . Then there exist  $\lambda_1, \dots, \lambda_m$  such that

$$\begin{cases} (\nabla f + \lambda_1 \nabla g_1 + \dots + \lambda_m \nabla g_m)(p^0) = 0 \\ g_i(p^0) = 0. \end{cases}$$

*Proof.* We prove this in the case of four unknowns and two constraints (the proof can be generalized). Consider

$$\text{minimize (or maximize) } f(x, y, z, t) \quad \text{subject to} \quad \begin{cases} \varphi(x, y, z, t) = 0 \\ \psi(x, y, z, t) = 0. \end{cases}$$

Suppose the minimum (or maximum) of  $f$  is obtained at  $p^0 = (x_0, y_0, z_0, t_0)$ . Then by assumption, there are two variables (say  $z, t$ ) such that

$$\det \frac{\partial(\varphi, \psi)}{\partial(z, t)} \Big|_{p^0} \neq 0.$$

Then by the implicit function theorem, we can find  $z = g(x, y)$  and  $t = h(x, y)$  such that the constraint equations are satisfied. Then  $u(x, y) = f(x, y, z(x, y), t(x, y))$  obtains a minimum (or maximum) at  $(x_0, y_0)$ , so

$$\nabla u(x_0, y_0) = 0 \implies \begin{cases} f'_x + f'_z(\partial z/\partial x) + f'_t(\partial t/\partial x) = 0 \\ f'_y + f'_z(\partial z/\partial y) + f'_t(\partial t/\partial y) = 0. \end{cases} \quad (1, 2)$$



Then

$$\begin{cases} \varphi'_x + \varphi'_z(\partial z/\partial x) + \varphi'_t(\partial t/\partial x) = 0 \\ \psi'_x + \psi'_z(\partial z/\partial x) + \psi'_t(\partial t/\partial x) = 0. \end{cases} \quad \text{and} \quad \begin{cases} \varphi'_y + \varphi'_z(\partial z/\partial y) + \varphi'_t(\partial t/\partial y) = 0 \\ \psi'_y + \psi'_z(\partial z/\partial y) + \psi'_t(\partial t/\partial y) = 0. \end{cases} \quad (\star)$$

Since  $\det(\partial(\varphi, \psi)/\partial(z, t)) \neq 0$  at  $p^0$ , we can find  $\lambda, \mu$  such that

$$\begin{cases} f'_z + \lambda\varphi'_z + \mu\psi'_z = 0 \\ f'_t + \lambda\varphi'_t + \mu\psi'_t = 0, \end{cases}$$

which is always possible since the coefficient matrix is nonsingular. Let  $(A)$  and  $(B)$  be the two systems in  $(\star)$  and  $A_1, A_2$  and  $B_1, B_2$  be their first and second equations, respectively. Then look at the equation  $\lambda(A_1) + \mu(A_2) + (1)$ , which gives

$$f'_x + \lambda\varphi'_x + \mu\psi'_x + \underbrace{(f'_z + \lambda\varphi'_z + \mu\psi'_z)}_{=0} \frac{\partial z}{\partial x} + \underbrace{(f'_t + \lambda\varphi'_t + \mu\psi'_t)}_{=0} \frac{\partial t}{\partial x} = 0.$$

This implies that  $f'_x + \lambda\varphi'_x + \mu\psi'_x = 0$ , and we can similarly get that  $f'_y + \lambda\varphi'_y + \mu\psi'_y = 0$ . Thus we end up with  $\nabla f + \lambda\nabla\varphi + \mu\nabla\psi = 0$  at  $p^0$ , as desired.  $\square$