3

ADVANCED CALCULUS

3.1 Continuous functions

3.1.1 Continuous function on $\mathbb{R}$

3.1.1.1 *Basics*

Definition 3.1.1 (continuity). *[1, p. 122][2, p. 53]*

A function $f : D \to \mathbb{R}$ is continuous at $c \in D$ if for all $\epsilon > 0$, there exists a $\delta > 0$, such that for all $|x - c| < \delta, \left|f(x) - f(c)\right| < \epsilon$, or equivalently,

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

*f is said to be **continuous** on D if f is continuous on every point in D.*

Example 3.1.1. The function $f(x) = 1/x$ is continuous on real line except at $x = 0$.

Definition 3.1.2 (continuity, alternative definition). *[1, p. 123] Let $\{x_n\}$ be a real-valued sequence with limit x. Let f be a function defined on $\mathbb{R}$, then f is continous at x if*

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

Corollary 3.1.0.1 (criterion for discontinuity). *[1, p. 123]Let $f : A \to \mathbb{R}$, and let $c \in A$ be a limit point of A. If there exists a sequence $\{x_n\} \subseteq A$, and $x_n \to c$ but $f(x_n)$ not converges to $f(c)$*

Remark 3.1.1 (exchange of limit and function evaluation). From the perspective of exchange operations, continuity ensures that we can exchange the operation of taking limit and evaluation.

Lemma 3.1.1 (algebraic properties of continuity). *[2, p. 55] Suppose that the function $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ are continuous at the point $x_0 \in D$. Then*

- *the sum $f + g : D \to \mathbb{R}$ is continuous at x_0.*
- *the product $fg : D \to \mathbb{R}$ is continuous at x_0.*
- *If $g(x) \neq 0$ for all $x \in D$, then*

$$f/g : D \to \mathbb{R}$$

is continuous at x_0.

Proof. (1) Let $\{x_n\}$ be a sequence in D convergent to x_0. By the definition of continuity, we have

$$\lim_{n\to\infty} f(x_n) = f(x_0), \lim_{n\to\infty} g(x_n) = g(x_0).$$

The algebraic property of limits Lemma 1.4.1 gives

$$\lim_{n\to\infty} f(x_n) + g(x_n) = f(x_0) + g(x_0).$$

(2) Let $\{x_n\}$ be a sequence in D convergent to x_0. By the definition of continuity, we have

$$\lim_{n\to\infty} f(x_n) = f(x_0), \lim_{n\to\infty} g(x_n) = g(x_0).$$

The algebraic property of limits Lemma 1.4.1 gives

$$\lim_{n\to\infty} f(x_n)g(x_n) = f(x_0)g(x_0).$$

(3) similar to (1)(2). □

Lemma 3.1.2 (continuity of composition). *For functions $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ such that $f(D) \subset U$, assume f continuous at the point $x_0 \in D$ and $g : U \to \mathbb{R}$ is continuous at the point $f(x_0)$. Then the composition*

$$g \circ f : D \to \mathbb{R}$$

is continuous at x_0.

Proof. Let $\{x_n\}$ be a sequence in D convergent to x_0. By the definition of continuity of f, we have

$$\lim_{n\to\infty} f(x_n) = f(x_0)$$

. Then $\{f(x_n)\}$ is a sequence convergent to $f(x_0)$. By the definition of continuity of g, we have

$$\lim_{n\to\infty} g(f(x_n)) = g(f(x_0))$$

. Therefore $g \circ f$ is continuous at x_0. □

Theorem 3.1.1 (continuous via open sets). *[3, p. 12] If $A \subseteq \mathbb{R}^n$, a function $f : A \to \mathbb{R}^m$ is continuous if and only if for every open set $U \in \mathbb{R}^m$ there is some open set $V \subseteq \mathbb{R}^n$ such that $f^{-1}(U) = V \cap A$*

Proof. (1) Suppose f is continuous. If $a \in f^{-1}(U)$, then $f(a) \in U$. Since U is open, then there exist an open ball B in U that $f(a) \in B$. Since f is continuous at a, then there exist an open ball C in A such that $f(C) \subseteq B$. Do this for each $a \in A$, and let the union of all these C be V. Then V will be an open set, since union of open sets are open. And clearly $f^{-1}(U) = V \cap A$. (2) The converse is directly form the limit based definition of continuity. Given $\epsilon > 0$, let B be an open set around $f(a)$ such that $\|f(x) - f(a)\| < \epsilon$. From the assumption, there exist an open set $V \cap A = f^{-1}(B)$, such that every element $x \in V \cap A$ will have $\|f(x) - f(a)\| < \epsilon$. Then there exist an open ball C with radius δ, such that $C \subseteq V \cap A$ $\square$

Remark 3.1.2. The nature of continuity is that when select an open ball B in the range, no matter how small the open ball is, we can always an open ball C in the domain such that $f(C) \subseteq B$. Consider a discontinuous function having an isolated point $f(a)$, choose an small open ball B containing $f(a)$, then $f^{-1}(B)$ is finite sized set, which is closed(see metric space chapter for details.).

Theorem 3.1.2 (Intermediate value theorem). *[2, p. 63] If the real-valued function f is continuous on the closed interval $[a,b]$ and $f(a) < f(b)$. Then for any y in between, $f(a) < y < f(b)$, there exists $x \in (a,b)$ with $f(x) = y$.*

Proof. We use divide and conquer plus nested interval method to prove. We subset the interval $[a_n, b_n]$ half each time, such that the subset $[a_{n+1}, b_{n+1}]$ has the property that y is lying between $f(a_{n+1})$ and $f(b_{n+1})$. Continue this process, we know the intersection is not empty and the element within this interval can be made arbitrarily small to y. $\square$

Remark 3.1.3 (interpretation). The intermediate value theorem is an existence theorem, based on the real number property of completeness

Lemma 3.1.3 (map an interval to an interval). *[2, p. 65] Let I be an interval and suppose that the function $f : I \to \mathbb{R}$ is continuous. Then its image $f(I)$ also is an interval.*

Proof. To show $f(I)$ is an interval, we want to show $f(I)$ is a convex set. Take $y_1, y_2 \in f(I), y_1 < y_2$. From intermediate value theorem we have for any $c \in [y_1, y_2]$, there exists an x_0 between x_1 and x_2, where $f(x_1) = y_1, f(x_2) = y_2$. Therefore, $c \in f(I)$ since $x_0 \in I$. $\square$

Lemma 3.1.4 (preservation on compactness). *[1, p. 129][3, p. 12] If $f : A \to \mathbb{R}^m$ is continuous, where $A \subseteq \mathbb{R}^n$, and A is compact, then $f(A) \subseteq \mathbb{R}^m$ is compact.*

Proof. Let $\mathcal{O}$ be an open cover of $f(A)$. Because every element in $B \in \mathcal{O}$ has an open set $C \subseteq A$ and $C = f^{-1}(B) \cap A$. The pre-image of every element in $\mathcal{O}$ must be an open cover $\mathcal{O}_A$ for A. Because A is compact, then there exist an finite subcover, the image of the subcover must be the finite subcover of $\mathcal{O}$, therefore, every open cover of $f(A)$ has a finite subcover, therefore $f(A)$ is compact. □

3.1.1.2 *Continuity and inverse*

Lemma 3.1.5 (continuity and image topology). *[2, p. 76] Let $f : I \to \mathbb{R}, I = (a, b)$ be monotone increasing and with range $f(I)$ being an interval. Then f is continuous on I.*

conversely, if f is continuous on I, then its range $f(I)$ is also an interval.

Proof. Suppose f has a jump at x_0. Let y_0^- and y_0^+ be the two side limit values at the jump. Then at least $(y_0^-, f(x_0))$ or $(f(x_0), y_0^+)$ must be nonempty(otherwise there cannot be jumps). Pick the nonempty one and denote it J. Note that

$$J \subset (y_0^-, y_0^+) \subset (f(a), f(b)),$$

so the range $(f(a), f(b))$ contains an interval.

Therefore, f cannot contain jumps and must be continuous.

The second part directly from the fact that continuous function maps an interval to an intervalLemma 3.1.3. □

Remark 3.1.4 (geometric intuition). Suppose a monotone increasing function has a jump. Then if standing on the $x = +\infty$ and look into the direction of $-x$, then we will see an jump projecting on the y axis and make the range not an interval.

Lemma 3.1.6 (continuity of inverse). *If f is continuous and strictly monotone on an interval, then f^{-1} is also continuous.*

Proof. (1)first method. Consider strictly increasing function. Let x_0 be any point in the interval and $f(x_0) = y$. For any given $\epsilon > 0$, we take

$$\delta = \min(f(x_0) - f(x_0 - \epsilon), f(x_0 + \epsilon) - f(x_0)),$$

then when $|y - y_0| < \delta$, we have

$$y_0 - \delta < y < y + \delta,$$

Because f^{-1} is also strictly increasing, we have

$$f^{-1}(y_0 - \delta) < f^{-1}(y) < f^{-1}(y + \delta).$$

Further use the fact that

$$x_0 - \epsilon < f^{-1}(y_0 - \delta),$$

we have

$$x_0 - \epsilon < x < x_0 + \epsilon,$$

That is

$$|x - x_0| = \left| f^{-1}(y) - f^{-1}(y_0) \right| < \epsilon.$$

(2) second method. Use the fact that f^{-1} is also strictly monotone and the image of f^{-1} is an interval. Then use Lemma 3.1.5. □

3.1.2 Continuous function in metric space

Definition 3.1.3 (continuous in metric space). *[4, p. 136] Let (M_1, d_1) and (M_2, d_2) be metric spaces, let f be a function from M_1 to M_2. We say function f is continuous at some point c if for every $\epsilon > 0$, there exist a $\delta > 0$, such that for all points with $d_1(x, c) < \delta$, we have $d_2(f(x), f(c)) < \epsilon$. [4]*

Definition 3.1.4 (continuous in metric space, alternative). *[4, p. 136] Let f be a function from a metric space M_1 into a metric space M_2. Let $a \in M_1$. Then f is continuous at a if and only if whenever $\{x_n\}$ is a sequence in M_1 such that $\lim_{n\to\infty} x_n = a$, then $\lim_{n\to\infty} f(x_n) = f(a)$*

Theorem 3.1.3 (Preserving topological property via continuity). *[4, p. 139] Let f be a function from a metric space M_1 into a metric space M_2. The following are equivalent:*

- *f is continuous on M_1*
- *$f^{-1}(C)$ is closed whenever C is a closed subset of M_2*
- *$f^{-1}(C)$ is open whenever C is a open subset of M_2*

Proof. We first prove (1)(3) are equivalent: (a) Suppose f is continuous. If $a \in f^{-1}(U)$, then $f(a) \in U$. Since U is open, then there exist an open ball B in U that $f(a) \in B$. Since f is continuous at a, then there exist an open ball C in A such that $f(C) \subseteq B$. Do this for each $a \in A$, and let the union of all these C be V. Then V will be an open set, since union of open sets are open. And clearly $f^{-1}(U) = V \cap A$. (b) The converse is directly form the limit based definition of continuity. Given $\epsilon > 0$, let B be an open set around $f(a)$ such that $\|f(x) - f(a)\| < \epsilon$. From the assumption, there exist an open set $V \cap A = f^{-1}(B)$, such that every element $x \in V \cap A$ will have $\|f(x) - f(a)\| < \epsilon$. Then there exist an open ball C with radius δ, such that $C \subseteq V \cap A$ □

Lemma 3.1.7 (preservation on compactness). *[4, p. 152] If f is a continuous function from a compact metric space to M_1 to a metric space M_2, then $f(M_1)$ is compact.*

Proof. For any sequence in $f(M_1)$, the inverse image of the sequence in M_1 will have a convergent subsequence $\{x_i\}$. Then then image of this subsequence will be a convergent subsequence in $f(M_1)$ since f is continuous. Therefore, every sequence in $f(M_1)$ has a convergent subsequence and thus $f(M_1)$ is compact. □

Corollary 3.1.3.1. *If f is a continuous function from a compact metric space to M_1 to a metric space M_2, then $f(M_1)$ is closed and bounded.*

Proof. Use the fact that compact set is closed and bounded. □

3.1.3 Boundedness and extreme value theorems

Lemma 3.1.8 (Boundedness of continuous function). *If f is continuous at c in real line, there exists $\delta > 0$ such that f is bounded on $(c - \delta, c + \delta)$*

Proof. Because f is continuous at c, then given $\epsilon = 1$, there exist a $\delta > 0$, such that $|f(x) - f(c)| < 1$ on whenever $x \in (c - \delta, c + \delta)$. □

Theorem 3.1.4 (Boundedness of continuous function on closed interval). *[4, p. 114]If f is continuous at $[a, b]$, then f is bounded on $[a, b]$.*

Proof. Note that there exists an open interval I_c around c where f is bounded. The collections of $I_c, c \in [a, b]$ is an open cover for $[a, b]$, then there exist a finite subcover. Since f is bounded on each of the element in the subcover, f is bounded on $[a, b]$. □

Theorem 3.1.5 (Extreme value theorem in $\mathbb{R}$). *[4, p. 114][1, p. 130] If a function f is defined on a closed interval $A = [a, b]$ (or any closed and bounded set) and is continuous there, then the function attains its maximum; The same is true of the minimum of f.*

Proof. Because $f(A)$ is compact, then it has a maximum and minimum contained in $f(A)$. □

Lemma 3.1.9 (boundedness at continuous point). *[4, p. 146] Let f be a real-valued function on a metric space M. If f is continuous at $a \in M$, then there exist an open set $U \subset M$ containing a such that f is bounded on U.*

Proof. Since f is continuous at a, then for $\epsilon = 1$, there exists an open ball $B(a, \delta)(\delta > 0)$ such that

$$\left|f(x) - f(a)\right| < 1, \forall x \in B(a, \delta)$$

which implies

$$\left|f(x)\right| < \left|f(a)\right| + 1, \forall x \in B(a, \delta)$$

□

Theorem 3.1.6 (boundedness at compact set). *[4, p. 146] If f is continuous on a compact metric space M then f is bounded on M.*

Proof. Since f is continuous at every point $a \in M$, then the associated collection of open balls will cover M. Because M is compact, then there is a finite subcover covers M. Since f is bounded on each subset in the subcover, f will be bounded on the subcover. □

Corollary 3.1.6.1 (Weierstrass theorem). *[4, p. 146] A continuous real-valued function on a compact set M achieves a maximum and a minimum.*

Proof. Use the fact the f is bounded on M and the real line is complete. □

Corollary 3.1.6.2 (non-empty compact lower set implies existence of global minimizer). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous on all of $\mathbb{R}^n$. If there exists a scalar $a \in \mathbb{R}$ such that the level set*

$$S = \{x \in X : f(x) < a\}$$

is nonempty and compact, then there exists at least one global minimum.

Proof. There exists at least one global solution $f(x^*)$ on S. And since $S \subset \mathbb{R}^n$ and $f(x) > \alpha, \forall x \in \mathbb{R}^n - S$, therefore $f(x^*)$ will be the global minimizer. □

3.1.4 More on extreme values

Definition 3.1.5 (coercive function). *A function $\mathbb{R}^D \to \mathbb{R}$ is said to be coercive if for every sequence $\{x_n\} \in \mathbb{R}^D, \|x_n\| \to \infty$ implies $f(x_n) \to \infty$ as $n \to \infty$.*

Remark 3.1.5.

- Any monotonic function will not be coercive.
- Coercive functions have the general bowl shape that goes to positive infinite as $\|x\|$ goes to infinite.

Lemma 3.1.10 (coercity and compactness). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous on all of $\mathbb{R}^n$. The function f is coerice if and only if for every $\alpha \in \mathbb{R}$ the set $\{x|f(x) \leq \alpha\}$ is compact.*

Proof. (1) First the interval $(-\infty, \alpha]$ is a closed set, therefore the preimage $f^{-1}((-\infty, \alpha])$ is a closed set due to continuity of f [Theorem 3.1.3]. The set $f^{-1}((-\infty, \alpha])$ must be bounded, otherwise as $\|x\| \to \infty$, f can not be bounded by α by the definition of coercity of f. Therefore, for every $\alpha \in \mathbb{R}$ the set $\{x|f(x) \leq \alpha\}$ is closed and bounded, and hence compact. (2) Consider a sequence $\{x_n\}$ where $\|x_n\| \to \infty$ as $n \to \infty$. Suppose there exists a subsequence $\{x_{n_k}\}$ such that $\{f(x_{n_k})\}$ is bounded, then there will a compact set $S = \{x|f(x) \leq \alpha\}$ such that $\{x_{n_k}\} \subset S$. Since S is compact, then the subsequence $\{x_{n_k}\}$ must be bounded, which contradicts $\|x_n\| \to \infty$ as $n \to \infty$. Therefore, for any sequence $\|x_n\| \to \infty$ as $n \to \infty$, we have $f(x_n) \to \infty$. □

Theorem 3.1.7 (existence of global minimizer for coercive function). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous on all of $\mathbb{R}^n$. If f is coercive, then f has at least one global minimizer.*

Proof. Choose $\alpha \in \mathbb{R}$ such that the set $S = \{x|f(x) \leq \alpha\}$ is non-empty. By coercivity, this set is compact, and therefore there exists at least one global solution $f(x^*)$. And since $S \subset \mathbb{R}^n$ and $f(x) > \alpha, \forall x \in \mathbb{R}^n - S$, therefore $f(x^*)$ will be the global minimizer. □

3.1.5 Curves and surfaces

Definition 3.1.6 (curve). *A parameterized smooth curve in $\mathbb{R}^n$ is a differentiable map $r : I \subset \mathbb{R} \to \mathbb{R}^n$ from an open interval of the real line to $\mathbb{R}^n$, such that*

$$r(t) = (x_1(t), x_2(t), ..., x_n(t))$$

where $x_1(t), x_2(t), ..., x_n(t) \in C^\infty$ are the differentiable components or coordinate functions of $r(t)$.

A parameterized smooth curve $r(t) : I \to \mathbb{R}^n$ is called regular if $\dot{r}(t) \neq 0, \forall t \in I$.

Example 3.1.2.

- $n = 2$: the curve $r(t) = (x(t), y(t)) \in \mathbb{R}^2$ is called a plane curve.
- $n = 3$: the curve $r(t) = (x(t), y(t), z(t)) \in \mathbb{R}^2$ is called a space curve [Figure 3.1.1].

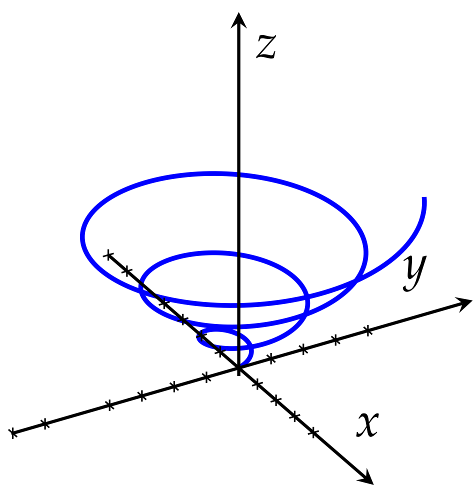

Figure 3.1.1: An example 3D curve generated by $z = t, x = t\cos(t), y = t\sin(t)$.

Definition 3.1.7 (unit tangent vector, normal). *Let $x(t)$ be a smooth curve, then its unit tangent vector at $x(t_0)$is defined as*

$$T(t_0) \triangleq \frac{\dot{x}(t_0)}{\|\dot{x}(t_0)\|};$$

it normal at $x(t_0)$is defined as

$$N(t_0) \triangleq \frac{\dot{T}(t_0)}{\left\|\dot{T}(t_0)\right\|}.$$

Lemma 3.1.11 (normal is perpendicular to tangent).

$$N(t) \perp T(t)$$

or equivalently,

$$N(t) \cdot T(t) = 0.$$

Proof. Note that

$$N(t) \cdot T(t) = \frac{dT}{dt} \cdot T / \|N\| = \frac{1}{2\|N\|} \frac{d\|T\|^2}{dt} = 0.$$

where we use the fact that $\|T\|^2 = 1$. □

Theorem 3.1.8 (length of a curve). *[5, p. 26] Let $c : [a, b] \to \mathbb{R}^3$ be a smooth, regular parametrized curve. Then length of c is defined to be*

$$l = \int_a^b \sqrt{(\frac{dc}{dt})^T \frac{dc}{dt}} dt,$$

where $dc/dt \in \mathbb{R}^3$.

Proof. Can be proven by starting from the definition that the length of a curve is the sum of the tiny segments. □

Lemma 3.1.12 (curve length is independent of parameterization). *[5, p. 27] Let $c : [a, b] \to \mathbb{R}^3, t \in [a, b]$ be a smooth, regular parametrized curve. Let $t' : [a, b] \to [a', b']$, such that $c(t') : [a', b'] \to \mathbb{R}^3$ is another smooth and regular parametrization of the same curve. Then length of c is defined to be*

$$l = \int_a^b \sqrt{(\frac{dc}{dt})^T \frac{dc}{dt}} dt = \int_{a'}^{b'} \sqrt{(\frac{dc}{dt'})^T \frac{dc}{dt'}} dt'.$$

Proof.

$$l = \int_a^b \sqrt{(\frac{dc}{dt})^T \frac{dc}{dt}} dt = \int_{a'}^{b'} \sqrt{(\frac{dc}{dt'} \frac{dt'}{dt})^T \frac{dc}{dt'} \frac{dt'}{dt}} \left| \frac{dt}{dt'} \right| dt' = \int_{a'}^{b'} \sqrt{(\frac{dc}{dt'})^T \frac{dc}{dt'}} dt'.$$

□

Definition 3.1.8 (arc length and its parameterization). *[5, p. 28] Let $x(t) \in \mathbb{R}^3, t \in [t_0, b]$ be a curve, then the function*

$$s(t) = \int_{t_0}^{t} \left\| \frac{dx}{dt} \right\| dt$$

is called the ***arc length of the curve****.*

The parameterization of a curve by its arc length is known as ***natural parameterization****. And we have*

$$ds = \left\| \frac{dx}{dt} \right\| dt$$

and

$$\frac{dx(s)}{ds} = \frac{dx(s)}{dt}\frac{dt}{ds} = \frac{\frac{dx}{dt}}{\left\| \frac{dx}{dt} \right\|}.$$

Definition 3.1.9 (unit tangent vector, alternative). *[5, p. 29] Let $x(s)$ be a curve with natural parameterization, then its unit tangent vector is defined as*

$$\frac{dx}{ds},$$

which is of unit length via

$$\frac{dx(s)}{ds} = \frac{dx(s)}{dt}\frac{dt}{ds} = \frac{\frac{dx}{dt}}{\left\| \frac{dx}{dt} \right\|}.$$

3.1.5.1 *Curvature*

Definition 3.1.10 (curvature of a curve). *The curvature of a curve $t(s)$ with natural parameterization is defined as*

$$\kappa(s) = \left\| \frac{dt(s)}{ds} \right\|_2 = \left\| \frac{d^2x}{ds^2} \right\|_2$$

The radius of curvature is defined as

$$\rho(s) = \frac{1}{\kappa(s)}.$$

3.1.5.2 *Surfaces*

Definition 3.1.11 (patch). *[6, p. 134] A patch is smooth map $r : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ from an open subset U of real plane to $\mathbb{R}^3$, such that*

$$r(u, v) = (x(u, v), y(u, v), z(u, v))$$

where the differentiable coordinate functions $x(u, v), y(u, v), z(u, v)$, which are functions from $U \to \mathbb{R}$ and in C^∞.

Remark 3.1.6. A patch is essentially a smooth map.

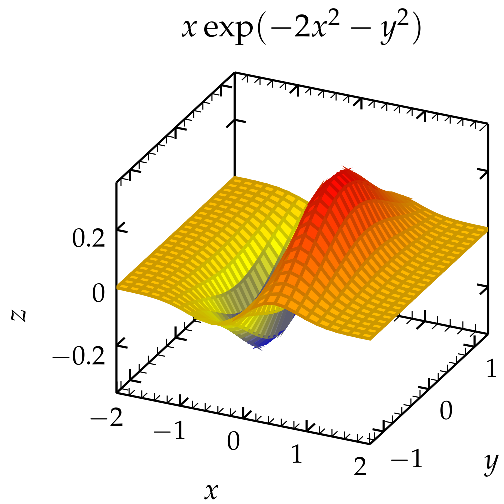

Figure 3.1.2: An example smooth surface generated by $z = x \exp(-2x^2 - y^2)$

Lemma 3.1.13 (linear independence). *The vector r_u and r_v are linearly independent at a point (u, v) if any one of the following holds:*

- $$det \begin{pmatrix} r_u \cdot r_u & r_u \cdot r_v \\ r_u \cdot r_v & r_v \cdot r_v \end{pmatrix} \neq 0$$
- $r_u \times r_v \neq 0$

Proof. (1) Note that $det(*) = (r_u \cdot r_u)(r_v \cdot r_v) - (r_u \cdot r_v)^2 \geq 0$ is the Cauchy inequality and the equality only holds when $r_u = \lambda r_v$(which is linear independence); (2) note that $\|r_u \times r_v\|^2 = det(*)$ □

Definition 3.1.12 (regular patch). *[6, p. 135] A patch $r : U \subset \mathbb{R}^2 \to \mathbb{R}^3$(where U is an open set) is called a* ***regular patch*** *if $r_u \times r_v \neq 0$*

Definition 3.1.13 (regular surface). *[6, p. 135] A subset M of $\mathbb{R}^3$ is called a regular surface if, for* ***any point*** *p in M there is an open neighborhood set V of p in $\mathbb{R}^3$ and a map $r : U \to V \cap M$, where U is an open set in $\mathbb{R}^2$, and $r(u, v) = (x(u, v), y(u, v), z(u, v))$, such that*

- *r is smooth*
- *r is a homeomorphism*
- *r is regular*

Definition 3.1.14 (tangent plane). *The tangent plane T_pM at a point p of a regular surface M is the set of tangent vectors at p of all curves in M passing through p.*

Remark 3.1.7 (tangent plane form a subspace)**.** The tangent plane T_pM is a vector subspace and it can be spanned by $\{r_u, r_v\}$

3.2 Uniform continuity

3.2.1 Uniform continuity on real line

3.2.1.1 *Concepts*

Definition 3.2.1 (uniformly continuous). *[2, p. 67] A function $f : D \to \mathbb{R}$ is said to be* ***uniformly continuous*** *provided that whenever $\{u_n\}$ and $\{v_n\}$ are sequences in D such that*

$$\lim_{n\to\infty} u_n - v_n = 0,$$

we have

$$\lim_{n\to\infty} f(u_n) - f(v_n) = 0.$$

Definition 3.2.2 (uniformly continuous, alternative). *[2, p. 73] A function $f : D \to \mathbb{R}$ is said to be* ***uniformly continuous*** *if given any $\epsilon > 0$, there exists a positive number δ such that for* ***all*** *u, v in D,*

$$\left|f(u) - f(v)\right| < \epsilon, \text{ whenever } |u - v| < \delta.$$

Note 3.2.1 (uniform continuity vs. continuity). Note that uniform continuity is a stronger continuity condition. If a function is uniformly continuous then it is continuous. It can be showed by replaced v to be any fixed $x_0 \in D$.

Example 3.2.1. Define $f(x) = x^2, x \in \mathbb{R}$. Then f is continuous on $\mathbb{R}$ but not uniformly continuous on $\mathbb{R}$. For example, set

$$u_n = n, v_n = n + 1/n.$$

Then

$$\lim_{n\to\infty} u_n - v_n = \lim_{n\to\infty} 1/n = 0,$$

but

$$\lim_{n\to\infty} f(u_n) - f(v_n) = \lim_{n\to\infty} 2 + 1/n^2 = 2 \neq 0.$$

Example 3.2.2. Define $f(x) = 1/x, x \in (0,1)$. Then f is continuous on $(0,1)$ but not uniformly continuous on $\mathbb{R}$. For example, set

$$u_n = 1/n, v_n = 1/2n.$$

Then

$$\lim_{n\to\infty} u_n - v_n = \lim_{n\to\infty} 1/2n = 0,$$

but

$$\lim_{n\to\infty} f(u_n) - f(v_n) = \lim_{n\to\infty} -n \neq 0.$$

Lemma 3.2.1 (uniform continuity preserving Cauchy sequence). *Suppose $D \subset R$ and $f : D \to R$ is uniformly continuous. If $\{x_n\}$ is a Cauchy sequence in D, then $\{f(x_n)\}$ is a Cauchy sequence in $f(D)$.*

Proof. Given any $\epsilon_1 > 0$, there exists $N > 0$, such that if $m, n > N$, then

$$|x_m - x_n| < \epsilon_1$$

Given any $\epsilon_2 > 0$, there exist a δ, such that if $|x_m - x_n| < \delta$, we have

$$|f(x_m) - f(x_n)| < \epsilon_2$$

Set $\epsilon_1 = \delta$, then we can find such $N > 0$, such that if $m, n > N$, then

$$|f(x_m) - f(x_n)| < \epsilon_2$$

□

Remark 3.2.1. Notes:

- Continuity is insufficient to preserve Cauchy sequences, but uniform continuity will.
- The function $f(x) = x/(1-x)$ is continuous but not uniformly continuous. The sequence $x_n = 1 - 1/n$ is a Cauchy sequence, its image $f(x_n) = n - 1$ is not. [7]

Lemma 3.2.2 (continuous function on a closed bounded interval is uniformly continuous). *[2, p. 68] Let $f : [a,b] \to \mathbb{R}$ be a continuous function on $[a,b]$, then f is uniformly continuous.*

Proof. Let $\{u_n\}$ and $\{v_n\}$ be two sequences in $[a,b]$ such that

$$\lim_{n\to\infty} u_n - v_n = 0.$$

For the purpose of contradiction, we assume there exist some $\epsilon > 0$ such that

$$\left|f(u_n) - f(v_n)\right| \geq \epsilon, \forall n.$$

From Theorem 1.6.2, there exists subsequences $\{u_{n_k}\}, \{v_{n_k}\}$ converge to the same limit x_0.

However, $f(x)$ is continuous, we have

$$\lim_{k\to\infty} f(u_{n_k}) = f(x_0), \lim_{k\to\infty} f(v_{n_k}) = f(x_0).f(x_0).$$

Therefore,

$$\lim_{k\to\infty} f(u_{n_k}) - f(v_{n_k}) = 0,$$

contradicts the fact that there exist some $\epsilon > 0$ such that

$$\left|f(u_n) - f(v_n)\right| \geq \epsilon, \forall n.$$

□

Lemma 3.2.3 (uniform continuity on open interval implies boundedness). *Suppose that the function $f : (a, b) \to \mathbb{R}$ is uniformly continuous. Then f is bounded on $A = (a, b)$.*

Proof. Assume that f is unbounded, and $sup_{x\in A} f(x) = \infty$. Then there exits a sequence $\{x_n\} \in A$ such that $\lim_{n\to\infty} f(x_A) = \infty$. Pick a subsequence $\{y_n\}$ from $\{x_n\}$ such that $f(y_{n+1}) - f(y_n) > 1, \forall n \in \mathcal{N}$.

Since f is uniformly continuous, we should have that there exists a δ such that

$$\left|x - y\right| < \delta, x, y \in A \implies \left|f(x) - f(y)\right| < 1.$$

From Theorem 1.6.2, the bounded sequence $\{y_n\}$ has a convergent subsequence $\{z_n\}$ such that for δ, there exists N such that for all $m, n > N$

$$|z_m - z_n| < \delta, \left|f(z_m) - f(z_n)\right| > 1,$$

which is a contradiction. □

Remark 3.2.2 (compare with continuous function). Note that a continuous function on an open interval cannot ensure boundedness; for example $f(x) = 1/x, x \in (0, 1)$.

3.2.1.2 *Lipschitz continuity*

Definition 3.2.3 (Lipschitz continuous).

- *A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}$ is* ***locally Lipschitz continuous*** *at $x_0 \in X$ if there is a constant $\beta \geq 0$ and $\delta > 0$ such that for every $\left|x - y\right| < \delta$*

$$\left|f(x) - f(x_0)\right| \leq \beta |x - x_0| \,.$$

- *A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}$ is* ***globally Lipschitz continuous*** *on X if there is a constant $\beta \geq 0$ such that for every $x, x_0 \in X$*

$$\left|f(x) - f(x_0)\right| \leq \beta |x - x_0|$$

Lemma 3.2.4 (Lipschitz continuity implies continuity). *Suppose a function $f : X \to \mathbb{R}$ is locally Lipschitz continuous at $x_0 \in X$ with a constant $\beta \geq 0$. Then f is continuous at x_0.*

Proof. Let $\{x_n\}$ be a sequence with limit of x_0. From definition of Lipschitz continuity, we have that any given $\epsilon > 0$, there exist a N such that for all $n > N$, we have

$$|x - x_n| < \epsilon/\beta,$$

and

$$\left|f(x_n) - f(x_0)\right| < \beta |x - x_n| < \epsilon.$$

That is

$$\lim_{n\to\infty} f(x_n) = f(x_0).$$

□

Lemma 3.2.5 (globally Lipschitz continuity implies uniformly continuity). *A globally Lipschitz function is uniformly continuous.*

Proof. Let β be the Lipschitz constant. Given any $\epsilon > 0$, we can find a $\delta = \epsilon/\beta$

$$\left|f(x) - f(y)\right| < \beta \left\|x - y\right\| < \epsilon$$

for $\left\|x - y\right\| \leq \delta$ □

Lemma 3.2.6 (differentiability and local Lipschitz continuity). *[4, p. 312]*

- *If $f : D \to \mathbb{R}$ is differential at x, then it is locally Lipschitz continuous at x;*
- *If f is locally Lipschitz continuous at x with constant K, then f is differentiable at x and with $\left|f'(x)\right| \leq K$.*

- *A function f is differential and $|f'(x)| \leq M, \forall x$, then f is globally Lipschitz and uniformly continuous.*
- *A continuously differentiable function f defined on a closed and bounded interval of $\mathbb{R}$ is globally Lipschitz continuous and uniformly continuously.*

Proof. (1)Because f is differential at x, then given $\epsilon = 1$, there exist a $\delta > 0, \forall 0 < |y - x| < \delta$, we have

$$\left| \frac{f(y) - f(x)}{y - x} - f'(x) \right| < 1$$

rearrange and we have

$$|f(y) - f(x)| \leq (1 + f'(x))|y - x|.$$

(2) If it is locally Lipschitz, then there exists a neighborhood around x such that

$$\frac{f(y) - f(x)}{y - x}$$

is bounded. When take the limit $y \to x, y \neq x$, the limit is bounded. (3) from (1). (4) A continuously differentiable function on a closed and bounded interval has bounded derivatives. □

Example 3.2.3.

- $f(x) = |x|$ is **globally Lipschitz continuous and uniformly continuous but not differentiable at** $x = 0$.
- $f(x) = \sqrt{x}$ defined on $[0, 1]$ is **continuous but not Lipschitz continuous at** $x = 0$.
- A differentiable function f on the real numbers need not be a continuously differentiable function, that is, f' is not continuous. Consider the function:

$$g(x) = \begin{cases} x^{3/2} \sin(1/x), x \neq 0 \\ 0, x = 0 \end{cases}$$

Then

$$g'(x) = \begin{cases} \frac{3}{2} x^{1/2} \sin(1/x) - x^{-1/2} \cos(1/x), x \neq 0 \\ 0, x = 0 \end{cases}.$$

Note that g has unbounded derivative when $x \to 0$, but $g'(0) = \lim_{x\to 0} \frac{x^{3/2}\sin(1/x)}{x} = 0$. g is not global Lipschitz continuous since g' is unbounded near $x = 0$.

Remark 3.2.3 (continuity, Lipschitz continuity, uniform continuity, differentiability). In general, for functions defined over a closed and bounded subset of real line, we have

$$continuously\ differentiable \subseteq Lipschitz\ continuous \subseteq uniformly\ continuous\ continuous$$

3.2.2 Uniform continuity on metric space

Definition 3.2.4. *A function $f : X \to Y$ is* uniformly continuous *if for every $\epsilon > 0$ there exist a $\delta > 0$ such that for every $x, x_0 \in X$,*

$$d(x, x_0) \leq \delta \Rightarrow d(f(x), f(x_0)) < \epsilon$$

Remark 3.2.4 (criterion for absence of uniform convergence). [1, p. 132] A function $f : A \to B$ fails to be uniformly continuous on A if there exists a particular $\epsilon > 0$ and two sequence $\{x_n\}$ and $\{y_n\}$ in A satisfying

$$d(x_n, y_n) \to 0, n \to \infty$$

but

$$d(f(x_n) - f(y_n)) \geq \epsilon$$

Theorem 3.2.1 (continuity implies uniform continuity on compact set). *[4, p. 154]If f is continuous function from a compact metric space M_1 into a metric space M_2, then f is uniformly continuous on M_1.*

Proof. Given a $\epsilon > 0$, for every $z \in M_1$, there exists a $delta(z) > 0$, such that for all $x \in B(z, \delta(z))$, $d(f(x), f(z)) < \epsilon$. These open balls will form an open cover on M_1, and since M_1 is compact, there exists a finite subcover consisting of a finite set J of open balls. Let δ be the smallest radius of open balls in J, then for any point $x \in M$, there exists an open ball with center z within δ distance to x. Then for any point x, y satisfying $d(x, y) < \delta/2$, there will exist a z such that $x, y \in B(z, \delta)$. Then we can use triangle inequality as

$$d(f(x), f(y)) \leq d(f(x), f(z)) + d(f(z), f(y)) = 2\delta$$

□

Corollary 3.2.1.1. *If f is a continuous real-valued function on a closed and bounded subset X of $\mathbb{R}^n$, then f is uniformly continuous on X.*

Remark 3.2.5.

- uniform continuity implies continuity, but not converse.

3.2.3 Locally and globally Lipschitz continuous

Definition 3.2.5 (locally Lipschitz continuous). *A function $f : X \to Y$ is locally* Lipschitz *continuous at $x_0 \in X$ if there is a constant $\beta \geq 0$ and $\delta > 0$ such that for every $\|x - y\| < \delta$*

$$\|f(x) - f(x_0)\| \leq \beta \|x - x_0\|$$

Lemma 3.2.7 (sufficient condition for local Lipschitz). *[4, p. 312] If f is differential at x, then it is locally Lipschitz continuous at x.*

Proof. Because f is differential at x, then given $\epsilon = 1$, there exist a $\delta > 0, \forall 0 < |y - x| < \delta$, we have

$$\left| \frac{f(y) - f(x)}{y - x} - f'(x) \right| < 1$$

rearrange and we have

$$|f(y) - f(x)| \leq (1 + f'(x)) \|y - x\|$$

□

Lemma 3.2.8 (necessary and sufficient condition for local Lipschitz). *A differentiable function $g : \mathbb{R} \to \mathbb{R}$ is locally Lipschitz continuous at $x$if and only if it has bounded first derivative at x*

Proof. (1) The sufficient part is the same as above theorem. (2) If it is locally Lipschitz, then there exists a neighborhood around x such that

$$\frac{f(y) - f(x)}{y - x}$$

is bounded. When take the limit $y \to x, y \neq x$, the limit is bounded. □

Example 3.2.4. The function $f(x) = x^{1/3}$ has unbounded first derivative at $x = 0$, and therefore it is not locally Lipschitz there.

Definition 3.2.6 (globally Lipschitz continuous). *A function $f : X \to Y$ is globally* Lipschitz *continuous if there is a constant $\beta \geq 0$ such that for every $x, x_0 \in X$*

$$\left|f(x) - f(x_0)\right| \leq \beta \|x - x_0\|$$

Lemma 3.2.9 (globally Lipschitz continuous implies uniformly continuous). *A globally Lipschitz function is uniformly continuous.*

Proof. Let β be the Lipschitz constant. Given any $\epsilon > 0$, we can find a $\delta = \epsilon/\beta$

$$\left|f(x) - f(y)\right| < \beta \left\|x - y\right\| < \epsilon$$

for $\left\|x - y\right\| \leq \delta$ □

Lemma 3.2.10 (bounded differential implies global Lipschitz). *A function f is differential and $\left|f'(x)\right| \leq M, \forall x$, then f is globally Lipschitz.*

Proof. from mean value theorem. □

3.3 Differentiation

3.3.1 Differential function concept

Definition 3.3.1 (differential function).

- *(differentiability at a point)Let f be a real-valued function define at $(a,b) \subset \mathbb{R}$. If the limit:*
$$f'(x) = \lim_{y\to x} \frac{f(y)-f(x)}{y-x}$$
exists, then f is differentiable at x.
- *(differentiability for an open interval)Let f be a real-valued function define at $(a,b) \subset \mathbb{R}$. We say f is a differentiable function in this interval (a,b) if f is differentiable for $\forall x \in (a,b)$.*

Lemma 3.3.1 (differentiability implies continuity). *[2, p. 91] If f is differentiable at a point c, then f is continuous at c.*

Proof.
$$\lim_{y\to c} f(y) - f(c) = f'(c) \lim_{y\to c}(y-c) = 0.$$
□

Definition 3.3.2 (continuously differentiable function). *A function f is continuously differentiable at x if*
$$\lim_{y\to x} \frac{f(y)-f(x)}{y-x} = f'(x) = \lim_{y\to x} f'(y)$$
that is, the derivative f' is continuous at x.

Note 3.3.1 (caution!). A differentiable function f on the real numbers need not be a continuously differentiable function, that is, f' is not continuous. Consider the function:
$$g(x) = \begin{cases} x^2 \sin(1/x), x \neq 0 \\ 0, x = 0 \end{cases}$$
Then
$$g'(x) = \begin{cases} 2x\sin(1/x) - \cos(1/x), x \neq 0 \\ 0, x = 0 \end{cases}$$

In particular, $g'(0) = 0$, but $\lim_{x\to 0} g'(x)$ does not exist. Note that the $g'(0)$ can be calculated using the definition of derivative.

3.3.2 Differential rules

Lemma 3.3.2 (algebraic rules of differentiation). *[2, p. 91] Let I be a neighborhood of x_0 and suppose that the function $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ are differentiable at x_0. Then*

- *the sum $f + g : I \to \mathbb{R}$ is differentiable at x_0 and*
$$(f+g)'(x_0) = f'(x_0) + g'(x_0).$$
- *the product $fg : I \to \mathbb{R}$ is differentiable at x_0 and*
$$(fg)'(x_0) = f(x_0)g'(x_0) + f'(x_0)g(x_0).$$
- *If $g(x) \neq 0 \forall x \in I$, then the reciprocal $1/g : I \to \mathbb{R}$ is differentiable at x_0 and*
$$(\frac{1}{g})'(x_0) = \frac{-g'(x_0)}{(g(x_0))^2}.$$

Proof. (1)
$$\lim_{x\to x_0} \frac{(f+g)(x) - (f+g)(x_0)}{x - x_0} = \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} + \frac{g(x) - g(x_0)}{x - x_0} = f'(x_0) + g'(x_0).$$

(2)
$$\begin{aligned} & \frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0} \\ = & \frac{f(x)g(x) - f(x)g(x_0) + f(x)g(x_0) - f(x_0)g(x_0)}{x - x_0} \\ = & f(x)\frac{g(x) - g(x_0)}{x - x_0} + g(x_0)\frac{f(x) - f(x_0)}{x - x_0} \end{aligned}$$

Take limits and use the algebraic properties of limitsLemma 1.4.1. (3) Note that
$$\begin{aligned} & \frac{1/g(x) - 1/g(x_0)}{x - x_0} \\ = & -\frac{1}{g(x)g(x_0)}\frac{g(x) - g(x_0)}{x - x_0} \end{aligned}$$

Take limits and use the algebraic properties of limits Lemma 1.4.1. □

Lemma 3.3.3 (derivative of the composition, chain rule). *[2, p. 99] Let I be a neighborhood of x_0 and suppose that the function $f: I \to \mathbb{R}$ and $g: I \to \mathbb{R}$ are differentiable at x_0. Let J be an open interval such that $f(I) \subseteq J$ and suppose that the function $g: J \to \mathbb{R}$ is differentiable at $f(x_0)$. Then the composition $g \circ f: I \to \mathbb{R}$ is differentiable at x_0 and*

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

Proof. (informal) Define $y_0 = f(x_0), y = f(x)$. And we have

$$\frac{(g \circ f)(x) - (g \circ f)(x_0)}{x - x_0} = \frac{(g(y) - g(y_0)}{x - x_0} = \frac{(g(y) - g(y_0)}{y - y_0} \frac{f(x) - f(x_0)}{x - x_0},$$

where we use the fact that

$$\frac{f(y) - f(y_0)}{x - x_0} = 1.$$

□

Lemma 3.3.4 (derivative of the inverse). *[2, p. 97] Let I be a neighborhood of x_0 and suppose that the function $f: I \to \mathbb{R}$ be strictly monotone and continuous. Suppose that f is differentiable at x_0 and that $f'(x_0) \neq 0$.*

- *Define $J = f(I)$. The inverse $f^{-1}: J \to \mathbb{R}$ is differentiable at $y_0 = f(x_0)$.*
- $$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}.$$

Proof. Take $y_0 = f(x_0)$. For $y \in J, y \neq y_0$, define $x = f^{-1}(y)$.

Note that

$$\frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0} = \frac{x - x_0}{f(x) - f(x_0)} = \frac{1}{\frac{f(x) - f(x_0)}{x - x_0}}.$$

Use the fact that f^{-1} is continuous [Lemma 3.1.6]: when $y \to y_0$, $x \to x_0$. Then we can take the limits of the above and use the algebraic properties of limits [Lemma 1.4.1]. □

Lemma 3.3.5 (leibniz rule). *Let u, v be n-times differentiable functions, then the product is also n-times differentiable and its nth derivative is given by*

$$(uv)^n = \sum_{k=0}^{n} \binom{n}{k} u^{(k)} v^{(n-k)}$$

Proof. This can be proof directly using Pascal's triangle. □

3.3.3 Mean value theorem

Lemma 3.3.6 (derivative characterization of extreme values). *[2, p. 103] Let I be a neighborhood of x_0 and suppose that the function $f : I \to \mathbb{R}$ is differentiable at x_0. If the point x_0 is either a maximizer or a minimizer of the function f, then $f'(x_0) = 0$.*

Proof. Let x_0 be the point $f(x_0)$ takes the maximum value. By the definitions of derivative and limits, we have

$$f'(x) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \to x_0, x < x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \to x_0, x > x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

From

$$f(x_0) = \lim_{x \to x_0, x < x_0} \frac{f(x) - f(x_0)}{x - x_0} \geq 0,$$

and

$$f(x_0) = \lim_{x \to x_0, x > x_0} \frac{f(x) - f(x_0)}{x - x_0} \leq 0.$$

Therefore, $f(x_0) \leq 0, f(x_0) \geq 0 \implies f(x_0) = 0$.

□

Theorem 3.3.1 (mean-value theorem). *[4][2, p. 103]*

- ***(Rolle's Theorem)*** *Suppose that the function $f : [a, b] \to \mathbb{R}$ is continuous the that the restriction of f to the open interval (a, b) is differentiable. Assume that*

$$f(a) = f(b),$$

then there exists a point x_0 in the open interval (a, b) at which $f'(x_0) = 0$.

- *If f and g are continuous real functions on $[a,b]$, and differentiable on (a,b), then there is a point $c \in (a,b)$ such that*

$$[f(b) - f(a)]g'(c) = [g(b) - b(a)]f'(c);$$

if $g'(x) \neq 0 \forall x \in (a,b)$, we can write as

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

- *If f is real-valued and continuous on $[a,b]$ and differentiable in (a,b), then there exists a point $c \in (a,b)$ such that*

$$f(b) - f(a) = f(c)'(b-a)$$

Proof. (1) If the maximum value and minimum value occur at the end point, then $f(x)$ is a constant and therefore $f'(x) = 0$. If maximum value or minimum value occur inside the interval, then $f' = 0$ at which f takes extreme values [Lemma 3.3.6]. (2)Construct $h(x) = [f(b) - f(a)]g(x) - [g(b) - b(a)]f(x)$ and $h(a) = h(b)$ implies the existence of extreme value at (a,b) based on extreme value theorem, then $f' = 0$ at this extreme point. (3)using above theorem and set $g(x) = x$. □

Lemma 3.3.7 (The identity criterion for differentiable functions). *[2, p. 104]*

- *Let I be an open interval and suppose that the function $f : I \to \mathbb{R}$ is differentiable. Then $f : I \to \mathbb{R}$ is constant if and only if*

$$f'(x) = 0, \forall x \in I.$$

- *Let I be an open interval and suppose that the function $g : I \to \mathbb{R}$ and $h : I \to \mathbb{R}$ are differentiable. Then g and h differ by a constant if and only if*

$$g'(x) = h'(x), \forall x \in I.$$

In particular, $g = h, \forall x$ if and only if

$$g'(x) = h'(x), \forall x \in I,$$

and there exists some point $x_0 \in I$ at which

$$g(x_0) = h(x_0).$$

Proof. (1) Let $c = f(x_0)$, where x_0 is an arbitrary number in I. Then any $x \neq x_0, x \in I$, we have(using mean value theorem Theorem 3.3.1]

$$f'(c) = \frac{f(x) - f(x_0)}{x - x_0} = 0 \implies f(x) = f(x_0) = c,$$

where c is a number between x_0 and x. (2) Let $f = g - h$ and $f(x_0) = c$. And then use (1). $\square$

3.4 Function sequence and series

3.4.1 Pointwise convergence, uniform convergence

Definition 3.4.1 (sequence and series of functions).

- *A set of functions $\{f_n\}$ are called a sequence of functions if for each positive integer n, a function f_n is given.*
- *A series of functions is given by*

$$\sum_{i=1}^{\infty} f_i(x)$$

with its partial sum defined as $S_n(x) = \sum_{i=1}^{n} f_i(x)$.

Definition 3.4.2 (pointwise and uniformly convergence). *[4] Let $\{f_n\}$ be a sequence of function on a set X. Let f be a function of X.*

- *We say that $\{f_n\}$ **converges pointwise** to f on X if*

$$\lim_{n\to\infty} f_n(x) = f(x) \forall x \in X.$$

- *We say that $\{f_n\}$ **converges uniformly** to f on X if for every $\epsilon > 0$, there exists a positive integer N such that if $n \geq N$, then*

$$\left| f_n(x) - f(x) \right| < \epsilon, \forall x \in X.$$

Remark 3.4.1. Note that the key difference between point-wise and uniform is in point-wise, N depends on x and ϵ, whereas in uniform, N only depends on ϵ. **Note that N cannot be ∞ (If N is ∞, then no $n > N$ can be found based on the definition of infinity.), otherwise every pointwise convergence will be uniform convergence**

Example 3.4.1. The geometric series $\sum_{i=1}^{\infty} x^n$ converges pointwise for $-1 < x < 1$. It does not uniformly converge for $-1 < x < 1$, because as $x \to 1, N \to \infty$ to converge.

Example 3.4.2. The sequence $f_n = x^n$ converge pointwise to 0 at $-1 < x < 1$. It is not uniform convergence because given $\epsilon > 0$, we require $N > \frac{\log \epsilon}{\log x} \to \infty$ as $x \to 1$.

Definition 3.4.3 (Cauchy condition). *Let $\{f_n\}$ be a sequence of function on a set X. Let f be a function of X. We say that $\{f_n\}$ converges uniformly to f on X if and only if for every $\epsilon > 0$, there exists a N, such that if $m, n > N$, we have*[4]

$$\left|f_n(x) - f_m(x)\right| < \epsilon, \forall x \in X$$

Theorem 3.4.1 (Weierstrass M-test). *[8, p. 416] Let $\sum_{i=1}^{\infty} u_n(x), x \in E$ be given. If there is a convergent series of constants $\sum_{i=1}^{n} M_n$ such that*

$$\left|u_n(x)\right| \leq M_n, \forall x \in E$$

then the series $\sum_{n=1}^{\infty} u_n(x)$ converges absolutely for all $x \in E$ and is uniformly convergent in E.

Proof: First it is easy to show that $\sum_{n=1}^{\infty} u_n(x)$ converges absolutely pointwise using the comparison test. Let $S(x)$ be the limit, $R_n(x) = \left|S_n(x) - S(x)\right|$, our goal is to show $R_n(x)$ will converge uniformly to zero. Note that $R_n(x) \leq M_{n+1} + M_{n+2} + ...$ Because $\sum_{i=1}^{\infty} M_i$ converge, then for any $\epsilon > 0$, there must exist an N(independent of x) such that the residue of $\sum_{i=1}^{\infty} M_i$ will be smaller than ϵ(use the partial sum of $\sum_{i=1}^{\infty} M_i$ to prove). Then there exists an N, $R_n(x) < \epsilon, \forall n > N$.

Example 3.4.3. $\sum_{n=1}^{\infty} cos(nx)/2^n$ will converge unfirmly by using the M series $\sum_{i=1}^{\infty} 1/2^n$

3.4.2 Properties of uniform convergence

3.4.2.1 *Uniform convergence preserve continuity*

Theorem 3.4.2. *Let $\{f_n\}$ a sequence of functions which converges uniformly to f on a metric space M. If each f_n is continuous at $a \in M$, then f is continuous at a.*[4]

Proof. Use triangle inequality

$$\left|f(x) - f(a)\right| \leq \left|f(x) - f_N(x)\right| + \left|f_N(x) - f_N(a)\right| + \left|f_N(a) - f(a)\right|.$$

□

Remark 3.4.2. The implication is exchange of limits:

$$\lim_{x\to a}\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty}\lim_{x\to a} f_n(x)$$

3.4.2.2 *Exchange limits and integration*

Theorem 3.4.3 (integration term by term). *[8, p. 421]A uniformly convergent series of continuous functions $\sum_{i=1}^{\infty} u_i(x)$ can be integrated term by term; that is, if each $u_i(x)$ is continuous for $a < x < b$, and $\sum_{i=1}^{\infty} u_i(x)$ converges uniformly to $f(x)$, then*

$$\int_a^b f(x)dx = \int_a^b u_1(x)dx + \int_a^b u_2(x)dx + ...$$

Proof. Let S_n be the partial sum. Because of uniform convergence, for any $\epsilon > 0$, we can find N, such that

$$\left|S_n(x) - f(x)\right| < \epsilon/(b-a), \forall x$$

Then $\left|\int_a^b f(x) - \int_a^b S_n(x)dx\right| < \epsilon/(b-a) \times (b-a) = \epsilon$, therefore

$$\int_a^b f(x)dx \to \int_a^b S(x)dx.$$

□

3.4.2.3 *Exchange limits and differential*

Theorem 3.4.4 (differential term by term). *A convergent series can be differentiated term by term, provided that the functions of the series have continuous derivatives and that the series of derivatives is uniformly convergent; that is, if $u_n'(x)$ is continuous, if the series $\sum_{i=1}^{\infty} u_i(x)$ converge to $f(x)$, and if the series $\sum_{i=1}^{\infty} u_i'(x)$ convergent uniformly for $a < x < b$, then*

$$f'(x) = \sum_{i=1}^{\infty} u_i'(x)$$

3.4.2.4 *Linearity of uniform convergence*

Theorem 3.4.5. *If $\sum_{i=1}^{\infty} u_i(x)$ and $\sum_{i=1}^{\infty} v_i(x)$ are uniformly convergent for $x \in [a,b]$, and $g(x), h(x)$ are continuous for $x \in [a,b]$, then the series*

1. $\sum_{i=1}^{\infty} u_i(x) + v_i(x)$
2. $\sum_{i=1}^{\infty} h(x)u_i(x)$
3. $\sum_{i=1}^{\infty} h(x)u_i(x) + g(x)v_i(x)$

are uniformly convergent.

Proof. (1)Let S_n, T_n be the partial sums, then

$$|S_n + T_n - S - T| \leq |S_n - S| + |T_n - T| \to 0$$

(2) $h(x)$ is bounded. Let $h(x) \leq M$, then

$$\big|h(x)S_n(x) - h(x)S(x)\big| \leq M\big|S_n(x) - S(x)\big| \to 0$$

(3) directly from (1) and (2). $\square$

3.5 Power series

3.5.1 Fundamentals

Definition 3.5.1 (power series). *Let t be a fixed real number. A power series expanded about t is an infinite series of the form*

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

Most often, we refers to

$$\sum_{n=0}^{\infty} a_n x^n$$

as a power series.

Definition 3.5.2 (domain of convergence for a power series). *[2, p. 255] Given a sequence of real numbers $\{c_k\}$ indexed by the nonnegative integers, we define the* ***domain of convergence*** *of the series $\sum_{k=0}^{\infty} c_k x^k$ be the set of all numbers x such that the series $\sum_{k=0}^{\infty} c_k x^k$ converges.*

Denote the domain of convergence by D. We can define a function $f : D \to \mathbb{R}$ by

$$f(x) \triangleq \lim_{n\to\infty} \sum_{k=1}^{n} c_k x^k = \sum_{k=1}^{\infty} c_k x^k.$$

Theorem 3.5.1 (radius of convergence and uniform convergence). *[8, p. 423]*

- *Every power series*

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

  has a radius of convergence of r^* such that the series converges absolutely when $|x-a| < r^*$ and diverges when $|x-a| > r^*$.
- *If $\sum_{n=0}^{\infty} c_n (x-a)^n$ converges absolutely at x_0, then $\sum_{n=0}^{\infty} c_n (x-a)^n$ converges.*
- *If $r^* = 0$, then the series only converge for $x = a$; If $r^* = \infty$, then the series converge everywhere. If $0 < r^* < \infty$, for any $0 < r_1 < r^*$, then the series* ***converge uniformly*** *for $|x-a| \leq r_1$.*

Proof. (1)Let $M_n = |c_n| r_1^n$, then based on definition of convergence radius the series

$$\sum_{n=1}^{\infty} M_n = \sum_{n=1}^{\infty} |c_n(x_1 - a)^n|, x_1 = a + r_1$$

will converge. Because

$$c_n(x-a)^n \leq |c_n| r_1^n, \forall (x-a) \leq r_1$$

therefore the series $\sum_{n=0}^{\infty} c_n(x-a)^n$ converge uniformly based on M-test. (2) Directly from the property of absolute convergence implies convergence [Theorem 1.7.1]. □

Remark 3.5.1 (Why we care about power series). We care about power series is because they usually have uniformly convergence, and therefore have many good properties, e.g., integration term-by-term,....

Lemma 3.5.1 (calculation of radius of convergence). *[8, p. 423] For the power series $\sum_{n=0}^{\infty} c_n(x-a)^n$ the radius of convergence of r^* can be calcuated as*

$$r^* = \lim_{n\to\infty} 1/|c_{n+1}/c_n|$$

or

$$r^* = \lim_{n\to\infty} 1/|c_n|^{1/n}$$

Proof. use root test and ratio test. □

Corollary 3.5.1.1. *[2, p. 256]*

- *The power series*

$$\sum_{k=0}^{\infty} k! x^k$$

has convergence of radius being 0; that is, the power series does not converge for any non-zero x.

- *The power series*

$$\sum_{k=0}^{\infty} \frac{x^k}{k!}$$

has convergence of radius being ∞; that is, the power series converges for any non-zero x.

3.5.2 Term-by-term operation

Theorem 3.5.2 (uniform convergent power series term-by-term operation properties). *[8, p. 424]A power series with radius of convergence r^* has the following properties:*

- *A power series represents a continuous function $f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n in |x-a| < r^*$.*
- *A power series can be integrated term by term, i.e.*

$$\int_{x_1}^{x_2} \sum_{n=0}^{\infty} c_n (x-a)^n$$

  for $a - r^* < x_1 < x_2 < a + r^*$.
- *A power series can be differentiated term by term,i.e., $f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n, |x-a| < r^*$ then $f(x) = \sum_{n=0}^{\infty} n c_n (x-a)^{n-1}, |x-a| < r^*$*

Proof. directly from the consequence of the uniformly convergence within radius of convergence. Note that when differentiated, the radius of convergence will not change. □

Example 3.5.1 (application of term by term differentiation). From

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n, |x| < 1$$

we can differentiate both sides, and get

$$\frac{1}{(1-x)^2} = \sum_{n=0}^{\infty} (n+1) x^n, |x| < 1.$$

3.5.3 Power series and analytic function

Definition 3.5.3 (analytical function). *[9, p. 285]* ***Analytical functions*** *are* ***infinitely differentiable*** *functions such that the Taylor series at any point x_0 in its domain*

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

converges to $f(x)$ in a neighborhood of x_0 pointwise.

Example 3.5.2. Common analytical functions are:

- polynomials
- trigonometric functions
- exponential functions
- logarithm and power functions

Example 3.5.3. Common non-analytical functions are:

- absolute value function
- piecewise functions (due to the meeting point of different pieces)

Example 3.5.4 (A nonanalytic, infinitely differentiable function). [2, p. 221] Define

$$f(x) = \begin{cases} \exp(-1/x^2), if \ x \neq 0 \\ 0, if \ x = 0. \end{cases}$$

Then the function $f : \mathbb{R} \to \mathbb{R}$ has derivatives of all orders. However, the only at which

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

is at $x = 0$.

3.5.4 Approximation by polynomials

Lemma 3.5.2 (Approximate finite set of data). *[9, p. 297]Given a set of data $\{(x_1, a_1), (x_2, a_2), ..., (x_k, a_k)\}$(assuming the data are consistent, i.e., no same x with different a), then there exist a polynomial of degree $k - 1$ that passing all the points. This polynomial is given as*

$$P(x) = \sum_{i=1}^{k} a_i Q_i(x)$$

where

$$Q_i(x) = \frac{\prod_{j \neq i}(x - x_j)}{\prod_{j \neq i}(x_i - x_j)}$$

This polynomial is known as Lagrange polynomial.

Proof. Notice that $Q_i(x_j) = \delta_{ij}$ □

Remark 3.5.2 (implications).

- The fitting error is 0 at existing data points; However, the polynomial can have terrible prediction error at new x. For example, given a data set of size m generated by linear functions. A $m-1$ polynomial can achieve perfect fit.
- If the degree number of polynomial chosen is smaller than the data number, then perfect fit might not be achieved, but might have better prediction. This is an example of bias-variance trade-off.

Lemma 3.5.3 (Approximation accuracy). *[10, p. 135] Let $x_0, x_1, ..., x_n$ be distinct real numbers, and let f be a given real-valued function with $n+1$ continuous derivatives on $H(t, x_0, ..., x_n)$(denotes the smallest interval containing $t, x_0, ..., x_n$). Then there exists $\xi \in I$ such that*

$$f(t) - \sum_{i=1}^{n} f(x_i)Q_i(t) = \frac{(t-x_0)\cdots(t-x_n)}{(n+1)!} f^{(n+1)}(\xi).$$

Theorem 3.5.3 (Weierstrass Approximation Theorem). *[9, p. 301]Let f be any continuous function on a compact interval $[a, b]$. Then there exists a sequence of polynomials converging* ***uniformly*** to *f on $[a, b]$.*

3.6 Taylor polynomial and Taylor series

3.6.1 Taylor polynomial and approximation

Definition 3.6.1 (order of contact of two functions). *Let I be a neighborhood of the point x_0. Two functions $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ are said to **have contact of order o** at x_0 provided that $f(x_0) = g(x_0)$.*

*In particular, the function f and g are said to **have contact of order** n at x_0 if*

$$f^{(k)}(x_0) = g^{(k)}(x_0), \forall 0 \leq k \leq n.$$

Definition 3.6.2 (Taylor polynomial). *A nth degree **Taylor polynomial** P_n for f at c is given as*

$$P_n(x) = f(c) + f^{(1)}(c)(x-c) + \frac{f^{(2)}}{2!}(x-c)^2 + ... + \frac{f^{(n)}}{n!}(x-c)^n.$$

Lemma 3.6.1 (existence of unique Taylor polynomial with contact). *[2, p. 200] Let I be an open interval and let n be a nonnegative integer and suppose that the function $f : I \to \mathbb{R}$ has n derivatives.*

Then there is a unique polynomial of degree at most n that has contact of order n with f. This polynomial is given by

$$p_n(x) = f(x_0) + f'(x_0)(x-x_0) + \cdots + \frac{f^{(k)}}{k!}(x-x_0)^k + \cdots \frac{f^{(n)}}{n!}(x-x_0)^n.$$

Theorem 3.6.1 (Lagrange Remainder Theorem, Taylor's theorem). *[4, p. 186][2, p. 203] Let I be an open interval and let n be a nonnegative integer and suppose that the function $f : I \to \mathbb{R}$ has $n+1$ derivatives.*

- *consider a special f such that at the point x_0 in I,*

$$f^{(k)}(x_0) = 0, \forall 0 \leq k \leq n.$$

Then for each point $x \neq x_0$ in I, there is a point c strictly between x and x_0 which

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

- *For general f, for each point $x \neq x_0$ in I, there is a point c strictly between x and x_0 such that*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}.$$

Proof. (1) Define $g(x) = (x-x_0)^{n+1}$ such that $g(x_0) = g^{(1)}(x_0) = g^{(2)}(x_0)... = g^{(n)}(x_0) = 0$. Based on Cauchy mean value theorem [Theorem 3.3.1], we know that there exists a x_1 between x_0 and x such that

$$\frac{f(x)-f(x_0)}{g(x)-g(x_0)} = \frac{f(x)}{g(x)} = \frac{f^{(1)}(x_1)}{g^{(1)}(x_1)}.$$

Similarly, use the mean value theorem again, there exists a x_2 between x_0 and x_1 such that

$$\frac{f^{(1)}(x_1)-f^{(1)}(x_0)}{g^{(1)}(x_1)-g^{(1)}(x_0)} = \frac{f^{(1)}(x_1)}{g^{(1)}(x_1)} = \frac{f^{(2)}(x_2)}{g^{(2)}(x_2)}.$$

Continue this process, we will get

$$\frac{f(x)}{g(x)} = \frac{f^{(n+1)}(x_{n+1})}{g^{(n+1)}(x_{n+1})} = \frac{f^{(n+1)}(x_{n+1})}{(n+1)!}.$$

and finally

$$f(x) = g(x)\frac{f^{(n+1)}(x_{n+1})}{(n+1)!}$$

(2) For general f, construct the function $F = f - p_n$ and use (1).

□

Theorem 3.6.2 (Cauchy integral Remainder Theorem). *[2, p. 216] Let I be an open interval and let n be a nonnegative integer and suppose that the function $f : I \to \mathbb{R}$ has $n+1$ derivatives. Then for each point $x \neq x_0$ in I, there is a point c strictly between x and x_0 such that*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + \frac{f^{(n+1)}}{(n+1)!}(x-x_0)^{n+1}.$$

Proof. By the first fundamental theorem [Theorem 3.7.5],we have

$$f(x) = f(x_0) + \int_{x_0}^{x} f'(t)dt.$$

Integrating by parts, we have

$$\begin{aligned}\int_{x_0}^{x} f'(t)dt &= -\int_{x_0}^{x} f'(t)\frac{d}{dt}(x-t)dt \\ &= -f'(t)(x-t)|_{t=x_0}^{t=x} + \int_{x_0}^{x} f''(t)(x-t)dt \\ &= f'(x_0)(x-x_0) + \int_{x_0}^{x} f''(t)(x-t)dt.\end{aligned}$$

More generally, we have

$$\begin{aligned}&\frac{1}{k!}\int_{x_0}^{x} f^{(k+1)}(t)(x-t)^k dt \\ &= -\frac{1}{(k+1)!}\int_{x_0}^{x} f^{(k+1)}(t)\frac{d}{dt}(x-t)^{k+1}dt \\ &= \frac{1}{(k+1)!}f^{(k+1)}(x_0)(x-x_0)^{k+1} + \frac{1}{(k+1)!}\int_{x_0}^{x} f^{(k+2)}(t)\frac{d}{dt}(x-t)^{k+1}dt.\end{aligned}$$

□

3.6.2 Taylor series and Taylor's theorem

Definition 3.6.3 (Taylor series). *[8, p. 428] We denote*

$$f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n, |x-a| < r^*$$

as the Taylor series *of $f(x)$ at $x = a$ if the coefficients x_n are given by the rule*

$$c_0 = f(a), c_1 = f'(a), c_2 = \frac{f''(a)}{2!}...$$

Lemma 3.6.2 (convergent power series is its own Taylor series). *Every power series with nonzero convergence radius is the Taylor series of its sum.*

Proof. Because we can differentiate term by term, we can verify

$$c_0 = f(a), c_1 = f'(a), c_2 = \frac{f''(a)}{2!}...$$

□

Lemma 3.6.3 (uniqueness of coefficients). *[8, p. 429] If two power series*

$$\sum_{n=0}^{\infty} c_n(x-a)^n, \sum_{n=0}^{\infty} C_n(x-a)^n$$

have non-zero convergence radii and have equal sums wherever both series converge, then

$$c_n = C_n, n = 0, 1, 2, ...$$

Proof. By assumption

$$\sum_{n=0}^{\infty} c_n(x-a)^n = \sum_{n=0}^{\infty} C_n(x-a)^n = f(x)$$

then

$$f^{(n)}(a)/n! = c_n = C_n$$

□

Theorem 3.6.3 (convergence of Taylor series). *Let I be a neighborhood of the point x_0 and suppose that the function $f : I \to \mathbb{R}$ has derivatives of all orders. Suppose also that there are positive numbers r and M such that the interval $[x_0 - r, x_0 + r]$ contained in I and that*

$$\left|f^{(n)}(x)\right| \leq M^n, \forall x \in [x_0 - r, x_0 + r].$$

Then

- *The Taylor polynomial*

$$\lim_{n\to\infty} p_n(x) = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k$$

will converge absolutely.

- $$\lim_{n\to\infty} p_n(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k = f(x), , \forall x \in [x_0 - r, x_0 + r].$$

Proof. (1) Note that

$$\left|\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k\right| \leq \frac{M^k r^k}{k!},$$

and the series $\sum_{k=1}^{\infty}(Mr)^k/k!$ converges [Lemma 1.7.1].Then based on comparison test [Theorem 1.7.3], the series $sum_{k=0}^{\infty}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k$ will converge absolutely.

(2) To show that the limit is $f(x)$, from Lagrange Remainder Theorem [Theorem 3.6.1], we have

$$\left|f(x)-p_n(x)\right| = \left|\frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}\right| \leq \frac{(Mr)^{n+1}}{(n+1)!},$$

where c is between x and x_0. Note that for all $x \in [x_0-r, x_0+r]$, we can make the right hand side as small as we want when n is sufficiently large. Therefore, $\lim_{n\to\infty} p_n(x) = f(x)$. □

Example 3.6.1. The Taylor expansion of e^x and $\cos(x)$ are given by

$$e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}, \cos(x) = \sum_{k=0}^{\infty}\frac{(-1)^k x^k}{(2k)!}x^{2k},$$

with convergence domain being the whole $\mathbb{R}$. This is because for $x \in \mathbb{R}$, $f^{(k)}(x_0 = 0) \leq 1$.

3.6.3 Common Taylor series

Lemma 3.6.4 (common Taylor series).

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + ... \qquad R = 1$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + ... \qquad R = \infty$$

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} + ... \qquad R = \infty$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + ... \qquad R = \infty$$

$$\tan^{-1} x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)} = x - \frac{x^3}{3} + \frac{x^5}{5} + ... \qquad R = \infty$$

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ... \qquad R = 1$$

$$(1+x)^k = \sum_{n=0}^{\infty} \binom{k}{n} x^n = 1 + kx + \frac{k(k-1)}{2!} x^2 + \frac{k(k-1)(k-2)}{3!} + ... \qquad R = 1$$

$$\sqrt{1+x} = 1 + \frac{1}{2}x + \frac{x^2}{2!}(-\frac{1}{4}) + \frac{x^3}{3!}(-\frac{3}{8}) + ... \qquad R = 1$$

Remark 3.6.1 (How convergent radius are determined?). Note that the Taylor series in nature is power series. The convergence radius can be determined using ratio test for power series on

$$\frac{|a_{n+1}|}{|a_n|},$$

as given by Lemma 3.5.1.

Remark 3.6.2 (binomial series vs binomial theorem).

- If k is a positive integer, then binomial series reduce to binomial theorem, since all later terms $n > k$ are zero.
- If k is a positive integer, then the convergence radius is ∞.

Corollary 3.6.3.1.

- $$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + ...$$
- $$\lim_{n\to\infty} \frac{x^n}{n!} = 0, \forall x \in \mathbb{R}$$

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

Proof. (2) because e^x is convergent for all $x \in \mathbb{R}$. From Theorem 1.7.1 , we know that if the $\sum_{n=1}^{\infty} a_n$ converges, then $\lim_{n\to\infty} a_n = 0$. □

Example 3.6.2 (some conversion techniques).

- $$\frac{1}{\sqrt{4-x}} = \frac{1}{\sqrt{4(1-\frac{x}{4})}} = \frac{1}{2}(1-\frac{x}{4})^{-1/2}$$
- $$e^{-x^2} = \sum_{n=0}^{\infty} \frac{(-x^2)^n}{n!}$$
- Expand $\sqrt{x}$ at x_0 as
$$\sqrt{x - x_0 + x_0} = \sqrt{x_0}\sqrt{1 + \frac{x - x_0}{x_0}}$$

3.6.4 Useful approximations

Theorem 3.6.4. *Let $S \subset \mathbb{R}^n, s \in \mathbb{R}^n$, and suppose that $f : S \to \mathbb{R}$ is continuously differentiable and $g = \nabla f$ is Lipschitz continuous at x with Lipschitz constant $\gamma(x)$ for some appropriate vector norm. It follows that if the segment $x + \theta s \in S$ for all $\theta \in [0,1]$, then*
$$\left| f(x+s) - m^L(x+s) \right| \leq \frac{1}{2}\gamma(x)\|s\|^2$$
where $m^L(x+s) = f(x) + g(x)^T s$

Proof.

$$\begin{aligned} \left| f(x+s) - m^L(x+s) \right| &= \left| f(x+s) - f(x) - g(x)^T s \right| \\ &= \left| g(x')^T s - g(x)^T s \right| \leq \gamma(x) \left| x' - x \right| \end{aligned}$$

□

Theorem 3.6.5. *Let $S \subset \mathbb{R}^n, s \in \mathbb{R}^n$, and suppose that $f : S \to \mathbb{R}$ is twice continuously differentiable and $H = \nabla^2 f$ is Lipschitz continuous at x with Lipschitz constant $\gamma(x)$ for some appropriate vector norm. It follows that if the segment $x + \theta s \in S$ for all $\theta \in [0, 1]$, then*

$$\left| f(x+s) - m^Q(x+s) \right| \leq \frac{1}{6}\gamma(x)\|s\|^3$$

where $m^Q(x+s) = f(x) + g(x)^T s + \frac{1}{2} s^T H s$

3.7 Riemann Integral Theory

3.7.1 Construction of Riemann integral

Definition 3.7.1 (partition, upeer sums, and lower sums). *[1, p. 218]*

- *Let $[a,b]$ be a closed interval. A **partition** P of $[a,b]$ is a finite set $\{x_0, x_1, ..., x_n\}$ such that*
$$a = x_0 < x_1 ... < x_n = b.$$
- *Let f be a bounded function on $[a,b]$, let α be an **increasing** function on $[a,b]$, and let P be the partition. The **upper sum** of f with respect to α for the partition P is*
$$U(f,P) = \sum_{i=1}^{n} M_i(x_i - x_{i-1}), M_i = \inf\{f(x)|x \in [x_{i-1}, x_i]\}.$$
- *The **lower sum** is*
$$L(f,P) = \sum_{i=1}^{n} m_i(x_i - x_{i-1}), m_i = \sup\{f(x)|x \in [x_{i-1}, x_i]\}.$$

Definition 3.7.2. *A partition Q is a **refinement** of a partition P if Q contains all of the points P: that is $P \subseteq Q$.*

Definition 3.7.3 (upper integral, lower integral). *[1, p. 220] Let $\mathcal{P}$ be the collection of all possible partitions of the interval $[a,b]$. The **upper integral** of f is defined to be*
$$U(f) = \inf\{U(f,\mathcal{P}), P \in \mathcal{P}\}.$$

*Similarly, the **lower integral** of f is defined to be*
$$L(f) = \inf\{L(f,\mathcal{P}), P \in \mathcal{P}\}.$$

Lemma 3.7.1 (inequality of lower and upper sums). *[1, p. 220]*

- *If $P \subseteq Q$, then $L(f,P) \leq L(f,Q)$, and $U(f,P) \geq U(f,Q)$.*
- *For any bounded function f on $[a,b]$, it is always the case that $U(f) \geq L(f)$.*

Definition 3.7.4 (Riemann integrability). *A bounded function f defined on the interval $[a,b]$ is **Riemann integrable** if $U(f) = L(f)$ in this case we define*

$$\int_a^b f(x) = U(f) = L(f).$$

Theorem 3.7.1 (integrability criterion). *[1, p. 221] A bounded function f is integrable on $[a,b]$ if and only if for every $\epsilon > 0$, there exists a partition P_ϵ of $[a,b]$ such that*

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

3.7.2 Riemann integrability

3.7.2.1 *Basics*

Theorem 3.7.2 (sequential criterion for integrability). *[1, p. 223] A **bounded** function f is integrable on $[a,b]$ if and only if there exists a sequence of partitions $\{P_n\}_{n=1}^\infty$ satisfying*

$$\lim_{n\to\infty} [U(f, P_n) - L(f, P_n)] = 0,$$

and in this case

$$\int_a^b f dx = \lim_{n\to\infty} U(f, P_n) = \lim_{n\to\infty} L(f, P_n).$$

Remark 3.7.1 (how to use). [1, p. 223] We can let P_n be the partition of $[a,b]$ into n equal subintervals.

Lemma 3.7.2 (endpoints and integrability). *[1, p. 224] If $f : [a,b] \to \mathbb{R}$ is **bounded** and f is **integrable** on $[c,b]$ for all $c \in (a,b)$, then f is integrable on $[a,b]$. An analogous result holds at the other endpoints.*

Lemma 3.7.3 (continuous function on closed interval implies integrability). *[4][2, p. 156][1, p. 222] If f is continuous at interval $[a,b]$, then f is Riemann integrable.*

Proof. Because f is continuous on a closed interval,then it is **uniformly continuous**, given a ϵ, we can find fine enough partition such that the upper bound and lower upper in each partition component is small enough, such that $U(f,T) - L(f,T) < \epsilon$. $\square$

Lemma 3.7.4 (Integrability implies boundedness, necessary condition). *[11] If function f is integrable at interval $[a, b]$, then f is bounded.*

In other words, ***if a function is unbounded, then it is not integrable.***

Proof. Note that this is not **Riemann-Stieltjes** integral. To prove: suppose it is not bounded, then not matter how finer the partition will be, then there exist a component in the partition,say Δ_k, in which the function f is unbounded. For any number M, there exist a $x^* \in \Delta_k$, such that $f(x^*)\Delta_k > M$, and this make the Riemann sum $\sum f(\eta_i)\Delta_i$ always greater than any given large number. □

3.7.2.2 *Lebesgue characterization of integrability*

Definition 3.7.5 (zero measure set). *[3] A subset A of $\mathbb{R}^n$ has* ***measure 0*** *if for every $\epsilon > 0$ there is a cover $\{U_1, U_2, U_3, ...\}$ of A by closed rectangles such that $\sum_{i=0}^{\infty} v(U_i) < \epsilon$.*

Remark 3.7.2.

- A set with only finitely many points clearly has measure 0.
- If A has infinitely countable of points that arranged in sequence $a_1, a_2, a_3, ...$, then A also has measure 0, since we can use choose U_i to be a closed rectangle containing a_i with $v(U_i) < \epsilon/2^i$. Then $\sum_{i=0}^{\infty} \epsilon/2^i = \epsilon$.
- If $A = A_1 \cup A_2 \cup A_3 ...$ and each A_i has measure 0, then A has measure 0.

Theorem 3.7.3 (bounded function with zero-measure discontinuous points implies integrability, Lebesgue's theorem). *[1, p. 242] Let f be a* ***bounded*** *function defined on the interval $[a, b]$. Then f is Riemann-integrable if and only if the sets of points where f is not continuous has measure zero.*

Proof. by properly partition,i.e., let partition endpoint aligned with discontinuity. □

3.7.2.3 *limits and integrability*

Note 3.7.1 (limits and Riemann integrability). [12, p. 1] In general, the limit of Riemann integrable functions is not necessarily Riemann-integrable.

For example, let $\{x_n\}$ be the sequence of ordered rational number between 0 and 1. Then the function

$$D_n = \sum_{k=1}^{n} E_n(x)dx, E_n(x) = \begin{cases} 1, x = x_n \\ 0, x \neq x_n \end{cases}$$

Consider the integral

$$\int_0^1 D_n(x)dx.$$

- If n is any finite number, then $\int_0^1 D_n(x)dx = 0$ is Riemann integrable.
- As $n \to \infty$, $\int_0^1 D_\infty(x)dx$ is not Riemann integrable. Note that $D_\infty(x)$ is equivalent to the Dirichlet function defined by

  $$D(x) = \begin{cases} 1, x \in \mathbb{Q} \\ 0, x \in \mathbb{R}/\mathbb{Q} \end{cases}.$$

  Because no matter how fine the partition is, in each interval there will be a rational number and an irrational number since they are dense [Theorem 1.3.1]; therefore the upper sum will be 1 and the lower sum will be 0 for all possible partitions.
- That is, in general

  $$\lim_{n\to\infty} \int_X f_n(x)dx \neq \int_X \lim_{n\to\infty} f_n(x)dx.$$

3.7.2.4 *Algebraic properties*

Theorem 3.7.4 (algebraic properties of Riemann integral). *[1, p. 229] Assume f and g are integrable functions on the interval $[a,b]$. It follows that*

- *The function $f+g$ are integrable on $[a,b]$ with*

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

- *For $k \in \mathbb{R}$, the function kf is integrable with*

  $$\int_a^b kf = k\int_a^b f.$$

- *If $m \leq f(x) \leq M$ on $[a,b]$, then*

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

- *If $f(x) \leq g(x)$ on $[a,b]$, then*

$$\int_a^b f \leq \int_a^b g.$$

- *The function $|f|$ is integrable and*

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

3.7.3 First Fundamental Theorem of Calculus

Theorem 3.7.5 (First Fundamental Theorem of Calculus, integrating derivatives). *[2, p. 161]Let the function $F : [a,b] \to \mathbb{R}$ be continuous on the closed interval $[a,b]$ and be differentiable on the open interval (a,b). Moreover, suppose that its derivative $F' : (a,b) \to \mathbb{R}$ is both continuous and bounded. Then*

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

Remark 3.7.3 (When not to use). [2, p. 163] Define

$$f(x) = \begin{cases} 4, if\ 2 \leq x \leq 3 \\ 0, if\ 3 \leq x \leq 6 \end{cases}.$$

The function f is integrable; however, we cannot apply First Fundamental Theorem to calculate the integral $\int_2^4 f(x)dx$ because $f(x)$ is not continuous on $[2,4]$. Note that we can use it to calculate the integral on the continuous segment.

3.7.4 Second Fundamental Theorem of Calculus

3.7.4.1 *Fundamentals*

Lemma 3.7.5 (Continuity and differentiability of integral functions). *[2, p. 169] Suppose that the function $f : [a,b] \to \mathbb{R}$ is integrable. Define*

$$F(x) = \int_a^x f(t)dt, \forall x \in [a,b].$$

Then

- *the function $F : [a,b] \to \mathbb{R}$ is continuous.*
- *F is differentiable.*

Proof. (1)Let $u, v \in [a,b], u < v$. Then

$$F(v) = F(u) + \int_u^v f(t)dt,$$

such that

$$F(v) - F(u) = \int_u^v f(t)dt.$$

Because $f(x)$ is continuous on $[a,b]$, then it has a minimum value $f(x_m) = m$ and a maximum value $f(x_M) = M$ [Theorem 3.1.4,Theorem 3.1.5]. We have

$$m(v-u) \leq \int_u^v f(x)dx \leq M(v-u),$$

Let $K = \max(|m|, |M|)$, then

$$-K(v-u) \leq \int_u^v f(x)dx \leq K(v-u);$$

therefore

$$-K(v-u) \leq F(v) - F(u) \leq K(v-u) \implies \left|F(v) - F(u)\right| \leq K|v-u|.$$

When $v \to u$, we have $F(v) \to F(u)$;that is, $F(x)$ is continuous. (2) Directly from second fundamental theorem of calculus. □

Theorem 3.7.6 (second fundamental theorem of calculus, differentiate integral). *[2, p. 168] Suppose that the function $f : [a,b] \to \mathbb{R}$ is continuous. Then*

$$\frac{d}{dx}[\int_a^x f(t)dt] = f(x), \forall x \in (a,b).$$

If F is defined as

$$F(t) = \int_a^t f(x)dx$$

where f is continuous on an open interval I, and a is a point inside the interval, then

$$F'(t) = f(t)$$

Proof. Define

$$F(x) \triangleq \int_a^x f(t)dt, \forall x \in [a,b].$$

Then from Lemma 3.7.5, we know that $F(x)$ is continuous.

Let $x_0 \in (a,b)$, we have

$$\int_{x_0}^x f(t)dt = F(x) - F(x_0) = f(c(x))(x - x_0),$$

where $c(x)$ is between x_0 and x, and we use mean value theorem for integrals [Theorem 3.7.7]. Therefore,

$$F'(x_0) = \lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0} = \lim_{x \to x_0} f(c(x)) = f(x_0),$$

where we use the fact that $f(x)$ is continuous. □

Lemma 3.7.6 (basic generalization of second fundamental theorem). *[2, p. 170]*

- *Suppose that the function $f : [a,b] \to \mathbb{R}$ is continuous. Then*

$$\frac{d}{dx} \int_x^b f(t)dt = -f(x), \forall x \in (a,b).$$

- *Let I be an open interval and suppose that the function $f : I \to \mathbb{R}$ is continuous. Fix a point x_0 in I. Then*

$$\frac{d}{dx}[\int_{x_0}^x f(t)dt] = f(x), \forall x \in I.$$

Proof. (1) Note that

$$\int_a^b f(t)dt = \int_a^x f(t)dt + \int_x^b f(t)dt.$$

Differentiating both sides and we will get the result. (2) Let $[x_0, b]$ be the closed interval such that $[x_0, b] \subset I, x \in [x_0, b]$. Then we use second fundamental theorem Theorem 3.7.6. □

3.7.4.2 *Differentiating definite integrals*

Lemma 3.7.7. *[2, p. 171] Let I be an open interval on $\mathbb{R}$ and suppose that the function $f : I \to \mathbb{R}$ is continuous. Let J be an open interval on $\mathbb{R}$ and suppose that the function $\psi : J \to \mathbb{R}$ is differentiable and that $\psi(J) \subseteq I$. Fix a point x_0 in I. Then*

- $$\frac{d}{dx}[\int_{x_0}^{\psi(x)} f(t)dt] = f(\psi(x))\psi'(x), \forall x \in J.$$
- $$\frac{d}{dx}[\int_{\psi(x)}^{x_0} f(t)dt] = -f(\psi(x))\psi'(x), \forall x \in J.$$
- *Let K be an open interval on $\mathbb{R}$ and suppose that the function $\psi : K \to \mathbb{R}$ is differentiable and that $\phi(K) \subseteq I$. Further assume $\psi(x)$ Then*

$$\frac{d}{dx}[\int_{\phi(x)}^{\psi(x)} f(t)dt] = f(\psi(x))\psi'(x) - f(\phi(x))\phi'(x), \forall x \in J.$$

Proof. (1) Define

$$G(x) = \int_{x_0}^{\psi(x)} f(t)dt,$$

and

$$F(x) = \int_{x_0}^{x} f(t)dt.$$

Then

$$G = F \circ \psi,$$

and

$$G'(x) = F'(\psi(x))\psi'(x) = f(\psi(x))\psi'(x).$$

(2) Define

$$G(x) = \int_{\psi(x)}^{x_0} f(t)dt,$$

and

$$F(x) = \int_{x}^{x_0} f(t)dt.$$

Then

$$G = F \circ \psi,$$

and

$$G'(x) = F'(\psi(x))\psi'(x) = -f(\psi(x))\psi'(x).$$

(3) Let $\phi(x) \leq x_0 \leq \psi(x)$. Then

$$\int_{\phi(x)}^{\psi(x)} f(t)dt = \int_{\phi(x)}^{x_0} f(t)dt + \int_{\phi(x)}^{x_0} f(t)dt.$$

Then we use (1) and (2). □

Lemma 3.7.8 (exchange derivative and integral). *Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a continuous function such that partial derivative $\frac{\partial f}{\partial t}(x,t)$ exists and is continuous in both x, t. Then*

$$\frac{d}{dt}(\int_a^b f(x,t)dx) = (\int_a^b \frac{d}{dt} f(x,t)dx)$$

Proof. see appendix for exchange of derivative and integral. □

Corollary 3.7.6.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that partial derivative $\frac{\partial f}{\partial t}(x,t)$ exists and is continuous in both x, t. Then*

$$\frac{d}{dt}(\int_{a(t)}^{b(t)} f(x,t)dx) = f(b(t),t)b'(t) - f(a(t),t)a'(t) + (\int_a^b \frac{d}{dt} f(x,t)dx)$$

3.7.4.3 *Application to differential equation*

Lemma 3.7.9. *Let I be an open interval containing the point x_0 and suppose that the function $f : I \to \mathbb{R}$ is continuous. For any number y_0, the differential equation*

$$F'(x) = f(x), \forall x \in I; F(x_0) = y_0;$$

has a unique solution $F : I \to \mathbb{R}$ given by the formula

$$F(x) = y_0 + \int_{x_0}^{x} f(t)dt, \forall x \in I.$$

Proof. From the definition of F, we have $F(x_0) = y$. From the second fundamental theoremTheorem 3.7.6, we have

$$F'(x) = [y_0 + \int_{x_0}^{x} f(t)dt]' = f(x).$$

To prove uniqueness, suppose there is another function $G(x)$ satisfying $G' = F'$ and $G(x_0) = F(x_0)$. Then from the identity criterion Lemma 3.3.7, F and G must be the same function. □

3.7.5 Essential theorems

Theorem 3.7.7 (The mean value theorem for integrals). *[2, p. 166] Suppose that the function $f : [a, b] \to \mathbb{R}$ is continuous. Then there is a point x_0 in the interval $[a, b]$ such that*

$$f(x_0) = \frac{1}{b-a}\int_a^b f(x)dx = f(x_0).$$

Proof. Since f is continuous on the closed interval, then it is integrable. Also, since $f(x)$ is continuous on $[a, b]$, then it has a minimum value $f(x_m) = m$ and a maximum value $f(x_M) = M$ [Theorem 3.1.4,Theorem 3.1.5]. We have

$$f(x_m)(b-a) \leq \int_a^b f(x)dx \leq f(x_M)(b-a),$$

or equivalently,

$$f(x_m) \leq \frac{1}{b-a}\int_a^b f(x)dx \leq f(x_M),$$

From intermediate value theorem [Theorem 3.1.2], there exists an $x_0 \in [x_m, x_M]$ such that

$$f(x_0) = \frac{1}{b-a}\int_a^b f(x)dx = f(x_0).$$

□

Lemma 3.7.10 (integral mean value theorem, generalization). *Let $w(x)$ be nonnegative and integrable on $[a, b]$, and let $f(x)$ be continuous on $[a, b]$, then*

$$\int_a^b w(x)f(x)dx = f(\xi)\int_a^b w(x)dx,$$

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

Proof. Because $f(x)$ is continuous on a closed interval, then it has a minimum value m and a maximum value M [Theorem 3.1.4,Theorem 3.1.5]. Then

$$m\int_a^b w(x)dx \leq \int_a^b w(x)f(x)dx \leq M\int_a^b w(x)dx.$$

Now let $I = \int_a^b w(x)dx$. If $I = 0$, then we are done since

$$\int_a^b f(x)w(x)dx = f(c)I = 0,$$

where c is an arbitrary number in $[a,b]$. If $I \neq 0$, we have

$$m \leq \frac{1}{I}\int_a^b f(x)w(x)dx \leq M.$$

From intermediate value theorem [Theorem 3.1.2], there exists an $\xi \in [a,b]$ such that

$$f(\xi) = \frac{1}{I}\int_a^b f(x)w(x)dx.$$

□

3.7.6 Integration rules

Lemma 3.7.11 (integration by substitution). *[2, p. 179] Let the function $f : [d,c] \to \mathbb{R}$ be continuous. Suppose function g has a bounded continuous derivative and has inverse g^{-1}. Then*

$$\int_c^d f(x)dx = \int_{g^{-1}(c)}^{g^{-1}(d)} f(g(y))g'(y)dy$$

Proof. Define

$$H(t) = \int_c^t f(x)dx - \int_{g^{-1}(c)}^{g^{-1}(t)} f(g(y))g'(y)dy.$$

Then

$$\begin{aligned} H'(t) &= f(t) - f(g(g^{-1}(t)))g'(g^{-1}(t))(g^{-1}(t))' \\ &= f(t) - f(t)g'(g^{-1}(t))(1/g'(g^{-1}(t))) \\ &= 0 \end{aligned}$$

Note that $H(c) = 0$, therefore $H(d) = H(c) = 0$ based on the identity criterion [Lemma 3.3.7]. □

Remark 3.7.4. When we substitute x by $g(y)$(we can do the substitution since x is dummy variable), the integral domain is changed from $[c,d]$ to $g^{-1}([c,d])$.

3.7.7 Improper Riemann integrals

Remark 3.7.5. There are two types of improper integrals: infinite interval and discontinuous integrand.

Definition 3.7.6 (improper integral with infinite integration limits). *[13, p. 580]*

- *Given a $f(x)$ defined on $[a, +\infty]$, and f is integrable on any finite interval $[a, u]$. Then*
$$\int_a^\infty f(x)dx \triangleq \lim_{u\to\infty} \int_a^u f(x)dx = J$$
if the limit exists.
- *Given a $f(x)$ defined on $(-\infty, b]$, and $f(x)$ is integrable on any finite interval $[u, b]$. Then*
$$\int_{-\infty}^b f(x)dx \triangleq \lim_{u\to-\infty} \int_u^b f(x)dx = J$$
if the limit exists.
- *Given a $f(x)$ defined on $(-\infty, \infty]$, and $f(x)$ is integrable on any finite interval $[a, b]$. Then*
$$\int_{-\infty}^\infty f(x)dx \triangleq \lim_{u\to-\infty} \int_u^c f(x)dx + \lim_{u\to\infty} \int_c^u f(x)dx$$
if the both limits exist.

If the limits do not exist, we say the improper integral diverge.

Example 3.7.1. [13, p. 581]

- (diverging integral)
$$\int_1^\infty \frac{dx}{x} = \lim_{b\to\infty} \int_1^b \frac{dx}{x} = \lim_{b\to\infty} \ln b = \infty.$$
- (convergent integral)
$$\int_0^\infty \exp(-x)dx = \lim_{b\to\infty} \int_0^b \exp(-x)dx = \lim_{b\to\infty} 1 - \exp(-b) = 1.$$

Definition 3.7.7 (improper integral with infinite discontinuity). *[13, p. 583]*

- *Consider a $f(x)$ define on $[a, b)$, and f is unbounded in the neighborhood of b. Assume $f(x)$ is bounded and integrable on any finite interval $[a, u] \subseteq [a, b)$. Then*
$$\int_a^b f(x)dx \triangleq \lim_{c\to b^-} \int_a^c f(x)dx,$$
if the limit exists.

- *Consider a $f(x)$ define on $(a,b]$, and f is unbounded in the neighborhood of a. Assume $f(x)$ is bounded and integrable on any finite interval $[u,b] \subseteq (a,b]$. Then*

$$\int_a^b f(x)dx \triangleq \lim_{c \to a^+} \int_c^b f(x)dx,$$

if the limit exists.
- *Consider a $f(x)$ define on (a,b), and f is unbounded in the neighborhood of a and b. Assume $f(x)$ is bounded and integrable on any finite interval $[u,v] \subseteq (a,b)$. Then*

$$\int_a^b f(x)dx \triangleq \lim_{u \to b^-} \int_c^u f(x)dx + \lim_{u \to a^+} \int_u^c f(x)dx, c \in (a,b)$$

if both limits exist.

If the limits do not exist, we say the improper integral diverge.

Example 3.7.2. [13, p. 584]

- (diverging integral)

$$\int_0^2 \frac{dx}{x^3} = \lim_{b \to 0^+} \int_b^2 \frac{dx}{x^3} = \lim_{b \to 0^+} (-\frac{1}{8} + \frac{1}{2b^2}) = \infty.$$

- (convergent integral)

$$\int_0^1 \frac{1}{\sqrt[3]{x}} dx = \lim_{b \to 0^+} \int_0^b \frac{1}{\sqrt[3]{x}} dx = \lim_{b \to 0^+} \frac{3}{2}(1 - b^{2/3}) = \frac{3}{2}.$$

3.8 Basic measure theory

3.8.1 Measurable space

3.8.1.1 *σ algebra*

Definition 3.8.1 (σ algebra). *Given a set Ω, a σ-field, or σ-algebra is a collection $\mathcal{F}$ of subsets of Ω, with the following properties:*

1. $\emptyset \in \mathcal{F}$.
2. *if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.*
3. *if $A \in \mathcal{F}$, then $\cup_{i=0}^{\infty} A_i \in \mathcal{F}$.*

Example 3.8.1.

- The trivial σ-field is $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
- Let $\mathcal{P}(\Omega)$ denote the collection of all subsets of Ω, then $\mathcal{P}(\Omega)$ is a σ-field.
- For any σ algebra defined on Ω, it is always between $\mathcal{F}_0$ and $\mathcal{P}(\Omega)$.
- The collection $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$, where A is a fixed subset of Ω.
- The set of all the subsets of finite set Ω.
- For a finite sample space Ω, the power set of Ω is the largest σ field, $\{\emptyset, \Omega\}$ is the smallest σ field.

Lemma 3.8.1 (σ algebra is closed under countable intersection and relative complement). *Let $\mathcal{F}$ be the σ algebra be a σ algebra defined on set X. Let $A_1, A_2, ...$ be a countable sequence subsets in $\mathcal{F}$. Then*

- $$\cap_{n=1}^{\infty} A_i \in \mathcal{F}.$$
  In particular, it is closed under finite intersection; that is
  $$\cap_{n=1}^{N} A_i \in \mathcal{F}.$$
- *Let $B, C \in \mathcal{F}$ and $B \subseteq C$, then*
  $$C - B \in \mathcal{F}.$$

Proof. (1) Use DeMorgan's law [Lemma 1.1.2], we have

$$\cup_{n=1}^{\infty} A_i \in \mathcal{F} \implies \cap_{n=1}^{\infty} A_i = (\cup_{n=1}^{\infty} A_i)^C \in \mathcal{F} \in \mathcal{F}.$$

(2) Note that $C - B = C \cap (B^C)$. Since $B, C, B^C \in \mathcal{F}$, from (1) $C - B$ is in $\mathcal{F}$. □

Definition 3.8.2 (generated σ algebra from a collection of subsets). *[12, p. 6] Let X be a set. Let $\mathcal{G}$ be a collection of the subsets of X, or equivalently we write $\mathcal{G} \subseteq \mathcal{P}(X)$. Then the σ **algebra** generated by $\mathcal{G}$ is the smallest σ algebra defined on X containing $\mathcal{G}$; it is denoted by $\sigma(\mathcal{G})$.*

Remark 3.8.1 (how to generate the σ algebra from subsets)**.** Given $\mathcal{G}$ a collection of the subsets of X, we can do complement, countable union, and countable intersection until we get a σ algebra.

3.8.1.2 *Measurable space and positive measure*

Definition 3.8.3 (measurable space, measurable sets). *Let X be a set and Σ be the σ algebra defined on X. The pair (X, Σ) is called **measurable space**, the members $A \in \Sigma$ are called **measurable sets** or $\Sigma-$measurable sets. A triple (X, Σ, μ) is called measure space.*

Definition 3.8.4 (positive measure, measure space). *[12, p. 7] Given a set X with its σ field Σ, a function $\mu : \Sigma \to \bar{\mathbb{R}}$ is called a positive measure if it satisfies:*

- *Non-negativity: For all $E \in \Sigma$, $\mu(E) >= 0$*
- $\mu(\emptyset) = 0$
- *Countable additivity: For all countable collections $\{E_i\}$ of pairwise disjoint sets in Σ:*

$$\mu(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$$

*A triple (X, Σ, μ) is called **measure space**.*

Example 3.8.2 (counting measure). Let X be an arbitrary set. Let $\mathcal{F}$ be its σ algebra. Define $\mu : \mathcal{F} \to [0, \infty]$ as

$$\mu(A) = \begin{cases} number\ of\ elements\ in\ A, A \in \mathcal{F}, A\ is\ finite\ set \\ \infty, A \in \mathcal{F}, A\ is\ infinite\ set \end{cases}.$$

μ is called **counting measure**.

3.8.1.3 *Borel algebra and Lebesgue measure*

Definition 3.8.5 (Borel subsets in $\mathbb{R}$). *[14][15, p. 7]*

- *A Borel subset in $\mathbb{R}$ is a set that can be formed from other open sets in $\mathbb{R}$ (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement.*
- *Every open interval in $\mathbb{R}$ is a Borel subset of $\mathbb{R}$ because an open interval can be written as the countable union of a sequence of closed intervals.*

Note 3.8.1 (open interval close interval conversion). Using countable union and intersection properties, we can convert between open interval and close intervals, for example

- $(a, b) = \cup_{n=1}^{\infty}[a + 1/n, b - 1/n]$
- $[a, b] = \cap_{n=1}^{\infty}(a - 1/n, b]$
- (countable union)$(a, \infty) = \cup_{n=1}^{\infty}[a, a + n]$
- singleton: $\{a\} = [a, a]$

Definition 3.8.6 (Borel algebra). *The collection of all Borel subsets on $\mathbb{R}$ forms a σ-algebra, known as the Borel σ-algebra, denoted by $\mathcal{B}(\mathbb{R})$. The Borel σ algebra on $\mathbb{R}$ is the smallest σ-algebra that containing all the closed and open intervals in $\mathbb{R}$.*

Definition 3.8.7 (Lebesgue measure on $\mathbb{R}$). *[15, p. 20] Let $\mathcal{B}(\mathbb{R})$ be the σ algebra of Borel subset of $\mathbb{R}$. The Lebesgue measure on $\mathbb{R}$, which we denote by $\mathcal{L}$, assigns to each set $B \in \mathcal{B}(\mathbb{R})$ a number $[0, \infty)$ or the value ∞ such that*

- $\mathcal{L}[a, b] = b - a$, *if* $a \leq b$
- *(countable additivity) if $B_1, B_2, ...$ are disjoint sets in $\mathcal{B}(\mathbb{R})$, then*

$$\mathcal{L}(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mathcal{L}(B_n)$$

Lemma 3.8.2 (Basic properties of Lebesgue measure on $\mathbb{R}$). *For the Lebesgue measure defined on $\mathbb{R}$, we have the following properties:*

- *Singleton sets have measure of 0.*
- *Empty set has measure 0.*
- *(finite additivity) If $B_1, B_2, ..., B_n$ are disjoint sets in $\mathcal{B}(\mathbb{R})$, then*

$$\mathcal{L}(\cup_{i=1}^{n} B_i) = \sum_{n=1}^{n} \mathcal{L}(B_i).$$

- *Any countable set of real numbers has Lebesgue measure of 0.*

Proof. (1) From the definition $\{a\} = \mathcal{L}[a,a] = a - a = 0$. (2) Use countable additivity

$$\mathcal{L}(\cup\emptyset) = \mathcal{L}(\emptyset) = \sum \mathcal{L}(\emptyset).$$

Note that $\mathcal{L}(\emptyset) \geq 0$ then $\mathcal{L}(\emptyset)$ must be zero. (3) Straight forward from countable additivity. (4) Use the union of singleton sets. □

Definition 3.8.8 (Borel algebra, Lebesgue measure on $\mathbb{R}^n$). *[16]*

- *In $\mathbb{R}^n$, let $a_i < b_i (i = 1, 2, ..., n)$, the **Borel σ algebra** $\mathcal{B}(\mathbb{R}^n)$ is the smallest σ-algebra generated [Lemma 11.1.2] by **all** intervals like*

$$(a_1, b_1) \times (a_2, b_2) ... (a_n, b_n) \in \mathcal{B}(\mathbb{R}^n)$$

- *The measure $\lambda : \mathcal{B}(\mathbb{R}^n) \to \mathbb{R}$ defined as:*

$$\lambda((a_1, b_1) \times (a_2, b_2) ... (a_n, b_n)) = \prod_{i=1}^{n} (b_i - a_i)$$

*is called a **Lebesgue measure** on $\mathcal{B}(\mathbb{R}^n)$.*

Remark 3.8.2. It can be easily showed that the Lebesgue measure satisfies the requirement for a measure: (1) non-negativity;(2) $\mu(\emptyset) = 0$; (3) countable additivity.

3.8.2 Measurable functions and properties

3.8.2.1 *Measurable function and measurability*

Definition 3.8.9 (measurable functions). *[4, p. 361] A function f from a set X, with a σ algebra $\mathcal{F}$, into $[-\infty, \infty]$ is called an extended real measurable function on X with respect to $\mathcal{F}$ if $f^{-1}(V) \in \mathcal{F}$, for every open set $V \subset [-\infty, \infty]$*

Example 3.8.3 (constant value function is measurable). A constant value function map from a measurable space $(X, \mathcal{F})$ to a constant c in $\mathbb{R}$ is always measurable, since

$$f^{-1}(V) = \begin{cases} \emptyset, if\ c \notin V \in \mathcal{B}(\mathbb{R}) \\ X, if\ c \in V \in \mathcal{B}(\mathbb{R}) \end{cases} \emptyset.$$

Definition 3.8.10 (Borel measurable function). *[15, p. 21] Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function defined on $\mathbb{R}$. If for every $B \in \mathcal{B}(\mathbb{R})$, the set $f^{-1}(B)$ is also in $\mathcal{B}(\mathbb{R})$, then f is said to be* ***Borel measurable function****.*

Remark 3.8.3.

- Every continuous and piecewise continuous function is Borel measurable.
- Many Borel measurable functions are not continuous. For example, step function.
- It is usually extremely difficult to find a function that is not Borel measurable.

Theorem 3.8.1 (criteria for measurable function). *[4, p. 361]Let X be a set, let A function f from a set X, with a σ algebra $\mathcal{F}$, into $[-\infty, \infty]$. Then the following are equivalent:*

- *f is measurable*
- $\{x | f(x) > a\} \in \mathcal{F}$ *for every* $a \in \mathbb{R}$
- $\{x | f(x) < a\} \in \mathcal{F}$ *for every* $a \in \mathbb{R}$
- $\{x | f(x) \geq a\} \in \mathcal{F}$ *for every* $a \in \mathbb{R}$
- $\{x | f(x) \leq a\} \in \mathcal{F}$ *for every* $a \in \mathbb{R}$

Proof. (1) implies (2) by definition;
(2) implies (3): first the set $\{x | f(x) \geq a\} = \cap_{n=1}^{\infty} \{x | f(x) > a - 1/n\}$, therefore $\{x | f(x) \geq a\} \in \mathcal{F}$, and therefore

$$\{x | f(x) < a\} = X - \{x | f(x) \geq a\} \in \mathcal{F}$$

other cases can be prove using similar techniques. □

3.8.2.2 *Properties*

Lemma 3.8.3 (function composition and measurability). *[4, p. 362]*

- *If $F : \mathbb{R} \to \mathbb{R}$ is Borel measurable and $f : X \to \mathbb{R}$ is Borel measurable, then $F \circ f$ is measurable.*

- *If $F : \mathbb{R} \to \mathbb{R}$ is continuous and $f : X \to \mathbb{R}$ is Borel measurable, then $F \circ f$ is measurable.*

Proof. (1) Let $B \in \mathcal{B}(\mathbb{R})$, then $F^{-1}(B)$ is in $\mathcal{B}(\mathbb{R})$ since F is Borel measurable. Since f is Borel measurable $(F \circ f)^{-1}(B) = f^{-1}(F^{-1}(U)) \in \mathcal{B}(\mathbb{R})$. (2) Let U be an open set in $\mathbb{R}$, the $F^{-1}(U)$ is open since F is continuous. Since f is measurable $(F \circ f)^{-1}(U) = f^{-1}(F^{-1}(U)) \in \mathcal{F}$. We can also use the fact every continuous function is Borel measurable. □

Corollary 3.8.1.1. *[4, p. 362] If $f : X \to \mathbb{R}$ is measurable, then the function $cf, f^2, |f|, 1/f$(if $f \neq 0, \forall x \in X$) are measurable.*

Proof. use continuous function $F = cx, x^2$... and function composition. □

Lemma 3.8.4 (algebraic properties of Borel measurable function). *[4, pp. 362–364] If $f, g : (X, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ are real-valued measurable functions and $k \in \mathbb{R}$, then*

- *kf is Borel measurable function.*
- *$f + g$ and $f - g$ are Borel measurable functions.*
- *f^2 is Borel measurable function.*
- *fg is Borel measurable function.*
- *$1/g, g \neq 0$ is Borel measurable function.*
- *f/g is Borel measurable function.*
- *$f^+(x) = \max(f(x), 0)$ and $f^-(x) = -\min(f(x), 0)$ are both measurable. Conversely, if f^+, f^- are measurable, so is f.*

Proof. We will emphasize the proof $f + g$ and fg. Others directly from Lemma 3.8.3. (1) If $k > 0$, then
$$\{x : kf(x) < b\} = \{x : f < b/k\} = f^{-1}((-\infty, b/k)).$$
the set $f^{-1}((-\infty, b/k))$ is in $\mathcal{F}$ because f is Borel measurable. Similarly we can prove the case $k \leq 0$. (2) Note that

$$\{x : f(x) + g(x) < b\} = \cup_{q+r<b;q,r\in\mathbb{Q}}\{x : f(x) < q\} \cap \{x : g(x) < r\}.$$

Since
$$\cup_{q+r<b;q,r\in\mathbb{Q}}(-\infty, q) \cap (-\infty, r)$$
are Borel subsets of $\mathbb{R}$, its inverse image is also in $\mathcal{F}$. Therefore, $f + g$ is Borel measurable. (3) The function f^2 is measurable since if $b > 0$,

$$\{x : f(x)^2 < b\} = \{x : -\sqrt{b} < f(x) < \sqrt{b}\}$$

and the set $(-\sqrt{b}, \sqrt{b}) \in \mathcal{B}$. (4) Note that

$$fg = \frac{1}{2}((f+g)^2 - f^2 - g^2),$$

and then use (1)(2)(3). (5) Note that if $g \neq 0$,

$$\{x : 1/g(x) < b\} = \begin{cases} \{x : 1/b < g(x) < 0\}, if b < 0 \\ \{x : -\infty < g(x) < 0\}, if b = 0 \\ \{x : -\infty < g(x) < 0\} \cup \{x : 1/b < g(x) < \infty\}, if b > 0 \end{cases}.$$

(6) Use (4)(5). (7)(a) For $B \subset [0, \infty]$, if B does not includes $\{0\}$, $(f^+)^{-1}(B) = f^{-1}(B) \in \mathcal{F}$; if if B includes $\{0\}$, $(f^+)^{-1}(B) = f^{-1}(B) \cup f^{-1}([-\infty, 0]) \in \mathcal{F}$. Therefore f^+ is measurable. Similarly we can prove the case of f^-. (b) conversely, note that $f = f^+ - f^-$. □

Lemma 3.8.5 (measurability and limits). *Let $f_1, f_2, ...$ be a sequence of measurable $\bar{\mathbb{R}}$-valued functions. Then the functions*

$$\sup_n f_n(x), \inf_n f_n(x), \limsup_n f_n(x), \liminf_n f_n(x),$$

are all measurable.

Proof. (1) Denote $g(x) = \sup_n f_n(x)$. Then if there exist an n such that $f_n(x) > c$, we have $g(x) > c$. Therefore,

$$\{x | g(x) > c\} = [\cap_{n=1}^{\infty} \{x : f(x) \leq c\}]^C = \cup_{n=1}^{\infty} \{x : f_n(x) > c\}.$$

Since each $\{x : f_n(x) > c\}$ is measurable, its countable union is also measurable. (2) Similar to (1).

□

3.8.3 Convergence of measurable functions

3.8.4 Almost everywhere convergence

Definition 3.8.11 (Almost everywhere convergence of Borel measurable function). *[15, p. 24] Let $f_1, f_2, f_3, \ldots$ be a sequence of real-valued, Borel-measurable functions defined on $\mathbb{R}$. Let f be another real-valued, Borel-measurable functions defined on $\mathbb{R}$. We say that $f_1, f_2, \ldots$ converges to f almost everywhere and write*

$$\lim_{n\to\infty} f_n = f \text{ almost everywhere}$$

if the set of $x \in \mathbb{R}$ for which the sequence of numbers $f_1(x), f_2(x), \ldots$ does not have limit $f(x)$ is a set with Lebesgue measure zero.

Remark 3.8.4 (interpretation).

- For every fixed $x \in \mathbb{R}$, we can evaluate the limit of the sequence $f_1(x), f_2(x), \ldots$. If $\lim_{n\to\infty} f_n(x) = f(x)$, then we say $f_1(x), f_2(x), \ldots$ converges to $f(x)$ at x.
- When we say almost everywhere convergence, we mean the subset of $\mathbb{R}$ **where** $\lim_{n\to\infty} f_n(x) \neq f(x)$ **has measure of zero.**

Example 3.8.4. [15, p. 24] Consider a sequence of normal density functions

$$f_n(x) = \sqrt{\frac{n}{2\pi}} \exp(-\frac{nx^2}{2}).$$

We have

- If $x \neq 0$, then $\lim_{n\to\infty} f_n(x) = 0$.
- If $x = 0$, then $\lim_{n\to\infty} f_n(0) = \lim_{n\to\infty} \sqrt{\frac{n}{2\pi}} = \infty$.

Therefore, the sequence $f_1, f_2, \ldots$ converge almost everywhere to the function $g(x) = 0, \forall x \in \mathbb{R}$. Only at $x = 0$, , which has zero Lebesgue measure, the sequence $f_1(0), f_2(0), \ldots$ not converges to $g(0)$.

3.9 Lebesgue integral

3.9.1 Simple function and its Lebesgue integral

Definition 3.9.1 (simple function). *[4, p. 364]*

- *A **simple function** is a function taking finite number of values.*
- *Let (X, Σ) be a measurable space. A **simple measurable function** is a function $f : X \to \mathbb{R}$:*

$$f(x) = \sum_{k=1}^{n} a_k I_{A_k}(x),$$

*where $A_1, A_2, ...A_n$ are disjoint subsets in Σ that **partitions** X.*

Remark 3.9.1 (representation of simple functions on X)**.**

- A simple function is a finite linear combination of indicator functions of measurable sets.
- When we represent a simple function by $f(x) = \sum_{k=1}^{n} a_k I_{A_k}(x)$, we require $A_1, A_2, ..., A_n$ to partition X; otherwise for some $x \in X$, its mapping $f(x)$ is not defined.

Lemma 3.9.1 (different representations of simple functions). *[12, p. 16] Let f be a simple function mapping from X to $\mathbb{R}$. Let $A_1, A_2, ..., A_N$ be a partition of X such that $f(x) = \sum_{k=1}^{N} a_k I_{A_k}(x)$. It follows that*

- *(coarsest representation) Suppose f will take Q different values $q_1, q_2, ..., q_Q$. Then the collection of subsets $f^{-1}(q_1), f^{-1}(q_2), ..., f^{-1}(q_Q)$ form the coarsest partition, denoted by P_0.*
- *Any partition P used to represent f must be obtained from subdividing subsets in P_0; that is, $P \subseteq P_0$. Any partition $P \subseteq P_0$ is called **compatible partition**.*
- *Let $B_1, B_2, ..., B_M$ be an arbitrary partition(not necessarily compatible), then*

$$f(x) = \sum_{i=1}^{N} \sum_{j=1}^{M} a_i I_{A_i \cap B_j}(x).$$

Note that $\{A_i \cap B_j, i = 1, ..., N, j = 1, ..., M\}$ is also an compatible partition.

- *Let $B_1, B_2, ..., B_M$ be another compatible partition of X, then $f(x) = \sum_{k=1}^M b_k I_{B_k}(x)$. where $a_i = b_j$ if $A_i \cap B_j \neq \emptyset$. That is, f can also be represented by*

$$f(x) = \sum_{i=1}^N \sum_{j=1}^M a_i I_{A_i \cap B_j}(x) = \sum_{i=1}^N \sum_{j=1}^M b_j I_{A_i \cap B_j}(x).$$

Proof. (1)(2) obviously. (3) Note that sets $A_i \cap B_j, j = 1, ..., M$ is obtained from subdividing A_i. (4)

$$f(x) = \sum_{i=1}^N a_i I_{A_i}(x) = \sum_{i=1}^N \sum_{j=1}^M a_i I_{A_i \cap B_j}(x) = \sum_{i=1}^N \sum_{j=1}^M b_j I_{A_i \cap B_j}(x) = \sum_{j=1}^M b_i I_{B_j}(x).$$

Note that we use the fact that $a_i = b_j$ must hold on the common set $A_i \cap B_j$;otherwise f will map an element to different values. □

Definition 3.9.2 (Lebesgue integral of simple functions). *[4, p. 368][12, p. 16] Let s be a non-negative simple measurable function in X, $s = \sum_{i=1}^n c_i I_{A_i}$, where $A_i \in \Sigma$, and partitions X. We define the integral of s over X with respect to a measure μ as:*

$$\int_X s d\mu = \sum_{i=1}^n c_i \mu(A_i)$$

Theorem 3.9.1 (algebraic properties of Lebesgue integral of simple functions).

- *If $f, g : X \to [0, \infty)$ are measurable simple functions and $\alpha in X$. Then $\alpha f, f + g$ are measurable simple functions and*

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu, \int_X f + g d\mu = \int_X f d\mu + \int_X g d\mu.$$

- *If $f(x) \leq g(x)$ for each $x \in X$ we have*

$$\int_X f d\mu \leq \int_X g d\mu.$$

Proof. (1) The scaling is straight forward. For the sum rule, we have

$$\begin{aligned}
\int_X f+g d\mu &= \int_X \sum_{i=1}^N a_i I_{A_i} + \sum_{j=1}^M b_j I_{B_j} d\mu \\
&= \int_X \sum_{i=1}^N \sum_{j=1}^M a_i I_{A_i \cap B_j} + \sum_{i=1}^N \sum_{j=1}^M b_j I_{A_i \cap B_j} d\mu \\
&= \int_X \sum_{i=1}^N \sum_{j=1}^M a_i I_{A_i \cap B_j} + b_j I_{A_i \cap B_j} d\mu \\
&= \sum_{i=1}^N \sum_{j=1}^M a_i \mu(A_i \cap B_j) + b_j \mu(A_i \cap B_j) \\
&= \sum_{i=1}^N a_i \mu(A_i) + \sum_{j=1}^M b_j \mu(B_j) \\
&= \int_X f d\mu + \int_X g d\mu
\end{aligned}$$

where we use results from Lemma 3.9.1. (2) Note that $f-g$ is a simple function only takes non-positive values. Therefore

$$\int_X f d\mu - \int_X g d\mu = \int_X (f-g) d\mu \leq 0.$$

□

3.9.2 Lebesgue integral of measurable functions

3.9.2.1 *Integral of non-negative functions*

Theorem 3.9.2 (any non-negative measurable functions as the limit of simple functions). *If f is any measurable function from X into $[0,\infty]$, then there exists a sequence $\{s_n\}$ simple functions such that [4]*

$$0 \leq s_1 \leq s_2 \leq s_3 \leq ...$$

and

$$\lim_{n\to\infty} s_n(x) = f(x), \forall x \in X.$$

Note 3.9.1 (notations of Lebesgue integral)**.** The notation for Lebesgue integral is given by

$$\int_{x\in X} f(x)d\mu, \int_{x\in X} f(x)d\mu, \int_{x\in X} f(x)d\mu(x), \int_X f(x)\mu(dx)$$

Definition 3.9.3 (Lebesgue integral of non-negative functions). *Let f be a non-negative extended real-valued function on X. We define the integral of f over X with respect to a measure μ to be*

$$\int_X f d\mu = \sup\{\int_X s d\mu | s \text{ is a simple measurable function on } X, 0 \leq s \leq f\}.$$

*If the integral is a finite value, we say f **is Lebesgue integrable**.*

Remark 3.9.2 (Lebesgue integral of non-negative functions are well-defined)**.**

- Note that for any non-negative simple function, its Lebesgue integral is wells defined.
- sup operation on a set of real numbers will produce a unique number.

Theorem 3.9.3 (monotone convergence theorem). *[12, p. 25]Let $(X, \mathcal{F}, \mu)$ be a measure space. Let $f_n : X \to [0, \infty]$ be non-negative measurable functions increasing pointwise to f. Then*

$$\int_X f(x)d\mu = \int_X (\lim_{n\to\infty} f_n(x))d\mu = \lim_{n\to\infty} \int_X f_n(x)d\mu$$

Remark 3.9.3 (limit of infinity value)**.** note that for any x, $f_1(x) \leq f_2(x) \leq f_3(x) \leq ....$ If $\lim_{n\to\infty} f_n(x) = \infty$. If for a nonzero measure set A we have $f(x) = \infty, \forall x \in A$, then $\int_X f d\mu = \infty$

Remark 3.9.4 (interpretation of limits)**.** Note that $f(x) = \lim_{n\to\infty} f_n(x) \in [0, \infty]$ is guaranteed to exist since the sequence $\{f_n(x)\}$ is monotone increasing sequence. Moreover, $f(x)$ is measurable via Lemma 3.8.5 since for monotone sequence, taking sup is equivalent to taking sequential limits [1.5.1].

Theorem 3.9.4 (algebraic properties of Lebesgue integral of non-negative measurable functions). *[4, p. 374]*

- *If $f, g : X \to [0,\infty)$ are measurable simple functions and $\alpha in X$. Then $\alpha f, f+g$ are measurable simple functions and*

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu, \int_X f + g d\mu = \int_X f d\mu + \int_X g d\mu.$$

- *If $f(x) \leq g(x)$ for each $x \in X$ we have*

$$\int_X f d\mu \leq \int_X g d\mu.$$

3.9.2.2 *Integral of general functions*

Definition 3.9.4 (Lebesgue integral of general functions). *[4, p. 369] Let f be a measurable function on $(X, \mathcal{M})$. Define $f^+(x) = \max(f(x), 0)$ and $f^-(x) = -\min(f(x), 0)$(both are non-negative and measurable [Lemma 3.8.4]). It follows that*

- *If **at least one** of the numbers $\int_X f^+ d\mu$ and $\int_X f^- d\mu$ is finite, then **the integral of** f **over** X **with respect to** μ is given by*

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

- *If **both** of the numbers $\int_X f^+ d\mu$ and $\int_X f^- d\mu$ are finite, then f is said to be **Lebesgue integrable**. The set of Lebesgue integrable functions are denoted by $\mathcal{L}(X, \mathcal{M}, \mu)$.*
- *If **both** of the numbers $\int_X f^+ d\mu$ and $\int_X f^- d\mu$ are ∞, then the integral of f is **not defined**.*

Note 3.9.2 (existence of Lebesgue integral vs. Lebesgue integrability)**.**

- For a general measurable function, the Lebesgue integral might not exist.
- If the Lebesgue integral exists and it is finite, then it is Lebesgue integrable.

Theorem 3.9.5 (algebraic properties of Lebesgue integral of general measurable functions). *[4, p. 376]*

- *If $f, g : X \to [0,\infty)$ are measurable simple functions and $\alpha in X$. Then $\alpha f, f+g$ are measurable simple functions and*

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu, \int_X f + g d\mu = \int_X f d\mu + \int_X g d\mu.$$

- *If $f(x) \leq g(x)$ for each $x \in X$ we have*

$$\int_X f d\mu \leq \int_X g d\mu.$$

3.9.3 Riemann vs. Lebesgue integrals

Lemma 3.9.2. *If a function is Riemann integrable then it is Lebesgue integrable.*

Let f be a bounded function defined $\mathbb{R}$, let $a < b$ be numbers.[15]

1. The Riemann integral $\int_a^b f(x)dx$ is defined (i.e. the lower sum and upper Riemann sums converge to the same limit) if and only if the set of points $x \in [a, b]$ where $f(x)$ is not continuous has Lebesgue measure zero.
2. If the Riemann integral $\int_a^b f(x)dx$ is defined, then f is Borel measurable (so the Lebesgue integral $\int_{[a,b]} f(x)d\mathcal{L}(x)$ is also defined), and the Rieman and Lebesgue integral agree.

Example 3.9.1 (integration of Dirichlet function). [12, p. 1] Define Dirichlet function $D : [0, 1] \to \{0, 1\}$as

$$D(x) = \begin{cases} 1, x \in \mathbb{Q} \\ 0, x \in \mathbb{R}/\mathbb{Q} \end{cases}.$$

For the integral

$$\int_0^1 D(x)dx,$$

we have

- Riemann integral does not exists. Because no matter how fine the partition is, in each interval there will be a rational number and an irrational number since they are dense [Theorem 1.3.1]; therefore the upper sum will be 1 and the lower sum will be 0 for all possible partitions.
- Lebesgue integral will give integral value of 1 since we can decompose as

$$\int_0^1 D(x)d\mu(x) = \int_{[0,1]\cap Q} D(x)d\mu(x) + \int_{[0,1]-Q} D(x)d\mu(x) = 0 + 1 = 1.$$

3.9.4 Convergence theorems

Theorem 3.9.6 (Fatou's lemma). *[12, p. 27]Let $(X, \mathcal{F}, \mu)$ be a measure space. Let $f_n : X \to [0, \infty]$ be non-negative measurable functions increasing pointwise to f. Then*

$$\int_X f(x)d\mu = \int_X (\lim_{n\to\infty} f_n(x))d\mu = \lim_{n\to\infty} \int_X f_n(x)d\mu$$

Theorem 3.9.7 (dominated convergence theorem). *[12, p. 27]Let $(X, \mathcal{F}, \mu)$ be a measure space. Let $f_n : X \to [0, \infty]$ be non-negative measurable functions increasing pointwise to f. Then*

$$\int_X f(x)d\mu = \int_X (\lim_{n\to\infty} f_n(x))d\mu = \lim_{n\to\infty} \int_X f_n(x)d\mu$$

Example 3.9.2 (application of dominated convergence theorem). Compute the following integral

$$\lim_{n\to\infty} \int_{\mathbb{R}} \frac{n\sin(x/n)}{x(x^2+1)}dx.$$

Solution:
Define $f_n(x) = \frac{n\sin(x/n)}{x(x^2+1)}$. Then

$$\begin{aligned} |f_n(x)| &= \left|\frac{n\sin(x/n)}{x(x^2+1)}\right| \\ &= \left|\frac{\sin(x/n)}{x/n}\frac{1}{1+x^2}\right| \\ &\leq \frac{1}{1+x^2} = g(x). \end{aligned}$$

Therefore, we have

$$\lim_{n\to\infty} \int_{\mathbb{R}} \frac{n\sin(x/n)}{x(x^2+1)}dx = \int_{\mathbb{R}} \lim_{n\to\infty} \frac{n\sin(x/n)}{x(x^2+1)}dx = \int_{\mathbb{R}} \frac{1}{1+x^2} = \pi.$$

Example 3.9.3 (when dominated convergence condition is not satisfied). Define a sequence of functions $\{f_n\}$ via

$$f_n(x) = nI_{(0,1/n]} = \begin{cases} n, if\ 0 < x \leq 1/n \\ 0, else \end{cases}.$$

There exists no function g such that $|f_n(x)| \leq g(x)$; therefore the dominated convergence theorem condition is not satisfied. Then it is easy to see that

$$1 = \lim_{n\to\infty} \int_0^1 f_n(x)dx.$$

However,

$$\int_0^1 \lim_{n\to\infty} f_n(x)dx = \int_0^1 0dx = 0.$$

where we use the fact that $f_n(x)$ converges to 0 pointwise.

3.9.4.1 *Applications*

Theorem 3.9.8 (exchange summation and integral). *[12, p. 30]*

- *Let $f_n : X \to [0, \infty]$ be **non-negative** function. Then*

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

- *Let $f_n : X \to \mathbb{R}$ be measurable function, with $\int_X \sum_{n=1}^{\infty} |f_n(x)| dx = \sum_{n=1}^{\infty} \int_X |f_n(x)| dx$ being finite. Then*

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

Proof. (1)Let $g_n(x) = \sum_{i=1}^{n} f_n(x)$. Then $g_n(x)$ is monotone increasing sequence, then it has limit (even though the limit is ∞). Then from monotone convergence theorem [Theorem 3.9.3], we have

$$\lim_{n\to\infty} \int g_n = \int \lim_{n\to\infty} g_n.$$

(2)Let $g_n(x) = \sum_{i=1}^{n} f_n(x)$, then $|g_n(x)| \leq H(x) = \sum_{n=1}^{\infty} \int_X |f_n(x)| dx$. Then from dominated convergence theorem [Theorem 3.9.7], we have

$$\lim_{n\to\infty} \int g_n = \int \lim_{n\to\infty} g_n.$$

□

Theorem 3.9.9 (exchange summation in double series). *[4, pp. 373, 94]*

- *Let $f_n : X \to [0, \infty]$ be **non-negative** function. Then*

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

- *Let $f_n : X \to \mathbb{R}$ be measurable function, with $\int_X \sum_{n=1}^{\infty} |f_n(x)|\, dx = \sum_{n=1}^{\infty} \int_X |f_n(x)|\, dx$ being finite. Then*

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

Proof. (1)Let $g_n(x) = \sum_{i=1}^{n} f_n(x)$. Then $g_n(x)$ is monotone increasing sequence, then it has limit (even though the limit is ∞). Then from monotone convergence theorem [Theorem 3.9.3], we have

$$\lim_{n\to\infty} \int g_n = \int \lim_{n\to\infty} g_n.$$

(2)Let $g_n(x) = \sum_{i=1}^{n} f_n(x)$, then $|g_n(x)| \leq H(x) = \sum_{n=1}^{\infty} \int_X |f_n(x)|\, dx$. Then from dominated convergence theorem [Theorem 3.9.7], we have

$$\lim_{n\to\infty} \int g_n = \int \lim_{n\to\infty} g_n.$$

□

Theorem 3.9.10 (differentiation under the integral sign). *[12, p. 35] Let $(X, \mathcal{F}, \mu)$ be a measurable space. Let T be an open set of $\mathbb{R}^n$, and $f : X \times T \to \mathbb{R}$, with $f(\cdot, t)$ being measurable for each $t \in T$. Then*

$$F(t) = \int_X f(x,t)dx,$$

is differentiable with the derivative

$$F'(t) = \frac{d}{dt} \int_X f(x,t)dx = \int_X \frac{\partial}{\partial t} f(x,t)dx$$

provided the following condition satisfied

- *for each $x \in X$, $\frac{\partial}{\partial t} f(x,t)$ exists for all $t \in T$.*
- *there is an integrable function g such that $\left|\frac{\partial}{\partial t} f(x,t)\right| \leq g(x)$ for all $t \in T$.*

Proof. Let $\{h_n\}$ be a sequence converging to 0. We have

$$\begin{aligned}
F'(t) &= \lim_{n\to\infty} \frac{F(t+h_n)-F(t)}{h_n} \\
&= \lim_{n\to\infty} \int_X \frac{f(x,t+h_n)-f(x,t)}{h_n} dx \\
&= \int_X \lim_{n\to\infty} \frac{f(x,t+h_n)-f(x,t)}{h_n} dx \\
&= \int_X \frac{\partial}{\partial t} f(x,t) dx
\end{aligned}$$

where we use Theorem 3.9.7 to justify the exchange of limit and integral in the following way:

$$\left| \frac{f(x,t+h_n)-f(x,t)}{h_n} \right| = \left| \frac{\partial}{\partial t} f(x,z) \right| \le g(x)$$

□

3.10 Notes on bibliography

The key references for this chapter are intermediate level real analysis textbooks[4][1][17][18].

A quite friendly introduction to measure theory and Lebesgue integral can be found in [19].

BIBLIOGRAPHY

1. Abbott, S. *Understanding analysis* (Springer, 2001).
2. Fitzpatrick, P. *Advanced calculus* (American Mathematical Soc., 2006).
3. Spivak, M. *Calculus on Manifolds* (WA Benjamin New York, 1965).
4. Johnsonbaugh, R. & Pfaffenberger, W. Foundations of mathematical analysis (2010).
5. Kreyszig, E. *Differential Geometry* ISBN: 9780486318622 (Dover Publications, 2013).
6. Krim, H. & Hamza, A. B. *Geometric methods in signal and image analysis* (Cambridge University Press, 2015).
7. Carter, M. *Foundations of mathematical economics* (MIT Press, 2001).
8. Kaplan, W. *Advanced calculus* (Addison Wesley Publishing Company, 1973).
9. Strichartz, R. S. *The way of analysis* (Jones & Bartlett Learning, 2000).
10. Atkinson, K. E. *An introduction to numerical analysis* (John Wiley & Sons, 1989).
11. Normal university, X. *mathematical analysis (Chinese)* (higher education press, 2001).
12. Cheng, S. *A crash course on the Lebesgue integral and measure theory* 2008.
13. Larson, R. & Edwards, B. H. *Calculus, 9th Edition* (Cengage Learning, 2009).
14. Wikipedia. *Borel set — Wikipedia, The Free Encyclopedia* [Online; accessed 18-May-2016]. 2016.
15. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
16. Fries, C. *Mathematical finance: theory, modeling, implementation* (John Wiley & Sons, 2007).
17. Thomson, B., Bruckner, J. & Bruckner, A. *Elementary Real Analysis* ISBN: 9780130190758 (Prentice-Hall, 2001).
18. Duren, P. L. *Invitation to classical analysis* (American Mathematical Soc., 2012).
19. Nelson, G. S. *A user-friendly introduction to Lebesgue measure and integration* (American Mathematical Soc., 2015).