MATH20410 (W25): Analysis in Rn II (accelerated)

Lecturer: Joe Jackson
Notes by: Aden Chen

Friday 7th March, 2025

Contents

1 Single-Variable Differential Calculus

In this chapter, we consider mainly functions of the form $f : I \to \mathbb{R}$, where I is an interval, e.g., (a, b), $[a, b]$, $(a, b]$, (a, ∞), $\mathbb{R}$. This is the function we have in mind unless otherwise stated.

Definition 1.1 (Differentiability)**.** We say f is **differentiable at** $x \in I$ if the limit

$$f'(x) \coloneqq \lim_{t \to x} \frac{f(t) - f(x)}{t - x} = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

exists. In this case, we call $f'(x)$ the derivative of f at x. Moreover:

- We say that f is **differentiable** if $f'(x)$ exists for each $x \in I$.
- We say f is **continuously differentiable** ($f \in C^1$) if $f' : I \to \mathbb{R}$ is continuous.

Example 1.2.

- $f(x) = |x|$. Differentiable on $\mathbb{R} \setminus \{0\}$.
- $f(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$. Continuous but not differentiable at 0.
- $f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$. Differentiable everywhere (in particular at 0), but $f \notin C^1$.

Proposition 1.3 (Rules for computing derivatives)**.**

(i) Linearity. $(af + bg)' = af' + bg'$ *(if f' and g' exist, such requirements are hereafter omitted).*

(ii) Product rule. $(fg)' = f'g + fg'$.

(iii) Quotient rule. $(f/g)' = (f'g - fg')/g^2$.[1]

(iv) Chain rule. $(f \circ g)' = (f' \circ g) \cdot g'$.

[1] Low dhigh minus high dlow. Not Haidilao...

Proof. We prove the quotient rule; the remaining are left as exercises. Starting from the definition

$$\left(\frac{f}{g}\right)'(x) = \lim_{t\to x}\frac{\frac{f}{g}(t) - \frac{f}{g}(x)}{t-x}$$
$$= \lim_{t\to x}\frac{\frac{f(t)}{f(t)} + \frac{f(x)}{g(t)} - \frac{f(x)}{g(t)} + \frac{f(x)}{g(x)}}{t-x}.$$

Note that

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

and we have

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g^2(x)}$$

□

Theorem 1.4. *If f is differentiable at x then f is continuous at x.*

Proof. Note that

$$\lim_{t\to x} f(t) - f(x) = \lim_{t\to x}\frac{f(t)-f(x)}{t-x}(t-x) = f'(x)\cdot 0 = 0.$$

□

1.1 The Mean Value Theorem

Lemma 1.5. *Suppose $f : [a,b] \to \mathbb{R}$ has a local maximum or minimum at $x \in (a,b)$. If $f'(x)$ exists, then $f'(x) = 0$.*

Proof. From the definition of the derivative, consider the limits from the left and right; one is non-positive and the other is non-negative. □

Theorem 1.6 (Rolle's Theorem)**.** *Suppose $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$, differentiable on (a,b), and such that $f(a) = f(b)$. Then there exists $x \in (a,b)$ such that $f'(x) = 0$.*

Proof. Consider the global maximum or minimum (exist since f is a continuous function defined on a compact set) and apply the previous lemma. (If both the maximum and minimum is at a or b, f is constant.) □

Theorem 1.7 (Mean Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be such that f is continuous on $[a, b]$ and differentiable on (a, b). Then there exists $x \in (a, b)$ such that $f(b) - f(a) = f'(x)(b - a)$.*

Proof. Apply Rolle's to $\tilde{f} = f - [f(b) - f(a)] \cdot \frac{x-a}{b-a}$. □

Theorem 1.8. *Let $f : (a, b) \to \mathbb{R}$ be differentiable.*

(a) if $f' = 0$, then f is constant.

(b) if $f' \geq 0$, then f is increasing.

(c) if $f' \leq 0$, then f is decreasing.

Proof. Apply the mean value theorem. □

Theorem 1.9 (The Intermediate Value Property of Derivatives)**.** *Let $f : [a, b] \to \mathbb{R}$ be differentiable[2] and suppose $f'(a) < \lambda < f'(b)$ Then there exists $x \in (a, b)$ such that $f'(x) = \lambda$.*

[2] f need not be C^1!

Proof *(à la Pugh)***.** Slide a small secant of length so small that the slope around a and b is separated also by λ. By continuity of the slope, there exists a secant between a and b with slope λ. Apply the mean value theorem to this slope. □

Proof *(à la Joe/Rudin)***.** We start with $\lambda = 0$. Then $f'(a), f'(b) \neq 0$ and the global min/max of f cannot be at the endpoints. At the global extrema we have the desired result. When $\lambda \neq 0$, consider $\tilde{f} \coloneqq f - \lambda x$. □

Example 1.10. Consider

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

We have

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

which has the intermediate value property.

Theorem 1.11 (Generalized Mean Value Theorem)**.** *Let $f, g : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $x \in (a, b)$ such that*

$$\big(f(a) - f(b)\big)g'(x) = \big(g(a) - g(b)\big)f'(x).$$

Remark 1.12. When the above is not zero,

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

Proof. Define

$$h(t) \coloneqq \big(f(b) - f(a)\big)g(t) - \big(g(b) - g(a)\big)f(t).$$

Note that

$$h(a) = f(b)g(a) - f(a)g(b) = h(b)$$

and apply Rolle's. □

1.2 L'Hôpital's Rule

Theorem 1.13 (L'Hôpital's Rule, a particular case)**.** *Let $f, g : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If $g(x) \neq 0$ in a neighborhood of a and $f(x) = g(x) = 0$, then*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)},$$

if the last limit exists.

Proof. Consider some small $\delta > 0$. The generalized MVT gives some $x \in (a, a+\delta)$ such that

$$\frac{f(a+\delta)}{g(a+\delta)} = \frac{f'(x)}{g'(x)} \approx \lim_{t\to a} \frac{f'(t)}{g'(t)},$$

where the last approximation follows from the existence of the limit. Note that as $\delta \to 0$, $x \to a$, and the approximation error shrinks to 0. □

Refer to Rudin or something for the general case.

1.3 Higher Derivatives

If $f : I \to \mathbb{R}$ is differentiable, then we can define the second derivative $f'' \coloneqq (f')'$ if f' is differentiable. Higher derivatives can be defined similarly. We usually write $f^{(n)}$ for the n-th derivative of f.

Example 1.14. $L(x) = f(x_0) + f'(x_0)(x - x_0)$ is a (first order) linear approximation of f at x_0. How good is this approximation? A first answer is

$$f(x) = L(x) + o(|x - x_0|),$$

since we have as $x \longrightarrow x_0$ that

$$\frac{f(x) - L(x)}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \longrightarrow 0.$$

But can we say even more about the quality of the approximation? – Yes, if f is twice differentiable.

Proposition 1.15 (First-order Taylor's Theorem)**.** *Suppose f' exists and is continuous on $[a, b]$ and f'' exists on (a, b). Let $x_0, x \in [a, b]$ with $x_0 \neq x$. Then*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(y)(x - x_0)^2,$$

where y is between x_0 and x. In particular, we have

$$|f(x) - f(x_0) - f'(x_0)(x - x_0)| \leq \frac{1}{2} \sup_{y \in (a,b)} |f''(y)| \cdot |x - x_0|^2.$$

Proof. Find M such that we have

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{M}{2}(x - x_0)^2.$$

We need only find y such that $M = f''(y)$. Define

$$g(t) \coloneqq f(t) - f(x_0) - f'(x_0)(t - x_0) - \frac{M}{2}(t - x_0)^2.$$

Note that $g''(t) = f''(t) - M$, so we need only find a point at which g'' vanishes. Since $g(x_0) = g(x) = 0$, by the MVT there exists y' between x_0 and x such that $g(y') = 0$. Observe that $g'(x_0) = 0$, and so by the MVT again, there exists y between x_0 and y' (and by extension between x_0 and x) such that $g''(y) = 0$. □

The more general story: given $f : [a, b] \to \mathbb{R}$ and $x_0 \in [a, b]$, we may define

$$\begin{aligned}
P_0(x) &\coloneqq f(x_0), \\
P_1(x) &\coloneqq f(x_0) + f'(x_0)(x - x_0), \\
P_2(x) &\coloneqq f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2, \\
&\vdots \\
P_n(x) &\coloneqq \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k,
\end{aligned}$$

when the corresponding derivatives exist. Note that $P_n(x)$ is the unique degree n polynomial such that $P_n^{(k)}(x_0) = f^{(k)}(x_0)$ for $k = 1, \ldots, n$.

Theorem 1.16 (Taylor's Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be such that*

- *$f^{(k)}$ exists on $[a, b]$ for $k = 1, \ldots, n$; and*
- *$f^{(n+1)}$ exists on (a, b).*

Then, for any $x_0, x \in [a, b]$ with $x_0 \neq x$, there exists y between x_0 and x such that

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

for some y between x_0 and x.

We proof the case $n = 2$, the same idea can be used to prove the general case.

Proof. Define

$$g(t) = f(t) - P_2(t) - \frac{M}{6}(t - x_0)^3.$$

Since $g''' = f''' - M$, we need only find y such that $g'''(y) = 0$. Note that $g(x_0) = g(x) = 0$, and so by the MVT there exists y' between x_0 and x such that $g'(y') = 0$. Next, note that $g'(x_0) = 0$, and so by the MVT there exists y'' between x_0 and y' such that $g''(y'') = 0$. Finally, note that $g''(x_0) = 0$, and so by the MVT there exists y between x_0 and y'' such that $g'''(y) = 0$. □

2 Multivariable Differential Calculus

Some remainders about $\mathbb{R}^n$:

- $\mathbb{R}^n = \{x = (x_1, \dots, x_n) : x_i \in \mathbb{R}\}$.
- $\mathbb{R}^n$ is a vector space, with canonical basis $\{e_i, \dots, e_n\}$.
- $\mathbb{R}^n$ comes with an inner product $\langle x, y \rangle = x \cdot y = \sum x_i y_i$, a norm $|x| = \sqrt{x \cdot x} = (\sum x_i y_i)^{1/2}$, and a metric $d(x, y) = |x - y|$.

2.1 Higher Dimensional Codomains

Consider a function $f : \mathbb{R} \supset I \to \mathbb{R}^n$.

Definition 2.1. f is differentiable at x if the limit

$$f'(x) \coloneqq \lim_{t \to x} \frac{f(t) - f(x)}{t - x}$$

exists.

Remark 2.2. We may write $f(t) = (f_1(t), \dots, f_n(t))$, and $f'(x) = (f_1'(x), \dots, f_n'(x))$, since a sequence $x \in \mathbb{R}^n$ converges if and only if each of its components converges.

Theorem 2.3. *We have the following analog of the MVT:*

$$|f(b) - f(a)| \leq |f'(t)| \cdot |b - a|.$$

for some t between a and b.

Proof. Assume $a < b$. Define

$$h(t) \coloneqq \langle f(b) - f(a), f(t) \rangle .$$

The MVT gives

$$\begin{aligned} h(b) - h(a) = h'(t)(b - a) &= \langle f(b) - f(a), f'(t) \rangle (b - a) \\ &\leq (b - a)|f(b) - f(a)||f'(t)|, \end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality. Noting that

$$h(b) - h(a) = |f(b) - f(a)|^2,$$

we have the desired result. $\square$

2.2 Higher Dimensional Domain

We next consider functions $f : U \to \mathbb{R}$, where $U \subset \mathbb{R}^n$ is open.

Definition 2.4 (Partial Derivatives)**.**

$$\frac{\partial f}{\partial x_i}(x) = \mathrm{D}_i f(x) \coloneqq \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}.$$

Definition 2.5 (Directional Derivatives)**.** Fix $u \in \mathbb{R}^n$.

$$= \mathrm{D}_i u f(x) \coloneqq \lim_{h \to 0} \frac{f(x + hu) - f(x)}{h}.$$

2.2.1 The Derivative

Intuition: A function is differentiable if a first-order Taylor expansion holds. That is, if f is "well-approximated" by a linear function.

Definition 2.6. We denote the set of all linear maps from $\mathbb{R}^n$ to $\mathbb{R}$ as $L(\mathbb{R}^n, \mathbb{R})$.

Definition 2.7 (The Derivative)**.** A function f is differentiable at x if there exists a linear map $T \in L(\mathbb{R}^n, \mathbb{R})$ such that

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

In this case we write $\mathrm{D}f(x) = T$. In other words, $f(x + h) = f(x) + \mathrm{D}f(x)(h) + o(|h|)$.

Remark 2.8.

- If f is differentiable, then

$$\mathrm{D}f : U \longrightarrow L(\mathbb{R}^n, \mathbb{R}).$$

- If is easy to check that $\mathrm{D}f$ is well defined, that is, there is at most one T such that the limit holds.

We may think of the linear map $T : \mathbb{R}^n \to \mathbb{R}$ as

$$T(u) = \langle u, v \rangle, \tag{1}$$

where $v \coloneqq (Te_1, \ldots Te_n)$.

Definition 2.9 (The Gradient)**.** If f is differentiable at x, we define $\nabla f(x) = v$, where v satisfies (1). In other words,

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - \langle \nabla f(x), h \rangle}{|h|} = 0.$$

Theorem 2.10. *If f is differentiable at x, then $\mathrm{D}_u f(x)$ exists for all $u \in \mathbb{R}^n$ and $\mathrm{D}_u f(x) = \mathrm{D} f(x) u = \langle \nabla f(x), u \rangle$.*

Proof. Note that as $t \to 0$, we have

$$\begin{aligned} \left| \frac{f(x+tu) - f(x)}{t} - \mathrm{D} f(x) u \right| &= \left| \frac{f(x+tu) - f(x) - \mathrm{D} f(x)(tu)}{t} \right| \\ &= \left| \frac{f(x+tu) - f(x) - \mathrm{D} f(x)(tu)}{|tu|} \right| \cdot |u| \longrightarrow 0. \end{aligned}$$

□

Remark 2.11. In particular we have $\mathrm{D}_i f(x) = \mathrm{D}_{e_i} f(x) = \mathrm{D} f(x) e_i = \langle \nabla f(x), e_i \rangle$. In other words, if f is differentiable, then $\nabla f(x) = (\mathrm{D}_1 f, \ldots, \mathrm{D}_n f)$.

Remark 2.12.

- Differentiability holds if and only if the gradient exists.
- Differentiability implies the existence of directional derivatives, which then implies the existence of partial derivatives. The converse implications are not true.

Example 2.13. Consider

$$f(x_1, x_2) := \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}.$$

It is easy to see that $\mathrm{D}_1 f(0) = \mathrm{D}_2 f(0) = 0$ but $\mathrm{D}_{(1,1)} f(0)$ does not exist. Indeed, f is not even continuous on the line $t(1,1)$.

Example 2.14. Consider

$$f(x_1, x_2) := \begin{cases} \frac{x_1^3}{x_1^2 + x_2^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}.$$

Note that

$$\mathrm{D}_u f(0) = \lim_{t\to 0} \frac{t^3 u_1^3}{t^2(u_1^2+u_2^2)} \cdot \frac{1}{t} = \frac{u_1^3}{u_1^2+u_2^2}.$$

However, $\mathrm{D}f(0)$ cannot exist, since the above mapping is not linear.

Theorem 2.15. *If the partial derivatives $\mathrm{D}_1 f, \ldots, \mathrm{D}_n f$ exist and are continuous (in a neighborhood of x), then f is differentiable at x.*

Proof. Fix arbitrary $x \in E$ and define $Ah = \sum \mathrm{D}_i f(x) h_i$. We write $\omega_k \coloneqq \sum_{i=1}^k h_i e_i$ for $k = 1, \ldots, n$ and $\omega_0 \coloneqq x$. Note that $\omega_n = h$. By the MVT we can find δ_k between 0 and h_k such that

$$\begin{aligned} f(x+h) - f(x) - Ah &= \sum_{k=1}^n f(x+\omega_k) - f(x+\omega_{k-1}) - \mathrm{D}_k f(x) h_k \\ &= \sum_{k=1}^n h_k [\mathrm{D}_k(x + \omega_k + \delta_i e_i) - \mathrm{D}_k f(x)], \end{aligned}$$

which by continuity of D_i is sublinear. □

2.3 Extension to Functions with Higher Dimensional Codomains

Immediate.

We have

$$\mathrm{D}f(x) \in L(\mathbb{R}^n, \mathbb{R}^m), \quad \mathbb{R}^n \ni h \longmapsto \mathrm{D}f(x) \in L(\mathbb{R}^n, \mathbb{R}^m),$$

and

$$\mathrm{D}f : \mathcal{U} \longmapsto L(\mathbb{R}^n, \mathbb{R}^m).$$

Note that we may identify $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ with a unique matrix $A = [Te_1, \ldots, Te_n]$ such that we have $Th = Ah$ for each h.

Definition 2.16. If f is differentiable at x, we can define $[\mathrm{D}f(x)] \in \mathbb{R}^{n\times m}$ to be the unique matrix such that

$$\mathrm{D}f(x)h = [\mathrm{D}f(x)]h.$$

This is called the **Jacobian matrix**, and its determinant is called the **Jacobian**. More generally, for $T \in L(\mathbb{R}^n, \mathbb{R}^m)$, we use $[T]$ to denote the corresponding matrix.

Theorem 2.17. *If* $\mathrm{D}f(x)$ *exists, so do* $\mathrm{D}_i f_j$*, and we have*

$$[Df(x)] = \left[\mathrm{D}_i f_j\right] = \left[\nabla f_1(x) \dots \nabla f_m(x)\right]^\intercal .$$

It suffices to prove the following stronger proposition:

Proposition 2.18. *The function* f *is differentiable at* x *if and only if each* f_i *is differentiable at* x*. In this case,*

$$\mathrm{D}f(x)h = (\mathrm{D}f_1 h, \dots, \mathrm{D}f_m(x)h) = (\langle \nabla f_1(x), h\rangle, \dots, \langle \nabla f_m(x), h\rangle) = [\mathrm{D}f(x)]h.$$

Proof. Suppose f_i is differentiable. Define $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ by the formula

$$Th = (\mathrm{D}f_1 h, \dots, \mathrm{D}f_m(x)h).$$

Note that

$$\frac{|f(x+h) - f(x) - Th|}{|h|} = \left(\sum \frac{|f_i(x+h) - f_i(x) - \mathrm{D}f_i(x)h|^2}{|h|}\right)^{1/2} \longrightarrow 0.$$

The other direction is left as an exercise. □

Corollary 2.19. *If* $\mathrm{D}_j f_i$ *all exist and are continuous in a neighborhood of* x*, then* f *is differentiable at* x*.*

2.4 The Chain Rule

Consider

$$\mathbb{R}^n \supset \mathcal{U} \xrightarrow{g} \mathbb{R}^m \xrightarrow{f} \mathbb{R}^k.$$

Theorem 2.20 (Chain Rule)**.** *If* g *is differentiable at* x *and* f *is differentiable at* $g(x)$*, then* $f \circ g$ *is differentiable at* x *and*

$$\mathrm{D}(f \circ g)(x) = \mathrm{D}f(g(x)) \circ \mathrm{D}g(x).$$

A formal calculation:[3] We have

$$\begin{aligned} f \circ g(x+h) &= f \circ g(x) + \mathrm{D}f(g(x))\big(g(x+h) - g(x)\big) + o\big(g(x+h) - g(x)\big) \\ &= f \circ g(x) + \mathrm{D}f(g(x))\big(\mathrm{D}g(x)h + o(|h|)\big) + o(|h|) \\ &= f \circ g(x) + \mathrm{D}f(g(x))\big(\mathrm{D}g(x)h\big) + o(|h|). \end{aligned}$$

[3] In math, "formal calculation" often means calculation that is"systematic but without rigorous justification."

Proof. For small $h \in \mathbb{R}^p$, we write

$$g(x+h) = g(x) + Bh + R_g,$$

where $B = \mathrm{D}g(x)$ and $\lim_{h\to 0} R_g/h = 0$. Similarly, we write

$$f \circ g(x+h) = f(g(x) + Bh + R_g) = f \circ g(x) + ABh + AR_g + R_f,$$

where $A = \mathrm{D}f(g(x))$ and $\lim_{h\to 0} R_f/(Bh + R_g) \to 0$. It remains to note that the last two terms are sublinear. $\square$

2.5 Continuity of the Derivative

Let $f : \mathbb{R}^n \supset \mathcal{U} \to \mathbb{R}^M$, where $\mathcal{U}$ is open. Recall that if f is differentiable, we have defined

- $\mathcal{U} \ni x \to \mathrm{D}f(x) \in L(\mathbb{R}^n, \mathbb{R}^m)$.
- $\mathcal{U} \ni x \to [\mathrm{D}f(x)] \in \mathbb{R}^{m\times n}$.
- $\mathcal{U} \ni x \to \mathrm{D}_j f_i(x) \in \mathbb{R}$, $i = 1, \dots, m$, $j = 1, \dots, n$.

Definition 2.21. For $T \in L(\mathbb{R}^n, \mathbb{R}^m)$, we define the operator norm

$$\|T\| = \sup_{|v|=1} |Tv| = \sup_{|v| \in \mathbb{R}^n \setminus \{0\}} \frac{|Tv|}{|v|}.$$

This gives rise to the standard norm induced metric: for $T, S \in L(\mathbb{R}^n, \mathbb{R}^m)$, we have

$$d(T, S) = \|T - S\|.$$

Definition 2.22. For $A \in \mathbb{R}^{m\times n}$, we define the operator norm $\|A\|_{\mathrm{op}} = \sup_{|v|} |Av|$. Thus $\|T\| = \|[A]\|_{\mathrm{op}}$.

Definition 2.23. For $A \in \mathbb{R}^{m\times n}$, we define the Frobenius norm $\|A\|_{\mathrm{F}} = \left(\sum_{i,j} A_{ij}^2\right)^{1/2}$.

Proposition 2.24. *The following statements are equivalent:*

- $x \mapsto \mathrm{D}f(x)$ *is continuous (wrt d).*
- $x \mapsto [\mathrm{D}f(x)]$ *is continuous (wrt d_{op}).*
- $x \mapsto [\mathrm{D}f(x)]$ *is continuous (wrt d_F).*
- *Each* $x \mapsto \mathrm{D}_j f_i(x)$ *is continuous.*

Definition 2.25. The function f is C^1 if the above equivalent conditions hold.

2.6 The Inverse Function Theorem

Theorem 2.26 (The Inverse Function Theorem)**.** *Let $f : \mathbb{R}^n \supset E \to \mathbb{R}^n$ be C^1, where E is open. Suppose $x_0 \in E$ and $\mathrm{D}f(x_0)$ is invertible. Then there exists a neighborhood U of x_0 such that f is a bijection from U to $V := f(U)$, and $f^{-1} : V \to U$ is C^1 with derivative $\mathrm{D}(f^{-1}(y)) = \left[\mathrm{D}f(f^{-1}(y))\right]^{-1}$.*

Remark 2.27.

- Thus if the first order Taylor expansion is invertible, then f is invertible locally.
- Consider the identities

  $$x = f^{-1}(f(x)), \quad y = f(f^{-1}(y)).$$

  Differentiating

  $$I = \mathrm{D}f^{-1}(f(x)) \circ \mathrm{D}f(x), \quad I = \mathrm{D}f(f^{-1}(y)) \circ \mathrm{D}f^{-1}(y).$$

  This shows that $\mathrm{D}(f^{-1}(y))$ and $\mathrm{D}f(f^{-1}(x))$ are inverses of each other, provided that the functions are differentiable.
- Remember the one-dimensional case! We have that $(f^{-1})' = 1/f'$:

Proof *(Inverse Function Theorem, $n = 1$)***.** Let $\mathrm{D}f(x_0) \in L(\mathbb{R}, \mathbb{R})$ be invertible. Then $f'(x_0) \neq 0$, say $f'(x_0) > 0$ without loss of generality. By continuity of f', there exists an open interval U containing x_0 such that $f' > 0$ on U. Thus f is strictly increasing and thus one-to-one on U. It is easy to verify that $V := f(U) = (f(a), f(b))$, so V is open.

Next, we show that f^{-1} is continuous. For that, consider sequence $y_k \to y$. We seek to show that $f^{-1}(y_k) \to f^{-1}(y)$. Equivalently, given $f(x_k) \to f(x)$, we show $x_k \to x$. To that end, suppose not. Then, without loss of generality, there exists infinitely many x_k such that $x_k > x + \epsilon$ for some ϵ. Thus $f(x_k) > f(x+\epsilon) > f(x)$, a contradiction.

Finally, we show that f^{-1} is differentiable. Write $x := f^{-1}(y)$ and $f^{-1}(y+h) = x+k$, that is, define $k := f^{-1}(y+h) - f^{-1}(y)$. We have then that $h = f(x+k) - f(x)$. Then as $h \to 0$, we have $\lim_{h\to 0} k = 0$, by the continuity of f^{-1}, and so

$$\frac{f^{-1}(y+h) - f^{-1}(y)}{h} = \frac{k}{f(x+h) - f(x)} \longrightarrow \frac{1}{f'(x)}.$$

□

Before the general proof, we need the following result:

Theorem 2.28 (Contraction Mapping)**.** *Let (X, d) be a complete metric space. Let $\phi : X \to X$ be a* ***contraction****, that is, there exists $c < 1$ such that*

$$d\big(\phi(x), \phi(y)\big) \leq c d(x, y).$$

Then, there is a unique fixed point of ϕ.

Proof. Pick any $x_0 \in X$. Define $x_n \coloneqq \phi(x_{n-1})$ for $n \geq 1$. Note that

$$\phi\big(x_n, x_{n-1}\big) \leq c^n \phi\big(x_1, x_0\big).$$

Thus, for $n > m$, we have

$$d\big(x_n, x_m\big) \leq \sum_{k=m+1}^{n} d\big(x_k, x_{k-1}\big) \leq d\big(x_1, x_0\big) \sum_{k=m+1}^{n} c^{k-1}.$$

Since $\sum c^j$ is a converging series, the last term tends to 0 and so (x_n) is Cauchy. Then, setting $x = \lim x_n$, we have

$$\phi(x) = \lim \phi(x_n) = \lim x_{n+1} = x.$$

Uniqueness follows from the contraction property. □

We may now proceed with the general proof of the Inverse Function Theorem. We recall first the result:

Theorem 2.29 (The Inverse Function Theorem)**.** *Let $f : \mathbb{R}^n \supset E \to \mathbb{R}^n$ be C^1, where E is open. Suppose $x_0 \in E$ and $\mathrm{D}f(x_0)$ is invertible. Then there exists a neighborhood U of x_0 such that f is a bijection from U to $V \coloneqq f(U)$, and $f^{-1} : V \to U$ is C^1 with derivative $\mathrm{D}(f^{-1}(y)) = \big[\mathrm{D}f(f^{-1}(y))\big]^{-1}$.*

Proof *(Inverse Function Theorem, the General Case)***.**

Step 1: Local Invertibility. Choose δ small enough that

- $\big\|\mathrm{D}f(x)^{-1}\big\|$ is bounded in $B_\delta(x_0)$.[4]
- $\|\mathrm{D}f(x) - \mathrm{D}f(x')\|$ is "really small" if $x, x' \in B_\delta(x_0)$.

[4] Here, we used the fact that inversion is a continuous operation.

We check that f is injective on $U \coloneqq B_\delta(x)$. Note that $f(x) = y$ if and only if $\mathrm{D}f(x_0)^{-1}\big(y - f(x)\big) = 0$, which is equivalent to x being a fixed point of the function

$$\phi_y(x) \coloneqq x + \mathrm{D}f(x_0)^{-1}\big(y - f(x)\big).$$

Thus, to prove injectivity, we need only show that ϕ_y is a contraction. Observe that

$$\mathrm{D}\phi_y(x) = I - \mathrm{D}f(x_0)^{-1}\mathrm{D}f(x) = \mathrm{D}f(x_0)^{-1}\left[\mathrm{D}f(x_0) - \mathrm{D}f(x)\right].$$

Then,

$$\left\|\mathrm{D}\phi_y(x)\right\| \leq \left\|\mathrm{D}f(x_0)^{-1}\right\|\|\mathrm{D}f(x_0) - \mathrm{D}f(x)\|$$

can be made arbitrarily small, and in particular smaller than $1/2$, by choosing δ small enough. The function ϕ_y is then a contraction. While the image of ϕ_y may not be a subset of its domain U (and so Banach contraction does not apply), the same argument in the proof of the Banach contraction theorem shows that ϕ_y has at most one fixed point, if any, in U. Injectivity of f in U thus follows.

Set $V \coloneqq f(U)$. Note that f^{-1} is well defined on V.

Step 2: V is open. Fix $f(x_0) \in V$. Pick $r > 0$ such that $B_r(x_0) \subset U$. Note that

$$|x - x_0| \leq \left\|\mathrm{D}f(x_0)^{-1}\right\||f(x) - f(x_0)|.$$

Thus for $y = f(x)$ within $r/2\left\|\mathrm{D}f(x_0)^{-1}\right\|$ of $f(x_0)$, we have $x \in U$ and so $y \in V$.

Step 3: f^{-1} is continuous (Lipschitz). Recall that $\phi_y(x)$ is a contraction in x with Lipschitz constant $1/2$, and note that it is also Lipschitz in y, with Lipschitz constant say C. From

$$x - x' = \phi_y(x) - \phi_{y'}(x') = \phi_y(x) - \phi_y(x') + \phi_y(x') - \phi_{y'}(x')$$

we thus know

$$|x - x'| \leq \frac{1}{2}|x - x'| + C|y - y'|.$$

Then,

$$\left|f^{-1}(y) - f^{-1}(y')\right| = |x - x'| \leq 2C|y - y'|$$

and f^{-1} is Lipschitz.

Step 4: The formula for $\mathrm{D}f^{-1}$. Write $y = f(x)$. Set $h = f^{-1}(y+k) - f^{-1}(y)$. Note that $f^{-1}(y+k) = x + h$ and so $k = f(x+h) - f(x)$. We have then that

$$\begin{aligned}
&\frac{\left|f^{-1}(y+k) - f^{-1}(y) - \mathrm{D}f(x)^{-1}k\right|}{|k|} \\
&= \frac{\left|h - \mathrm{D}f(x)^{-1}\big(f(x+h) - f(x)\big)\right|}{|f(x+h) - f(x)|} \\
&\leq \frac{\left\|\mathrm{D}f(x)^{-1}\right\| \|\mathrm{D}f(x)h - f(x+h) + f(x)\|}{|h|} \cdot \frac{|h|}{|f(x+h) - f(x)|}.
\end{aligned}$$

Note that the first term tends to 0 and the second is bounded. We have established then that that $\mathrm{D}f^{-1}(y) = \mathrm{D}f(x)^{-1}$ is continuous. It remains to note that as a composition of continuous functions, $\mathrm{D}f^{-1}$ is continuous. □

2.7 The Implicit Function Theorem

Example 2.30. Consider function f and the equation $f(x,y) = 0$. What does it mean to "solve for x"? We seek a function g such that $f(g(y), y) = 0$.

We will deal with the more general case of $f : \mathbb{R}^{n+m} \supset E \to \mathbb{R}^n$. If f is differentiable at (x,y), then $\mathrm{D}f(x,y) \in L(\mathbb{R}^{n+m}, \mathbb{R}^n)$. For $(h,k) \in \mathbb{R}^{n+m}$, then $\mathrm{D}f(x,y)(h,k) \in \mathbb{R}^n$. Write $\mathrm{D}_x f(x,y)h = \mathrm{D}f(x,y)(h,0)$ and $\mathrm{D}_y f(x,y)k = \mathrm{D}f(x,y)(0,k)$. Note that $\mathrm{D}_x f \in (\mathbb{R}^n, \mathbb{R}^n)$ and $\mathrm{D}_y f \in (\mathbb{R}^m, \mathbb{R}^m)$.

Theorem 2.31 (Implicit Function Theorem)**.** *Let $f : \mathbb{R}^{n+m} \supset E \to \mathbb{R}^n$. Suppose f is C^1 in a neighborhood of some point (x_0, y_0) such that $f(x_0, y_0) = 0$. If $\mathrm{D}_x f(x_0, y_0)$ is invertible, then there exists a neighborhood U of x_0 and a neighborhood V of y_0 such that for each $y \in V$, there exist a unique x such that $f(x,y) = 0$. Moreover, the function g such that $f(g(y), y) = 0$ is C^1, with $\mathrm{D}g(y) = -\mathrm{D}_x f(g(y), y)^{-1} \mathrm{D}_y f(g(y), y)$.*

Remark 2.32.

- Consider the linear map $f(x,y) = A_x x + A_y y$. The condition $f(x,y) = 0$ is equivalent to $A_x x = -A_y y$. If A_x is invertible, then we have $g(y) = -A_x^{-1} A_y y$.
- If $h(y) \coloneqq f(g(y), y) = 0$, then $\mathrm{D}h(y) = \mathrm{D}_x f(g(y), y)\mathrm{D}g(y) + \mathrm{D}_y f(g(y), y) = 0$, giving $\mathrm{D}g = -(\mathrm{D}_x f)^{-1} \mathrm{D}_y f$.

- Remember the case of $n = 1$: when the partial derivative in the direction of x is nonzero, we can solve for x locally.

Proof. Define $F : E \to \mathbb{R}^{n+m}$ by $F(x, y) = (f(x, y), y)$. The Jacobian matrix of F at (x_0, y_0) is

$$[\mathrm{D}F(x_0, y_0)] = \begin{bmatrix} \mathrm{D}_x f(x_0, y_0) & \mathrm{D}_y f(x_0, y_0) \\ 0 & I \end{bmatrix}.$$

It turns out that

$$\det \mathrm{D}F(x_0, y_0) = \det \mathrm{D}_x f(x_0, y_0) \det I - \det 0 \det \mathrm{D}_y f(x_0, y_0) = \det \mathrm{D}_x f(x_0, y_0) \neq 0.$$

By the Inverse Function Theorem, then, F is invertible in a neighborhood of (x_0, y_0). By the construction of F, there then exists G such that $(G(x, y), y) = F^{-1}(x, y)$. Define then $g(y) \coloneqq G(0, y)$. We have

$$f(g(y), y) = f(G(0, y), y) = f(F^{-1}(0, y)) = 0.$$

□

Remark 2.33 (Using the Implicit Function Theorem). Consider the function $f : \mathbb{R}^{n+m} \to \mathbb{R}^n$ with $f(a, b) = 0$. Suppose we want to solve the equation $f(x, y) = 0$ for x in terms of y. This may be thought of as solving a system of n equations in n unknowns. We seek to find $g : V \to \mathbb{R}^n$ such that $f(g(y), y) = 0$.

By the Implicit Function Theorem, such g exists if $\mathrm{D}_x f(a, b)$ is invertible (and $f \in C^1$). Intuition: if the Jacobian of f is invertible, then we change the output of f to set $f = 0$ no matter how y is changed.

Example 2.34. Consider $f : \mathbb{R}^{2+3} \to \mathbb{R}^2$ with

$$f_1 \coloneqq 2e^{x_1} + x_2 y_1 - 4y_2 + 3, \quad f_2 = x_2 \cos(x_1) - 6x_1 + 2y_1 - y_3.$$

Set $a = (0, 1)$ and $b = (3, 2, 7)$. Note that we have $f(a, b) = 0$. We have

$$\mathrm{D}_x f(x, y) = \begin{bmatrix} 2x^{x_1} & y_1 \\ -x_2 \sin(x_1) & \cos(x_1) \end{bmatrix}.$$

At (a, b),

$$\det \mathrm{D}_x f(a, b) = \det \begin{bmatrix} 2 & 3 \\ -6 & 1 \end{bmatrix} = 20 \neq 0.$$

Then

$$\mathrm{D}g(b) = -\left[\mathrm{D}_x f(a, b)\right]^{-1} \left[\mathrm{D}_y f(x, b)\right] = \begin{bmatrix} 1/4 & 1/5 & -3/20 \\ -1/2 & 6/5 & 1/10 \end{bmatrix},$$

using which we can compute the first order approximation of g.

2.8 Higher Partial Derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}$. Note that $\mathrm{D}_i f : \mathbb{R}^n \to \mathbb{R}$.

Definition 2.35. Suppose $\mathrm{D}_i f$ exists. Define $\mathrm{D}_{ji} f(x) = \mathrm{D}_j[\mathrm{D}_i f](x)$ if the latter exists.

Definition 2.36. The function f is C^2 if all $\mathrm{D}_{ji} f$ exist and are continuous.

Theorem 2.37 (Clairaut's Theorem)**.** *If f is C^2, then $\mathrm{D}_{ji} f = \mathrm{D}_{ij} f$.*

Proof *($n = 2$)***.** By the MVT, we have

$$\begin{aligned}
\mathrm{D}_{12} f(x, y) &= \lim_{h \to 0} \frac{\mathrm{D}_2(x + h, y) - \mathrm{D}_2(x, y)}{h} \\
&= \lim_{h \to 0} \lim_{k \to 0} \frac{f(x + h, y + k) - f(x + h, y) - f(x, y + k) + f(x, y)}{hk} \\
&= \lim_{h \to 0} \lim_{k \to 0} \mathrm{D}_{21} f(t, s),
\end{aligned}$$

where t is between x and $x + h$ and s is between y and $y + k$. □

2.9 Higher Derivatives: An Informal Discussion

Recall that

$$f(x + h) = f(x) + \mathrm{D} f(x) h + o(h).$$

The "total" second order derivative of $f : \mathbb{R}^n \to \mathbb{R}$ should thus satisfy

$$f(x + h) = f(x) + \mathrm{D} f(x) h + \frac{1}{2} \mathrm{D}^2 f(x)(h, h) + o(h^2).$$

Consider then $\gamma(t) = x + tv$ and $f \circ \gamma$. We have

$$\begin{aligned}
(f \circ \gamma)''(0) &= \lim_{t \to 0} \frac{\mathrm{d}}{\mathrm{d}t} \left[\sum \mathrm{D}_i f(x + tv) v_i \right] \\
&= \lim_{t \to 0} \sum \langle \nabla \mathrm{D}_i f(x + tv) v_i, v \rangle \\
&= \lim_{t \to 0} \sum_{i,j} \mathrm{D}_{ij} f(x) v_i v_j = v^\intercal \mathrm{D}^2 f(x) v,
\end{aligned}$$

where $\mathrm{D}^2 f(x)$ is the Hessian. That is,

$$\begin{aligned}
\mathrm{D}^2 f : \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\
(h, k) &\longmapsto h^\intercal \operatorname{Hess}(f)(x) k.
\end{aligned}$$

3 Integration

Let $f : [a, b] \to \mathbb{R}$ be bounded. The goal is to define $\int_a^b f(x)\,\mathrm{d}x$ if it exists.

Definition 3.1. A **Partition** P of $[a, b]$ is a collection of points $x_0, \ldots, x_n$ such that $a = x_0 < x_1 < \cdots < x_n = b$. We say P^* is a **refinement** of P if $P \subset P^*$. We say $P_1 \vee P_2 \coloneqq P_1 \cup P_2$ is the **common refinement** of P_1 and P_2. Denote as $\Pi(a, b)$ the set of partitions of $[a, b]$.

Definition 3.2. Given $P \in \Pi(a, b)$, we define the **upper sum** and **lower sum** of f with respect to P by

- $U(P, f) \coloneqq \sum_{i=1}^n \left(\sup_{x_{i-1} \le x \le x_i} f(x)\right)(x_i - x_{i-1})$.
- $L(P, f) \coloneqq \sum_{i=1}^n \left(\inf_{x_{i-1} \le x \le x_i} f(x)\right)(x_i - x_{i-1})$.

We define

$$\overline{\int_a^b} f(x)\,\mathrm{d}x \coloneqq \inf_{P \in \Pi(a,b)} U(P, f), \quad \underline{\int_a^b} f(x)\,\mathrm{d}x \coloneqq \sup_{P \in \Pi(a,b)} L(P, f).$$

Definition 3.3. f is Riemann integrable if

$$\overline{\int_a^b} f(x)\,\mathrm{d}x = \underline{\int_a^b} f(x)\,\mathrm{d}x$$

and in this case, we define

$$\int_a^b f(x)\,\mathrm{d}x \coloneqq \overline{\int_a^b} f(x)\,\mathrm{d}x = \underline{\int_a^b} f(x)\,\mathrm{d}x.$$

Example 3.4. Let $f \coloneqq \int_{\mathbb{Q}}$.

Proposition 3.5. *If P^* is a refinement of P, then $U(P, f) \ge U(P^*, f)$ and $L(P, f) \le L(P^*, f)$.*

Corollary 3.6.

$$\underline{\int_a^b} f(x)\,\mathrm{d}x \le \overline{\int_a^b} f(x)\,\mathrm{d}x.$$

Proof. Consider for each P_1 and P_2 their common refinement to obtain

$$L(P_1, f) \leq L(P_1 \vee P_2, f) \leq U(P_1 \vee P_2, f) \leq U(P_2, f).$$

$\square$

Proposition 3.7. *The following are equivalent:*

- *f is Riemann integrable.*
- *For all $\epsilon > 0$, there exists a partition $P \in \Pi(a, b)$ such that $U(P, f) - L(P, f) < \epsilon$.*

Proof. For the forward direction, fix $\epsilon > 0$ and choose P_1, P_2 such that

$$U(P_1, f) < \int_a^b f \, \mathrm{d}x + \frac{\epsilon}{2}, \quad L(P_2, f) > \int_a^b f \, \mathrm{d}x - \frac{\epsilon}{2}.$$

Consider the common refinement $P_1 \vee P_2$. We have

$$U(P_1 \vee P_2, f) \leq U(P_1, f) < \int_a^b f \, \mathrm{d}x + \frac{\epsilon}{2} < L(P_2, f) + \epsilon < L(P_1 \vee P_2, f) + \epsilon.$$

For the reverse direction, note that

$$\overline{\int_a^b} f(x) \, \mathrm{d}x \leq U(P, f) < L(P, f) + \epsilon \leq \underline{\int_a^b} f(x) \, \mathrm{d}x + \epsilon.$$

Thus sending $\epsilon \to 0$ gives

$$\overline{\int_a^b} f(x) \, \mathrm{d}x = \underline{\int_a^b} f(x) \, \mathrm{d}x.$$

$\square$

Example 3.8. Let $f := \mathbb{1}_{>1/2}$ be defined on $[0, 1]$. For each $\epsilon > 0$, pick

$$P = \left\{0, \frac{1}{2} - \frac{\epsilon}{2}, \frac{1}{2} + \frac{\epsilon}{2}, 1\right\}.$$

3.1 What functions are Riemann integrable?

- continuous
- continuous, except for finitely many points,
- monotone.

Notation 3.9. Notation: given $P \in \Pi(x_0, \ldots, x_n)$, we define

- $\Delta x_i \coloneqq x_i - x_{i-1}$,
- $M_i \coloneqq \sup_{x_{i-1} \leq x \leq x_i} f(x)$,
- $m_i \coloneqq \inf_{x_{i-1} \leq x \leq x_i} f(x)$.

We may then write

$$U(P, f) = \sum M_i \Delta x_i, \quad L(P, f) = \sum m_i \Delta x_i, \quad U(P, f) - L(P, f) = \sum (M_i - m_i) \Delta x_i.$$

Proposition 3.10. *If $f : [a, b] \to \mathbb{R}$ is continuous, then f is Riemann integrable.*

Proof. Note that f is uniformly continuous. □

Corollary 3.11. *If $f : [a, b] \to \mathbb{R}$ is continuous except for finitely many points, then f is Riemann integrable.*

Proof *(Sketch)*. Use continuity to handle "most" of the $(M_i - m_i)\Delta x_i$ and use the fact that Δx_i is small for the otherwise. □

Proposition 3.12. *If $f : [a, b] \to \mathbb{R}$ is monotone, then f is Riemann integrable.*

Proof. Suppose without loss of generality that f is increasing. Fix $\epsilon > 0$ and choose P such that $\Delta x_i < \epsilon$ for each i. We have

$$\begin{aligned} U(P, f) - L(P, f) &= \sum (M_i - m_i) \Delta x_i \\ &\leq \sum \epsilon [f(x_i) - f(x_{i-1})] = \epsilon [f(b) - f(a)]. \end{aligned}$$

□

Theorem 3.13. *If $f : [a, b] \to \mathbb{R}$ is integrable, $f([a, b]) \subset [c, d]$, and $\phi : [c, d] \to \mathbb{R}$ is continuous, then $h = \phi \circ f$ is integrable.*

Proof. Fix $\epsilon > 0$ and choose $\delta > 0$ such that

- $|x - y| < \delta$ implies $|\phi(x) - \phi(y)| < \epsilon$,
- $\delta < \epsilon$.

Choose P such that $U(P, f) - L(P, f) < \delta^2$. We have then that

$$\begin{aligned} U(P,h) - L(P,h) &= \sum (M_i^h - m_i^h)\Delta x_i \\ &= \sum_{i:M_i^f - m_i^f < \delta} (M_i^h - m_i^h)\Delta x_i + \sum_{i:M_i^f - m_i^f \geq \delta} (M_i^h - m_i^h)\Delta x_i. \end{aligned}$$

For the first term, note that if $M_i^f - m_i^f < \delta$ then $M_i^h - m_i^h < \epsilon$. For the second term, note that

$$\delta \sum_{i:M_i^f - m_i^f \geq \delta} \Delta x_i \leq \sum_{i:M_i^f - m_i^f \geq \delta} (M_i^f - m_i^f)\Delta x_i \leq \delta^2 < \delta\epsilon,$$

from which it follows that

$$\sum_{i:M_i^f - m_i^f \geq \delta} (M_i^h - m_i^h)\Delta x_i \leq (d' - c')\epsilon,$$

where d' and c' are chosen such that $\phi([c, d]) \subset [c', d']$. Finally,

$$U(P,h) - L(P,h) \leq \epsilon(b - a) + \epsilon(d' - c').$$

$\square$

Proposition 3.14.

(i) The set of integrable functions is a vector space, and integration is a linear map.

(ii) If $a < b < c$ and f is integrable on $[a, c]$ then

$$\int_a^c f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_b^c f(x)\,\mathrm{d}x.$$

(iii) If $f \leq g$ then

$$\int_a^b f(x)\,\mathrm{d}x \leq \int_a^b g(x)\,\mathrm{d}x.$$

(iv) $\left|\int_a^b f \, \mathrm{d}x\right| \leq \int_a^b |f| \, \mathrm{d}x \leq (b-a)\sup|f|$.

(v) If f and g are integrable, then fg is integrable.

Theorem 3.15 (The Fundamental Theorem of Calculus)**.** *Let $f : [a,b] \to \mathbb{R}$ be differentiable. Suppose $f' : [a,b] \to \mathbb{R}$ is Riemann is integrable. Then*

$$f(b) - f(a) = \int_a^b f'(x) \, \mathrm{d}x.$$

Proof. Take any partition P. The mean value theorem gives

$$f(x_i) - f(x_{i-1}) = f'(\xi_i)\Delta x_i$$

for some $\xi_i \in [x_{i-1}, x_i]$. Summing over i, we have $f(b) - f(a) = \sum f'(\xi_i)\Delta x_i$. Noting that

$$L(P, f') \leq \sum f'(\xi_i)\Delta x_i \leq U(P, f')$$

we complete the proof by taking inf and sup over P. □

Theorem 3.16 (The Fundamental Theorem of Calculus 2)**.** *Let $f : [a,b] \to \mathbb{R}$ be Riemann integrable. Define $F(x) = \int_a^x f(t) \, \mathrm{d}t$. Then*

- *F is continuous*
- *if f is continuous at x, then F is differentiable at x and $F'(x) = f(x)$.*

Proof. For $x < y$, we have

$$|F(x) - F(y)| = \left|\int_x^y f(t) \, \mathrm{d}t\right| \leq \int_x^y |f(t)| \, \mathrm{d}t \leq (y-x)\sup|f|.$$

Since f, being integrable, is bounded, we have from the above that F is Lipschitz and thus continuous.

For the second result, note that for $h > 0$ we have

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t) \, \mathrm{d}t.$$

Fix $\epsilon > 0$ and choose $\delta > 0$ such that

$$|x - t| < \delta \implies |f(x) - f(t)| < \epsilon.$$

If $0 < h < \delta$,we have

$$\left|\frac{F(x+h)-F(x)}{h} - f(x)\right| = \frac{1}{h}\left|\int_x^{x+h} f(t) - f(x)\, \mathrm{d}t\right|$$
$$= \frac{1}{h}\int_x^{x+h} |f(t) - f(x)|\, \mathrm{d}t \leq \epsilon.$$

□

3.2 Inequalities

Definition 3.17. Given $1 < p < \infty$, define

$$\|f\|_p = \left(\int_a^b |f|^p\right)^{1/p}.$$

3.2.1 Cauchy-Schwarz Inequality

Theorem 3.18 (Cauchy-Schwarz Inequality)**.** *If f and g are Riemann integrable, then* $\left|\int_a^b fg\, \mathrm{d}x\right| \leq \|f\|_2\|g\|_2$.

Proof. For any $a, b \in \mathbb{R}$ and $\epsilon > 0$, we claim that

$$ab \leq \frac{a^2}{2\epsilon} + \frac{\epsilon b^2}{2}.$$

To see this, note merely that

$$\frac{a^2}{\epsilon} + \epsilon b^2 - 2ab = \left(\frac{a}{\sqrt{\epsilon}} - \sqrt{\epsilon} b\right)^2 \geq 0.$$

This then gives

$$\left|\int_a^b fg\, \mathrm{d}x\right| \leq \int_a^b |fg|\, \mathrm{d}x \leq \int_a^b \left(\frac{f^2}{2\epsilon} + \frac{\epsilon g^2}{2}\right) \mathrm{d}x$$
$$= \frac{1}{2\epsilon}\|f\|_2^2 + \frac{\epsilon}{2}\|g\|_2^2.$$

Setting $\epsilon = \|f\|_2/\|g\|_2$ gives the desired result. □

We can use this result to control the size of $|f(x) - f(y)|$.

Corollary 3.19.
$$\left|\int_a^b f \, \mathrm{d}x\right| \leq \sqrt{b-a}\|f\|_2.$$

Proof. Take $g = 1$ and note that $\|1\|_2 = \sqrt{b-a}$. □

Theorem 3.20. *If $f : [a,b] \to \mathbb{R}$ is differentiable and $f' : [a,b] \to \mathbb{R}$ is integrable, then*
$$|f(x) - f(y)| \leq \|f'\|_2 |x-y|^{1/2}.$$
That is, f is Hölder continuous with Hölder constant $1/2$*.*

Proof. By the previous result,
$$|f(x) - f(y)| = \left|\int_x^y f' \, \mathrm{d}t\right| \leq |x-y|^{1/2}\|f'\|_2.$$
□

3.2.2 Hölder's Inequality

Theorem 3.21 (Hölder's Inequality)**.** *If f and g are integrable and* $1/p + 1/q = 1$*, then*
$$\left|\int_a^b fg \, \mathrm{d}x\right| \leq \|f\|_p \|g\|_q$$

Proof. Homework. □

We can again use this result to control the size of $|f(x) - f(y)|$.

Corollary 3.22. *If* $1/p + 1/q = 1$*, then*
$$\left|\int_a^b f \, \mathrm{d}x\right| \leq \|f\|_p |b-a|^{1/q}.$$

Theorem 3.23. *If f' in integrable and p, q are conjugate exponents, then*
$$|f(x) - f(y)| \leq \|f'\|_p |x-y|^{1/q}.$$

Proof. We have
$$|f(x) - f(y)| = \left|\int_a^b f' \, \mathrm{d}t\right| = \|f'\|_p |x-y|^{1/q}.$$
□

Remark 3.24. Taking a really large p (and thus a q close to one) gives a result similar to that given by the MVT. Then $\|f'\| \approx f'(\xi)$, where ξ is given by the MVT.

3.2.3 Jensen's Inequality

Theorem 3.25 (Jensen's Inequality)**.** *Let $f : [0,1] \to \mathbb{R}$ be integrable and $\phi : \mathbb{R} \to \mathbb{R}$ be convex (and hence continuous). Then*

$$\phi\left(\int_0^1 f \, \mathrm{d}x\right) \leq \int_0^1 \phi(f(x)) \, \mathrm{d}x.$$

Intuition: if $\sum \lambda_i = 1$, we have

$$\phi\left(\sum x_i \lambda_i\right) \leq \sum \lambda_i \phi(x_i)$$

4 Curves

What does it mean to integrate a map $f : [a,b] \to \mathbb{R}^n$? We set

$$\int_a^b f(t)\,\mathrm{d}t := \left(\int_a^b f_1(t)\,\mathrm{d}t, \dots, \int_a^b f_n(t)\,\mathrm{d}t\right) \in \mathbb{R}^n.$$

Theorem 4.1. *If $f : [a,b] \to \mathbb{R}^n$ is integrable, then*

$$\left|\int_a^b f(t)\,\mathrm{d}t\right| \leq \int_a^b |f(t)|\,\mathrm{d}t.$$

Proof. Write $y := \int_a^b f\,\mathrm{d}t \in \mathbb{R}^n$. We have

$$\begin{aligned} |y|^2 = \sum y_i \left(\int_a^b f_i\,\mathrm{d}t\right) &= \int_a^b \sum y_i f_i\,\mathrm{d}t \\ &\leq \int_a^b |y||f|\,\mathrm{d}t = |y| \int_a^b |f|\,\mathrm{d}t, \end{aligned}$$

where the inequality comes from Cauchy-Schwarz. Dividing by $|y|$ gives the desired result. □

Definition 4.2. A curve in $\mathbb{R}^n$ is a continuous function $\gamma : [a,b] \to \mathbb{R}^n$

What is the length of a curve?

Definition 4.3. Given a partition P of $[a,b]$, set

$$\Lambda(P,\gamma) := \sum |\gamma(x_i) - \gamma(x_{i-1})|$$

We define the length of γ by

$$\Lambda(\gamma) := \sup_{P \in \Pi(a,b)} \Lambda(P,\gamma).$$

How can we compute $\Lambda(\gamma)$?

$$\Lambda(P,\gamma) = \sum |\gamma(x_i) - \gamma(x_{i-1})| \approx \sum |\gamma'(x_i)|\Delta x_i \approx \int_a^b |\gamma'(x)|\,\mathrm{d}x.$$

Theorem 4.4. *Suppose $\gamma : [a,b] \to \mathbb{R}^n$ is C^1. Then $\Lambda(\gamma) = \int_a^b |\gamma'(t)|\,\mathrm{d}t$.*

Remark 4.5. We will repeatedly use FTC to obtain $\int_x^y \gamma')t)\ \mathrm{d}t = \gamma(y) - \gamma(x)$.

Proof. We prove first that $\Lambda(\gamma) \le \int_a^b |\gamma'(t)|\ \mathrm{d}t$. Fix P. We have

$$\begin{aligned}\Lambda(P,\gamma) = \sum |\gamma(x_i) - \gamma(x_{i-1})| &= \sum \left| \int_{x_{i-1}}^{x_i} \gamma'(t)\ \mathrm{d}t \right| \\ &\le \sum \int_{x_{i-1}}^{x_i} |\gamma'(t)|\ \mathrm{d}t = \int_a^b |\gamma'(t)|\ \mathrm{d}t.\end{aligned}$$

We take sup over P.

It remains to prove that $\Lambda(\gamma) \ge \int_a^b |\gamma'(t)|\ \mathrm{d}t$. Fix $\epsilon > 0$ and choose $\delta > 0$ small enough so that $|t - s| < \delta$ implies $|\gamma'(t) - \gamma'(s)| < \epsilon$. Then choose P such that $\Delta x_i < \delta$ and $L(P, |\gamma'|) > \int_a^b |\gamma(t)|\ \mathrm{d}t - \epsilon$. We now have

$$\begin{aligned}\gamma(x_i) - \gamma(x_{i-1}) = \int_{x_{i-1}}^{x_i} \gamma'\ \mathrm{d}t &= \int_{x_{i-1}}^{x_i} \gamma'(x_{i-1})\ \mathrm{d}t + \int_{x_{i-1}}^{x_i} (\gamma'(t) - \gamma'(x_{i-1}))\ \mathrm{d}t \\ &= \gamma'(x_{i-1})\Delta x_i + \int_{x_{i-1}}^{x_i} (\gamma'(t) - \gamma'(x_{i-1}))\ \mathrm{d}t.\end{aligned}$$

Thus

$$|\gamma'(x_{i-1})|\Delta x_i \le |\gamma(x_i) - \gamma(x_{i-1})| + \epsilon \Delta x_i.$$

We have

$$\begin{aligned}L(P, |\gamma'|) &\le \sum |\gamma'(x_{i-1})|\Delta x_i \\ &\le \Lambda(P,\gamma) + \epsilon(b-a) \le \Lambda(\gamma) + \epsilon(b-a).\end{aligned}$$

Therefore,

$$\int_a^b \gamma'\ \mathrm{d}t \le L(P, |\gamma'|) + \epsilon \le \Lambda(\gamma) + \epsilon(b-a) + \epsilon.$$

□

Example 4.6. Curve with infinite length: the Koch snowflake.

5 The Riemann-Stieltjes Integral

Let $\alpha : [a, b] \to \mathbb{R}$ be monotone increasing. We are assigning a "size" or "weight" of $\alpha(x_i) - \alpha(x_{i-1})$ to the interval $[x_{i-1}, x_i]$.

Definition 5.1. Given a partition P, we define the upper sum and upper integral as

$$U(P, f, \alpha) \coloneqq \sum M_i^f \Delta\alpha_i, \quad \overline{\int_a^b} f \,\mathrm{d}\alpha \coloneqq \inf_{P \in \Pi(a,b)} U(P, f, \alpha).$$

where

$$\Delta\alpha_i \coloneqq \alpha(x_i) - \alpha(x_{i-1}).$$

The lower sum and lower integral are defined equivalently. We say f is integrable with respect to α and write $f \in \mathcal{R}(\alpha)$ if the upper and lower integrals are equal, and in this case we define $\int_a^b f \,\mathrm{d}\alpha$ to be this common value.

Example 5.2. Consider the interval $[0, 1]$ and the function $\alpha(x) = \mathbb{1}(x > 1/2)$. We have

$$U(P, f, \alpha) = \sum M_i^f \Delta\alpha_i = M_{i^*}^f \Delta\alpha_{i^*} = M_{i^*}^f$$

where $1/2 \in (x_{i^*-1}, x_{i^*}]$. Then $\int_a^b f \,\mathrm{d}\alpha = f(1/2)$ if f is continuous.

Theorem 5.3. *The following are equivalent:*

- $f \in \mathcal{R}(\alpha)$.
- *For all* $\epsilon > 0$*, there exists* $P \in \Pi(a, b)$ *such that* $U(P, f) - L(P, f) < \epsilon$.

Theorem 5.4. *If* f *is continuous, then* $f \in \mathcal{R}(\alpha)$ *for any* α.

Proof. Fix $\epsilon > 0$. Find $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$ and choose P such that $\Delta\alpha_i < \delta$. We have then that

$$\begin{aligned} U(P, f, \alpha) - L(P, f, \alpha) &= \sum (M_i - m_i)\Delta\alpha_i \\ &< \sum \epsilon\Delta\alpha_i < \epsilon[\alpha(b) - \alpha(a)]. \end{aligned}$$

□

Theorem 5.5. *If* α *is continuous and* f *is monotone, then* $f \in \mathcal{R}(\alpha)$.

Proof. Suppose without loss of generality that f is increasing. Fix $\epsilon > 0$ and choose P such that $\Delta\alpha_i < \epsilon$ for each i. We have

$$U(P, f, \alpha) - L(P, f, \alpha) = \sum [f(x_i) - f(x_{i-1})]\Delta\alpha_i \leq \epsilon \sum [f(x_i) - f(x_{i-1})] \leq \epsilon [f(b) - f(a)].$$

$\square$

Proposition 5.6.

- $f \mapsto \int f \, \mathrm{d}\alpha$ *is linear. In particular,* $\mathcal{R}(\alpha)$ *is a vector space.*
- $f_1 \leq f_2$ *implies* $\int_a^b f \, \mathrm{d}\alpha \leq \int_a^b g \, \mathrm{d}\alpha$.
- $\left| \int_a^b f \, \mathrm{d}\alpha \right| \leq \sup_{a \leq x \leq b} |f(x)| [\alpha(b) - \alpha(a)]$.
- $\alpha \mapsto \int f \, \mathrm{d}\alpha$ *is linear.*
- $\left| \int_a^b f \, \mathrm{d}\alpha \right| \leq \int_a^b |f| \, \mathrm{d}\alpha$.
- *If* $f, g \in \mathcal{R}(\alpha)$, *then* $fg \in \mathcal{R}(\alpha)$.

Suppose α is smooth (and in particular α' exists). We have

$$\sum f(y_i)\Delta\alpha_i \approx \sum f(y_i)\alpha'(y_i)\Delta x_i.$$

This suggests that if α is differentiable, then

$$\int_a^b f \, \mathrm{d}\alpha = \int_a^b f\alpha' \, \mathrm{d}x.$$

This is in fact true by the following result:

Theorem 5.7. *If* α *is differentiable and* $\alpha \in \mathcal{R}(\alpha)$, *then* $f \in \mathcal{R}(\alpha)$ *if and only if* $f\alpha' \in \mathcal{R}$, *in which case we have*

$$\int_a^b f \, \mathrm{d}\alpha = \int_a^b f\alpha' \, \mathrm{d}x.$$

We prove the following stronger lemma:

Lemma 5.8. *If $\alpha' \in \mathbb{R}$, then for any bounded f, we have*

$$\overline{\int_a^b} f \,\mathrm{d}\alpha = \overline{\int_a^b} f\alpha' \,\mathrm{d}x$$

and similarly for the lower integrals.

Proof. Fix $\epsilon > 0$ and let P_0 be any partition such that

$$\sum (M_i^{\alpha'} - m_i^{\alpha'})\Delta x_i = U(P_0, \alpha') - L(P_0, \alpha') < \epsilon$$

Now let $y_i \in [x_{i-1}, x_i]$. We have

$$\sum f(y_i)\Delta\alpha_i = \sum f(y_i)\alpha'(z_i)\Delta x_i,$$

where $z_i \in (x_{i-1}, x_i)$. Then,

$$\begin{aligned}\left|\sum f(y_i)\Delta\alpha_i - \sum f(y_i)\alpha'(y_i)\Delta x_i\right| &\leq \sum |f(y_i)| \cdot |\alpha'(y_i) - \alpha'(z_i)| \cdot \Delta x_i \\ &\leq \max|f| \sum (M_i^{\alpha'} - m_i^{\alpha'})\Delta x_i \\ &\leq \max|f|\epsilon.\end{aligned}$$

Note that

$$U(P_0, f, \alpha) = \sup_{y_i} \sum f(y_i)\Delta\alpha_i, \quad U(P_0, f\alpha') = \sup_{y_i} \sum f(y_i)\alpha'(y_i)\Delta x_i.$$

Thus,

$$|U(P_0, f, \alpha) - U(P_0, f\alpha')| \leq \max|f|\epsilon,$$

where we used the fact that $|f(t) - g(t)| < \epsilon$ for all t implies $|\sup f(t) - \sup g(t)| < \epsilon$. Next, since refinements does not increase upper sums, we have

$$\overline{\int_a^b} f \,\mathrm{d}\alpha = \inf_{P \in \Pi(a,b)} U(P, f, \alpha) = \inf_{P \in \Pi(a,b)} U(P \vee P_0, f, \alpha),$$

and similarly,

$$\overline{\int_a^b} f\alpha' \,\mathrm{d}x = \inf_{P \in \Pi(a,b)} U(P, f\alpha').$$

Thus, using the estimate above, we have

$$\left|\overline{\int_a^b} f \,\mathrm{d}\alpha - \overline{\int_a^b} f\alpha' \,\mathrm{d}x\right| \leq \max|f|\epsilon.$$

□

Example 5.9. Consider $\alpha : [0,1] \to \mathbb{R}$ defined by

$$\alpha(x) \coloneqq \begin{cases} x & x \leq 1/2 \\ x+2 & x > 1/2 \end{cases}.$$

Using linearity of the Riemann-Stieltjes integral in α, we have $\int_0^a f \, \mathrm{d}\alpha = 2f(1/2) + \int_0^1 f \, \mathrm{d}x$.

6 Sequences and Series of Functions

Let X be an arbitrary set and consider functions $f_n : X \to \mathbb{R}$, $n \in \mathbb{N}$ and $f : X \to \mathbb{R}$.

Definition 6.1.

- We say $f_n \to f$ pointwise if for any $x \in X$ we have $f_n(x) \to f(x)$ as $n \to \infty$.
- We say $f_n \to f$ uniformly if $\sup_{x \in X} |f_n(x) - f(x)| \to 0$ as $n \to \infty$.

Note that pointwise convergence does not imply uniform convergence:

Example 6.2. Let

$$f_n(x) = \begin{cases} 1, & x \in [n, n+1) \\ 0, & \text{otherwise} \end{cases}.$$

It is easy to see that $f_n \to 0$ pointwise but not uniformly. Note also that $\int f_n = 1$ for each n, but $\int f = 0$. The mass escapes to infinity.

Example 6.3. Let $X = [0, 1]$. Let f_n be piecewise affine that is 0 on $[0, 1 - 1/n]$ and 1 at 1. We have $f_n \to \mathbb{1}_{\{1\}}$ pointwise.

Question 6.4. How does uniform convergence interact with

- continuity (if X is a metric space),
- integration (if $X = [a, b]$),
- differentiation (if $X = [a, b]$).

Theorem 6.5. *If (X, d) is a metric space, $f_n : X \to \mathbb{R}$ is continuous, and $f_n \to f$ uniformly, then f is continuous.*

Proof. Fix $x \in X$. For $y \in X$, we have

$$\begin{aligned} |f(x) - f(y)| &\le |f_n(x) - f_n(y)| + |f_n(x) - f(x)| + |f_n(y) - f(y)| \\ &\le |f_n(x) - f_n(y)| + 2 \sup_z |f(z) - f_n(z)|. \end{aligned}$$

Thus fix $\epsilon > 0$ and pick n such that $2 \sup_z |f(z) - f_n(z)| < \epsilon/2$. Then choose $\delta > 0$ such that $d(x, y) < \delta$ implies

$$|f_n(x) - f_n(y)| < \frac{\epsilon}{2}.$$

Now, if $d(x, y) < \delta$, then

$$|f(x) - f(y)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

□

Theorem 6.6. *Suppose $X = [a, b]$ and $\alpha : [a, b] \to \mathbb{R}$ is monotone. If $f_n \in \mathcal{R}(\alpha)$ and $f_n \to f$ uniformly, then $f \in \mathcal{R}(\alpha)$ and*

$$\int_a^b f \,\mathrm{d}\alpha = \lim_{n\to\infty} \int_a^b f_n \,\mathrm{d}\alpha.$$

Proof. We compare first $U(P, f, \alpha)$ with $U(P, f_n, \alpha)$. We have

$$\begin{aligned} U(P, f, \alpha) - U(P, f_n, \alpha) &= \sum (M_i^f - M_i^{f_n})\Delta\alpha_i \\ &\le \sum \left| \sup_{x\in[x_{i-1},x_i]} f(x) - f_n(x) \right| \Delta\alpha_i \\ &\le \sup_{x\in[a,b]} |f(x) - f_n(x)| \sum \Delta\alpha_i, \end{aligned}$$

where we used the fact that $M_i^f - M_i^{f_n} \le \sup_x |f(x) - f_n(x)|$. Taking an inf over P gives

$$\left| \overline{\int_a^b} f \,\mathrm{d}\alpha - \overline{\int_a^b} f_n \,\mathrm{d}\alpha \right| \le \sup_x |f(x) - f_n(x)| [\alpha(b) - \alpha(a)].$$

Then, sending $n \to \infty$, we have

$$\overline{\int_a^b} f \,\mathrm{d}\alpha = \lim_{n\to\infty} \overline{\int_a^b} f_n \,\mathrm{d}\alpha = \lim_{n\to\infty} \int_a^b f_n \,\mathrm{d}\alpha.$$

A similar proof shows that the lower integrals converge as well. □

Theorem 6.7. *Suppose $f_n : [a, b] \to \mathbb{R}$ is differentiable on $[a, b]$ and*

- *for some $x_0 \in [a, b]$, $\lim_{n\to\infty} f(x_0) = y_0$,*
- *$f_n' \to g$ uniformly for some $g : [a, b] \to \mathbb{R}$.*

Then, there exists a function $f : [a, b] \to \mathbb{R}$ such that $f_n \to f$ uniformly and $f' = g$.

Lemma 6.8. *Suppose that $f_n : X \to \mathbb{R}$ is uniformly Cauchy, that is, for each $\epsilon > 0$, there exists N such that for $m, n > N$,*

$$\sup_{x\in X}|f_n(x) - f_m(x)| < \epsilon.$$

Then there exists some f such that $f_n \to f$ uniformly.

Proof *(of Theorem 6.7)***.** We have

$$f_n(x) - f_m(x) = f_n(x_0) - f_m(x_0) + \big(f_n(x) - f_n(x_0)\big) - \big(f_m(x) - f_m(x_0)\big).$$

Applying the MVT to $x \mapsto f_n(x) - f_m(x)$ gives

$$\begin{aligned} f_n(x) - f_m(x) &= f_n(x_0) - f_m(x_0) + (x - x_0)\big(f_n'(y) - f_m'(y)\big) \\ &\le |f_n(x_0) - f_m(x_0)| + (b-a) \sup_{y\in[a,b]}\left|f_n'(y) - f_m'(y)\right|. \end{aligned}$$

Thus f_n is uniformly Cauchy and, by Lemma, there exists some f such that $f_n \to f$ uniformly.

It remains to show that $f' = g$. Fix $x \in [a,b]$. We have

$$\begin{aligned} &\frac{1}{h}\left[f(x+h) - f(x)\right] \\ &= \frac{1}{h}\left[f_n(x+h) - f_n(x) + [f(x+h) - f(x)] - [f_n(x+h) - f_n(x)]\right] \\ &= f_n'(x) + \frac{1}{h}\big\{[f_n(x+h) - f_n(x) -] - f_n'(x) + [f(x+h) - f(x)] - [f_n(x+h) - f_n(x)]\big\} \\ &= g(x) + \big[f_n'(x) - g(x)\big] \\ &\quad + \left[\frac{1}{h}\left[f_n(x+h) - f_n(x)\right] - f_n'(x)\right] \\ &\quad + \frac{1}{h}\left\{[f(x+h) - f(x)] - [f_n(x+h) - f_n(x)]\right\}. \end{aligned}$$

We show the last three terms tend to 0. The first of them tends to 0 by uniform convergence of f_n'. The second tends to 0 when h is small. For the last term, note that it is

$$
\begin{aligned}
&\lim_{m\to\infty} \frac{1}{h} \{[f_m(x+h) - f_m(x)] - [f_n(x_h) - f_n(x)]\} \\
&= \lim_{m\to\infty} [f'_m(y) - f'_n(y)] \\
&\leq \lim_{m\to\infty} \sup_{y\in[a,b]} \left|f'_m(y) - f'_n(y)\right| \\
&= \lim_{m\to\infty} \sup_{y\in[a,b]} \left|f'_m(y) - g(y)\right| + \left|g(y) - f'_n(y)\right| \\
&= \sup_{y\in[a,b]} \left|g(y) - f'_n(y)\right| = 0,
\end{aligned}
$$

where we used the MVT in the first equality. Note that the first and the third term tends to 0 as $n \to \infty$, independent of h, while the second term tends to 0 as $h \to 0$ for fixed n. Thus we need only choose large n and then small h. □

Remark 6.9. We have that uniform convergence preserves continuity and integrability, but not differentiability. For a counterexample, consider $f(x) := |x|$ and its mollification.

7 Function Spaces

Fix a metric (X, d) (often require X to be compact).

Definition 7.1.

- $C(X) = \{\text{continuous and bounded function } f : X \to \mathbb{R}\}$.
- For $f, g \in C(X)$, define

$$d_\infty(f, g) \coloneqq \sup_{x \in X} |f(x) - g(x)|.$$

Remark 7.2.

- $(C(X), d_\infty)$ is a complete metric space, since uniform limit of continuous functions is continuous.
- $d_\infty(f_n, f) \to 0$ if and only if $f_n \to f$ uniformly. That is, the metric d_∞ induces uniform convergence.
- $(C(X), d_\infty)$ is separable if X is compact.
- Compact subsets are equibounded and equicontinuous.‘

Proposition 7.3. *$C([0, 1])$ is separable.*

Proof. Consider the set of piecewise affine functions such that for some $0 = q_0 < q_1 < \cdots < q_n = 1$, where $q_i \in \mathbb{Q}$ we have $f(q_i) \in \mathbb{Q}$ and f is affine on $[q_{i-1}, q_i]$. Prove that this is dense in the set of all piecewise affine functions, which is in turn dense in $C([0, 1])$. □

Definition 7.4. A collection of functions $\mathcal{A} \subset C(X)$ is called an **algebra** if

- $f, g \in \mathcal{A}$ implies $f + g \in \mathcal{A}$,
- $f \in \mathcal{A}$, $c \in \mathbb{R}$ implies $cf \in \mathcal{A}$,
- $f, g \in \mathcal{A}$ implies $fg \in \mathcal{A}$.

We say that an algebra $\mathcal{A}$ is

- **unital** if $1 \in \mathcal{A}$,

- **separates points** if for any $x \neq y$ in X, there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.

Definition 7.5. Given $B \in C(X)$, we write $\overline{B}$ is the closure of B with respect to d_∞. That is,

$$\overline{B} := \{f \in C(X) | \exists f_n \in B \text{ with } f_n \to f \text{ uniformly}\}$$

Theorem 7.6 (Stone-Weierstrass). *Let $\mathcal{A} \subset C(X)$ be a unital algebra that separates points. Then $\overline{\mathcal{A}} = C(X)$.*

Lemma 7.7. *For any $R > 0$, there exists polynomials p_n such that*

$$\sup_{x \in [-R,R]} |p_n(x) - |x|| \longrightarrow 0.$$

The absolute value function can be approximated by polynomials.

Proof. Rudin Exercise 23. □

Lemma 7.8. *For any $x_1, x_2 \in X$, $x_1 \neq x_2$, $c_1, c_2 \in \mathbb{R}$, there exists $f \in \mathcal{A}$ such that*

$$f(x_1) = c_1, \quad f(x_2) = c_2.$$

Proof. Let g be such that $g(x_1) \neq g(x_2)$. Define

$$f_1(x) := \frac{c_1(g(x) - g(x_2))}{g(x_1) - g(x_2)}, \quad f_2(x) := \frac{c_2(g(x_1) - g(x))}{g(x_1) - g(x_2)}.$$

Note that $f_1, f_2 \in \mathcal{A}$ and $f_1 + f_2$ satisfies the desired properties. □

Lemma 7.9. *$\overline{\mathcal{A}}$ is also an algebra.*

Proof. Take $f, g \in \overline{\mathcal{A}}$. There exists $f_n, g_n \in \mathcal{A}$ such that $f_n \to f$ uniformly and $g_n \to g$ uniformly. Since $f_n + g_n \to f + g$ uniformly, we know $f + g \in \overline{\mathcal{A}}$ by closure. □

Proof *(of Stone-Weierstrass)*.

Step 1: $f \in \overline{\mathcal{A}}$ implies $|f| \in \overline{\mathcal{A}}$. Set $R = \max|f|$ and p_n the polynomials from the previous lemma. Let $f_n := P_n \circ f$. We have

$$\sup_{x \in X} |p_n \circ f(x) - |f(x)|| \leq \sup_{-R \leq t \leq R} |P_n(t) - |t|| \longrightarrow 0.$$

Thus $|f| \in \overline{\mathcal{A}}$.

Step 2: If $f, g \in \overline{\mathcal{A}}$, so are $\min(f, g)$ and $\max(f, g)$. Note merely that

$$\max(f,g) = \frac{f+g}{2} + \frac{|f-g|}{2}, \quad \max(f,g) = \frac{f+g}{2} - \frac{|f-g|}{2}.$$

This can be generalized to min and max over finitely many functions by induction.

Step 3: Given $f \in C(X)$, $\epsilon > 0$, and $x_0 \in X$, we can find $g \in \overline{\mathcal{A}}$ such that

- $g(x_0) = f(x_0)$,
- $g(x) \geq f(x) - \epsilon$ for each $x \in X$.

Given $x \in X$, let g_x be such that $g_x(x_0) = f(x_0)$ and $g_x(x) = f(x)$. Let V_x be a neighborhood of x such that $g_x \geq f - \epsilon$ on V_x. By compactness there exists $x_1, \ldots, x_n$ such that $X = \cup V_{x_i}$. Set $g := \max(g_{x_1}, \ldots, g_{x_n})$.

Step 4: Given $f \in C(X)$ and $\epsilon > 0$, there exists $g \in \overline{\mathcal{A}}$ such that

$$\sup_{x \in X} |f(x) - g(x)| < \epsilon.$$

For each $x \in X$, let g_x be such that

- $g_x(x) = f(x)$,
- $g_x(y) > f(y) - \epsilon$ for each $y \in X$.

For each $x \in X$, find a neighborhood V_x of x such that $g_x(y) < f(y) + \epsilon$ for each $y \in V_x$. By compactness there exists $x_1, \ldots, x_n$ such that $X = \cup V_{x_i}$. Set $g := \min(g_{x_1}, \ldots, g_{x_n})$. □

Theorem 7.10. *Polynomials are dense in $C[a, b]$.*

Proof. The set of polynomials on $[a, b]$ is a unital algebra that separates points. □

Theorem 7.11. *if X is a compact metric space, then $C(X)$ is separable.*

Proof. Let $\{x_n\}_{n \in \mathbb{N}}$ be a countable dense subset of X. For each n, let

$$f_n(x) = d(x, x_n).$$

Define

$$\mathcal{A}^0 := \left\{ f = f_{n_1} \ldots f_{n_m} \middle| n_i \in \mathbb{N} \right\}$$

and

$$\mathcal{A}^{\mathbb{R}} := \left\{ f = r_0 + \sum_{i=1}^{n} r_i g_i \,\middle|\, n \in \mathbb{N}, r_i \in \mathbb{R}, g_i \in \mathcal{A}^0 \right\}.$$

We can apply Stone-Weierstrass to $\mathcal{A}^{\mathbb{R}}$ and note that the following set is dense in $\mathcal{A}^{\mathbb{R}}$ and countable:

$$\mathcal{A} := \left\{ q_0 + \sum_{i=1}^{n} q_i g_i \,\middle|\, n \in \mathbb{N}, q_i \in \mathbb{Q}, g_i \in \mathcal{A}^0 \right\}.$$

□

Definition 7.12. We say $K \subset C(X)$ is **uniformly bounded** if there exists $M > 0$ such that

$$\sup_{x \in X} |f(x)| \leq M, \quad \forall f \in K.$$

We say $K \subset C(X)$ is **equicontinuous** if for all $\epsilon > 0$, there exists $\delta > 0$ such that if $f \in K$ and $x, y \in X$ such that $d(x, y) < \epsilon$, then

$$|f(x) - f(y)| < \epsilon.$$

Theorem 7.13. *Let X a compact metric space. If $K \subset C(x)$, then the following are equivalent:*

(i) K is compact,

(ii) K is closed, uniformly bounded, and equicontinuous.

Proof. (i) implies (ii) is the easier direction. (ii) implies (i) comes from the following theorem, by recalling that that compactness is equivalent to sequential compactness in metric spaces. □

Theorem 7.14 (Arzelà–Ascoli)**.** *Let $\{f_n\}_{n \in \mathbb{N}} \subset C(X)$ and suppose $\{f_n\}$ is uniformly bounded and equicontinuous. Then there exists a subsequence $\{f_{n_k}\}_{k \in \mathbb{N}}$ and $f \in C(X)$ such that*

$$f_{n_k} \longrightarrow f.$$

Lemma 7.15 (The diagonal subsequence trick)**.** *Suppose for each $i \in \mathbb{N}$, $\{a^i_j\}_{j \in \mathbb{N}}$ is a sequence and*

$$\sup_{i, j \in \mathbb{N}} \left| a^i_j \right| < \infty.$$

Then, there exists increasing indexes $\{n_k\}_{k \in \mathbb{N}} \subset \mathbb{N}$ and $c^i \in \mathbb{R}$ such that

$$a^i_{n_k} \longrightarrow c^i, \quad \forall i \in \mathbb{N}.$$

Lemma 7.16 (Continuous extension)**.** *Let $S \subset X$ be dense and $f : S \to \mathbb{R}$ be uniformly continuous. Then there exists a unique continuous extension $\tilde{f} : X \to \mathbb{R}$. That is, there exists a unique $\tilde{f} \in C(X)$ such that $\tilde{f}|_S = f$.*

Remark 7.17. It is necessary that f is uniformly continuous. While continuity on a compact subset implies uniform continuity, continuity on a dense subset of a compact does not. Counterexample: $f : \mathbb{Q} \to \mathbb{R}$ defined by $f(x) = \mathbb{1}_{[0,1/2)}(x)$.

Proof. Define for each $x \in X$

$$\tilde{f}(x) := \lim_{k\to\infty} f(s_k),$$

where $\{s_k\}_{k\in\mathbb{N}} \subset S$ is a sequence converging to x. It is easy to see that this is well-defined, an extension of f, and continuous.

For example, for continuity, fix $\epsilon > 0$ and choose $\delta > 0$ such that $d(s, s') < \delta$ implies $|f(s) - f(s')| < \epsilon$ for any $s, s' \in S$. Let $x, y \in X$ be such that $d(x, y) < \delta/3$ and choose $s_k^1 \to x$ and $s_k^2 \to y$. Then for all k large enough, we have $d(s_k^1, s_k^2) < \delta$. Thus,

$$\left|\tilde{f}(x) - \tilde{f}(y)\right| = \lim_{k\to\infty}\left|f(s_k^1) - f(s_k^2)\right| \le \epsilon.$$

□

Proof *(pour Arzelà–Ascoli)***.**

Step 1: Defining f on a dense subset. Let $S \subset X$ be countable and dense. Consider for each $s \in S$ the sequence $\{f_n(s)\}_{n\in\mathbb{N}}$. Note that $\{f_n(s)\}$ is uniformly bounded. The first Lemma then gives increasing indices $\{n_k\}_{n\in\mathbb{N}} \subset \mathbb{N}$ and $\{c(s)\}_{n\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} f_{n_k}(s) = c(s).$$

Step 2: The function $c : S \to \mathbb{R}$ is uniformly continuous. Fix $\epsilon > 0$ and choose $\delta > 0$ such that $d(x, y) < \delta$ implies $|f_n(x) - f_n(y)| < \epsilon$ for each n. For $s, s' \in S$ such that $d(s, s') < \delta$, we have

$$|c(s) - c(s')| = \lim_{k\to\infty}\left|f_{n_k}(s) - f_{n_k}(s')\right| \le \epsilon.$$

Step 3: The second Lemma then gives a unique continuous extension of c, say f. Fix $\epsilon > 0$ and choose $\delta > 0$ such that

- $d(x, y) < \delta$ implies $|f_n(x) - f_n(y)| < \epsilon$ for each n, and

- $d(x, y) < \delta$ implies $|f(x) - f(y)| < \epsilon$.

Note that $\{B_\delta(s) : s \in S\}$ is an open cover of X. Let $B_\delta(s_1), \ldots, B_\delta(s_m)$. Choose $N \in \mathbb{N}$ large enough such that $\left|f_{n_k}(s_i) - f(s_i)\right| < \epsilon$ for each $i = 1, \ldots, m$. For $x \in X$, choose s_i such that $x \in B_\delta(s_i)$. Then,

$$\begin{aligned}\left|f_{n_k}(x) - f(x)\right| &\leq \left|f_{n_k}(x) - f_{n_k}(s_i)\right| + \left|f_{n_k}(s_i) - f(s_i)\right| + |f(s_i) - f(x)| \\ &\leq 3\epsilon.\end{aligned}$$

□

Corollary 7.18 (form Step 3 of the previous proof)**.** *Pointwise convergence of an equicontinuous sequence of functions on a dense subset of the domain propagates to uniform convergence on the whole domain.*

Remark 7.19 (Un bref résumé). Let (X, d) be a compact metric space. $(C(X), d_\infty)$ is a new metric space with the following properties:

- it is complete (from the completeness of $\mathbb{R}$),
- it is separable (Stone-Weierstrass),
- we know what compact subsets look like (Arzelà–Ascoli).