

MATH20410 (W25): Analysis in \mathbb{R}^n II (accelerated)

Lecturer: Joe Jackson

Notes by: Aden Chen

Sunday 23rd February, 2025

Contents

1	Single-Variable Differential Calculus	3
2	Multivariable Differential Calculus	9
3	Integration	21
4	Curves	29
5	The Riemann-Stieltjes Integral	31

1 Single-Variable Differential Calculus

In this chapter, we consider mainly functions of the form $f : I \rightarrow \mathbb{R}$, where I is an interval, e.g., (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

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

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 $f' : I \rightarrow \mathbb{R}$ 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.* $(f \circ g)' = (f' \circ g) \cdot g'$.

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

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

$$\begin{aligned}\left(\frac{f}{g}\right)'(x) &= \lim_{t \rightarrow x} \frac{\frac{f}{g}(t) - \frac{f}{g}(x)}{t - x} \\ &= \lim_{t \rightarrow x} \frac{\frac{f(t)}{g(t)} + \frac{f(x)}{g(t)} - \frac{f(x)}{g(t)} + \frac{f(x)}{g(x)}}{t - x}.\end{aligned}$$

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

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

□

1.1 The Mean Value Theorem

Lemma 1.5. *Suppose $f : [a, b] \rightarrow \mathbb{R}$ 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 $f : [a, b] \rightarrow \mathbb{R}$ 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 $f : [a, b] \rightarrow \mathbb{R}$ 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 $f : (a, b) \rightarrow \mathbb{R}$ 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 $f : [a, b] \rightarrow \mathbb{R}$ be differentiable² 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 $\tilde{f} := f - \lambda x$. □

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 $f, g : [a, b] \rightarrow \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b) . Then there exists $x \in (a, b)$ such that*

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

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

$$h(t) := (f(b) - f(a))g(t) - (g(b) - g(a))f(t).$$

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 $f, g : [a, b] \rightarrow \mathbb{R}$ 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*

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

if the last limit exists.

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

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

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

Refer to Rudin or something for the general case.

1.3 Higher Derivatives

If $f : I \rightarrow \mathbb{R}$ is differentiable, then we can define the second derivative $f'' := (f')'$ 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 $x \rightarrow x_0$ that

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

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

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

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 $g''(y) = 0$. \square

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

$$P_0(x) := f(x_0),$$

$$P_1(x) := f(x_0) + f'(x_0)(x - x_0),$$

$$P_2(x) := f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2,$$

\vdots

$$P_n(x) := \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k,$$

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 $k = 1, \dots, n$.

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

- $f^{(k)}$ exists on $[a, b]$ for $k = 1, \dots, n$; 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 $g'''(y) = 0$. \square

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 $\{e_1, \dots, e_n\}$.
- \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 $f : \mathbb{R} \supset I \rightarrow \mathbb{R}^n$.

Definition 2.1. f is differentiable at x if the limit

$$f'(x) := \lim_{t \rightarrow x} \frac{f(t) - f(x)}{t - x}$$

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

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

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. □

2.2 Higher Dimensional Domain

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

Definition 2.4 (Partial Derivatives).

$$\frac{\partial f}{\partial x_i}(x) = D_i f(x) := \lim_{h \rightarrow 0} \frac{f(x + h e_i) - f(x)}{h}.$$

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

$$= D_u f(x) := \lim_{h \rightarrow 0} \frac{f(x + hu) - f(x)}{h}.$$

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

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

In this case we write $Df(x) = T$. In other words, $f(x + h) = f(x) + Df(x)(h) + o(|h|)$.

Remark 2.8.

- If f is differentiable, then

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

- It is easy to check that Df is well defined, that is, there is at most one T such that the limit holds.

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

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

where $v := (Te_1, \dots, Te_n)$.

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

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

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

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

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

□

Remark 2.11. In particular we have $D_i f(x) = D_{e_i} f(x) = Df(x)e_i = \langle \nabla f(x), e_i \rangle$. In other words, if f is differentiable, then $\nabla f(x) = (D_1 f, \dots, D_n f)$.

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 $D_1 f(0) = D_2 f(0) = 0$ but $D_{(1,1)} f(0)$ 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

$$D_u f(0) = \lim_{t \rightarrow 0} \frac{t^3 u_1^3}{t^2(u_1^2 + u_2^2)} \cdot \frac{1}{t} = \frac{u_1^3}{u_1^2 + u_2^2}.$$

However, $Df(0)$ cannot exist, since the above mapping is not linear.

Theorem 2.15. *If the partial derivatives $D_1 f, \dots, D_n f$ exist and are continuous (in a neighborhood of x), then f is differentiable at x .*

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

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

which by continuity of D_i is sublinear. □

2.3 Extension to Functions with Higher Dimensional Codomains

Immediate.

We have

$$Df(x) \in L(\mathbb{R}^n, \mathbb{R}^m), \quad \mathbb{R}^n \ni h \mapsto Df(x)h \in L(\mathbb{R}^n, \mathbb{R}^m),$$

and

$$Df : \mathcal{U} \mapsto L(\mathbb{R}^n, \mathbb{R}^m).$$

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

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

$$Df(x)h = [Df(x)]h.$$

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 $Df(x)$ exists, so do $D_i f_j$, and we have*

$$[Df(x)] = [D_i f_j] = [\nabla f_1(x) \dots \nabla f_m(x)]^\top.$$

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,*

$$Df(x)h = (Df_1 h, \dots, Df_m(x)h) = (\langle \nabla f_1(x), h \rangle, \dots, \langle \nabla f_m(x), h \rangle) = [Df(x)]h.$$

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

$$Th = (Df_1 h, \dots, Df_m(x)h).$$

Note that

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

The other direction is left as an exercise. \square

Corollary 2.19. *If $D_j f_i$ 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*

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x).$$

A formal calculation:³ We have

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

³In 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 $B = Dg(x)$ and $\lim_{h \rightarrow 0} R_g/h = 0$. Similarly, we write

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

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

2.5 Continuity of the Derivative

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

- $\mathcal{U} \ni x \rightarrow Df(x) \in L(\mathbb{R}^n, \mathbb{R}^m)$.
- $\mathcal{U} \ni x \rightarrow [Df(x)] \in \mathbb{R}^{m \times n}$.
- $\mathcal{U} \ni x \rightarrow D_j f_i(x) \in \mathbb{R}, i = 1, \dots, m, j = 1, \dots, n$.

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 $\|A\|_{\text{op}} = \sup_{|v|} |Av|$. Thus $\|T\| = \|[A]\|_{\text{op}}$.

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

Proposition 2.24. *The following statements are equivalent:*

- $x \mapsto Df(x)$ is continuous (wrt d).
- $x \mapsto [Df(x)]$ is continuous (wrt d_{op}).
- $x \mapsto [Df(x)]$ is continuous (wrt d_F).
- Each $x \mapsto D_j f_i(x)$ 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 $f : \mathbb{R}^n \supset E \rightarrow \mathbb{R}^n$ be C^1 , where E is open. Suppose $x_0 \in E$ and $Df(x_0)$ is invertible. Then there exists a neighborhood U of x_0 such that f is a bijection from U to $V := f(U)$, and $f^{-1} : V \rightarrow U$ is C^1 with derivative $D(f^{-1}(y)) = [Df(f^{-1}(y))]^{-1}$.*

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

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

This shows that $D(f^{-1}(y))$ and $Df(f^{-1}(y))$ 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 $Df(x_0) \in L(\mathbb{R}, \mathbb{R})$ 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 $y_k \rightarrow y$. We seek to show that $f^{-1}(y_k) \rightarrow f^{-1}(y)$. Equivalently, given $f(x_k) \rightarrow f(x)$, we show $x_k \rightarrow x$. 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 $h \rightarrow 0$, we have $\lim_{h \rightarrow 0} k = 0$, by the continuity of f^{-1} , and so

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

□

Before the general proof, we need the following result:

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

$$d(\phi(x), \phi(y)) \leq cd(x, y).$$

Then, there is a unique fixed point of ϕ .

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

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

Thus, for $n > m$, we have

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

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

Proof (Inverse Function Theorem, the General Case).

Step 1: Local Invertibility. Choose δ small enough that

- $\|Df(x)^{-1}\|$ is bounded in $B_\delta(x_0)$.⁴
- $\|Df(x) - Df(x')\|$ is “really small” if $x, x' \in B_\delta(x_0)$.

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

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

$$\phi_y(x) := x + Df(x_0)^{-1}(y - f(x)).$$

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

$$D\phi_y(x) = I - Df(x_0)^{-1}Df(x) = Df(x_0)^{-1} [Df(x_0) - Df(x)].$$

Then,

$$\|D\phi_y(x)\| \leq \|Df(x_0)^{-1}\| \|Df(x_0) - Df(x)\|$$

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 $V := f(U)$. 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

$$|x - x_0| \leq \|Df(x_0)^{-1}\| |f(x) - f(x_0)|.$$

Thus for $y = f(x)$ within $r/2\|Df(x_0)^{-1}\|$ 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,

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

and f^{-1} is Lipschitz.

Step 4: The formula for Df^{-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

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

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

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 $f : \mathbb{R}^{n+m} \supset E \rightarrow \mathbb{R}^n$. If f is differentiable at (x, y) , then $Df(x, y) \in L(\mathbb{R}^{n+m}, \mathbb{R}^n)$. For $(h, k) \in \mathbb{R}^{n+m}$, then $Df(x, y)(h, k) \in \mathbb{R}^n$. Write $D_x f(x, y)h = Df(x, y)(h, 0)$ and $D_y f(x, y)k = Df(x, y)(0, k)$. Note that $D_x f \in (\mathbb{R}^n, \mathbb{R}^n)$ and $D_y f \in (\mathbb{R}^m, \mathbb{R}^m)$.

Theorem 2.31 (Implicit Function Theorem). *Let $f : \mathbb{R}^{n+m} \supset E \rightarrow \mathbb{R}^n$. Suppose f is C^1 in a neighborhood of some point (x_0, y_0) such that $f(x_0, y_0) = 0$. If $D_x f(x_0, y_0)$ 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 $Dg(y) = -D_x f(g(y), y)^{-1} D_y f(g(y), y)$.*

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 $h(y) := f(g(y), y) = 0$, then $Dh(y) = D_x f(g(y), y) Dg(y) + D_y f(g(y), y) = 0$, giving $Dg = -(D_x f)^{-1} D_y f$.

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

$$[DF(x_0, y_0)] = \begin{bmatrix} D_x f(x_0, y_0) & D_y f(x_0, y_0) \\ 0 & I \end{bmatrix}.$$

It turns out that

$$\det DF(x_0, y_0) = \det D_x f(x_0, y_0) \det I - \det 0 \det D_y f(x_0, y_0) = \det D_x f(x_0, y_0) \neq 0.$$

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 $g(y) := G(0, y)$. 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 $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^n$ 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 $g : V \rightarrow \mathbb{R}^n$ such that $f(g(y), y) = 0$.

By the Implicit Function Theorem, such g exists if $D_x f(a, b)$ 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 $f : \mathbb{R}^{2+3} \rightarrow \mathbb{R}^2$ with

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

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

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

At (a, b) ,

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

Then

$$Dg(b) = -[D_x f(a, b)]^{-1} [D_y f(x, b)] = \begin{bmatrix} 1/4 & 1/5 & -3/20 \\ -1/2 & 6/5 & 1/10 \end{bmatrix},$$

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

2.8 Higher Partial Derivatives

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Note that $D_i f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Definition 2.35. Suppose $D_i f$ exists. Define $D_{ji} f(x) = D_j[D_i f](x)$ if the latter exists.

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

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

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

$$\begin{aligned} D_{12} f(x, y) &= \lim_{h \rightarrow 0} \frac{D_2(x + h, y) - D_2(x, y)}{h} \\ &= \lim_{h \rightarrow 0} \lim_{k \rightarrow 0} \frac{f(x + h, y + k) - f(x + h, y) - f(x, y + k) + f(x, y)}{hk} \\ &= \lim_{h \rightarrow 0} \lim_{k \rightarrow 0} D_{21} f(t, s), \end{aligned}$$

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

$$f(x + h) = f(x) + Df(x)h + o(h).$$

The “total” second order derivative of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ should thus satisfy

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

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

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

where $D^2 f(x)$ is the Hessian. That is,

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

3 Integration

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

Definition 3.1. A **Partition** P of $[a, b]$ is a collection of points x_0, \dots, x_n such that $a = x_0 < x_1 < \dots < x_n = b$. We say P^* is a **refinement** of P if $P \subset P^*$. We say $P_1 \vee P_2 := P_1 \cup P_2$ 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

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

We define

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

Definition 3.3. f is Riemann integrable if

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

and in this case, we define

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

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

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

Corollary 3.6.

$$\underline{\int_a^b} f(x) \, dx \leq \overline{\int_a^b} f(x) \, dx.$$

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

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

□

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

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

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

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

For the reverse direction, note that

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

Thus sending $\epsilon \rightarrow 0$ gives

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

□

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 $P \in \Pi(x_0, \dots, x_n)$, we define

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

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

Proof. Note that f is uniformly continuous. □

Corollary 3.11. *If $f : [a, b] \rightarrow \mathbb{R}$ 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 $f : [a, b] \rightarrow \mathbb{R}$ 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 $f : [a, b] \rightarrow \mathbb{R}$ is integrable, $f([a, b]) \subset [c, d]$, and $\phi : [c, d] \rightarrow \mathbb{R}$ 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,

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

□

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*

$$\int_a^c f(x) \, dx = \int_a^b f(x) \, dx + \int_b^c f(x) \, dx.$$

- (iii) *If $f \leq g$ then*

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

$$(iv) \left| \int_a^b f \, dx \right| \leq \int_a^b |f| \, dx \leq (b-a) \sup|f|.$$

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

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

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

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 . \square

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

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

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

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

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

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

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

$$\begin{aligned} \left| \frac{F(x+h) - F(x)}{h} - f(x) \right| &= \frac{1}{h} \left| \int_x^{x+h} f(t) - f(x) \, dt \right| \\ &= \frac{1}{h} \int_x^{x+h} |f(t) - f(x)| \, dt \leq \epsilon. \end{aligned}$$

□

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 $\left| \int_a^b fg \, dx \right| \leq \|f\|_2 \|g\|_2$.*

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

$$\begin{aligned} \left| \int_a^b fg \, dx \right| &\leq \int_a^b |fg| \, dx \leq \int_a^b \left(\frac{f^2}{2\epsilon} + \frac{\epsilon g^2}{2} \right) dx \\ &= \frac{1}{2\epsilon} \|f\|_2^2 + \frac{\epsilon}{2} \|g\|_2^2. \end{aligned}$$

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.

$$\left| \int_a^b f \, dx \right| \leq \sqrt{b-a} \|f\|_2.$$

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

Theorem 3.20. *If $f : [a, b] \rightarrow \mathbb{R}$ is differentiable and $f' : [a, b] \rightarrow \mathbb{R}$ 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,

$$|f(x) - f(y)| = \left| \int_x^y f' \, dt \right| \leq |x - y|^{1/2} \|f'\|_2.$$

□

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*

$$\left| \int_a^b f g \, dx \right| \leq \|f\|_p \|g\|_q$$

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*

$$\left| \int_a^b f \, dx \right| \leq \|f\|_p |b - a|^{1/q}.$$

Theorem 3.23. *If f' is integrable and p, q are conjugate exponents, then*

$$|f(x) - f(y)| \leq \|f'\|_p |x - y|^{1/q}.$$

Proof. We have

$$|f(x) - f(y)| = \left| \int_a^b f' \, dt \right| = \|f'\|_p |x - y|^{1/q}.$$

□

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 $f : [0, 1] \rightarrow \mathbb{R}$ be integrable and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ be convex (and hence continuous). Then*

$$\phi\left(\int_0^1 f \, dx\right) \leq \int_0^1 \phi(f(x)) \, dx.$$

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 $f : [a, b] \rightarrow \mathbb{R}^n$? We set

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

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

$$\left| \int_a^b f(t) \, dt \right| \leq \int_a^b |f(t)| \, dt.$$

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

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

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

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

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 $\Lambda(\gamma)$?

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

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

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

Proof. We prove first that $\Lambda(\gamma) \leq \int_a^b |\gamma'(t)| dt$. Fix P . We have

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

We take sup over P .

It remains to prove that $\Lambda(\gamma) \geq \int_a^b |\gamma'(t)| dt$. 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 $L(P, |\gamma'|) > \int_a^b |\gamma'(t)| dt - \epsilon$. We now have

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

Thus

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

We have

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

Therefore,

$$\int_a^b |\gamma'| dt \leq L(P, |\gamma'|) + \epsilon \leq \Lambda(\gamma) + \epsilon(b - a) + \epsilon.$$

□

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

5 The Riemann-Stieltjes Integral

Let $\alpha : [a, b] \rightarrow \mathbb{R}$ 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

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

where

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

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 $\int_a^b f \, d\alpha$ 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 $\int_a^b f \, d\alpha = f(1/2)$ 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

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

□

Proposition 5.6.

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

$$\int_a^b f \, d\alpha = \int_a^b f \alpha' \, dx.$$

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*

$$\int_a^b f \, d\alpha = \int_a^b f \alpha' \, dx.$$

We prove the following stronger lemma:

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

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

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

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

and similarly,

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

Thus, using the estimate above, we have

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

□

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

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

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