Lectures on Mathematics

by Alejandro José Soto Franco

Last updated: July 29, 2025
Hosted on Github · `WIP v1.1 Lecture Notes`

Contents

Preface. This is a collection of mathematical notes written in lecture format. No claims are made beyond what is written. The work is presented as it is. As a work-in-progress draft, what is not yet rigorous may become so later. The author has tried to write only what he can verify.

I Foundations of Analysis and Linear Systems

This section introduces the foundational constructs of differential calculus, vector-valued mappings, linear operators, and their integration into matrix dynamics. We develop these through a sequence of explicit examples and analytic constructions.

I.1 Intuition Primer: From Slope to Structure

Before introducing rigorous definitions, we begin with a conceptual foundation accessible to those new to higher mathematics.

What is a function?

A **function** $f : X \to Y$ is a rule that assigns to each input $x \in X$ exactly one output $f(x) \in Y$.

Examples:

- $f(x) = 2x$ maps each real number to its double.
- $g(t) = \sin(t)$ outputs the vertical coordinate of a point rotating on the unit circle at time t.

What is a derivative? The derivative of a function f at a point x is the *instantaneous rate of change*, or how much $f(x)$ changes per unit change in x, *in the limit of small changes.*

If $f(x)$ is smooth (no sharp corners), then near x,

$$f(x + h) \approx f(x) + f'(x) \cdot h. \tag{1}$$

This is the best linear approximation to f near x.

What is a tangent line? For differentiable f, the tangent line at x is the line that just touches the curve and has slope $f'(x)$:

$$\ell(h) = f(x) + f'(x)h. \tag{2}$$

Geometric idea: You "zoom in" infinitely close and the function starts to look like a line.

What does differentiable mean? A function is differentiable at a point if it has a well-defined tangent line there. The function must be smooth and not "kinked" (like $|x|$ at $x = 0$).

Rule of thumb: If the graph has no jumps or corners, and you can draw a tangent line, it's probably differentiable.

Why do we care about derivatives?

- Physics: Velocity is the derivative of position.
- Biology: A changing population's growth rate is the derivative of its size.
- Economics: Marginal cost is the derivative of total cost.

Derivatives make static quantities dynamic. They can be said to "measure" change; I will delay saying anything too deeply on that until Lecture II.

What is the notation?

- Leibniz: $\frac{\mathrm{d}f}{\mathrm{d}x}$ emphasizes the ratio of small changes.

- Prime: $f'(x)$ is a common shorthand.
- Lagrange form: $(f(x+h) - f(x))/h$ is the difference quotient.

All converge to the same quantity in the limit $h \to 0$.

Example I.1 (Example to Ground the Intuition)**.** Let $f(x) = x^2$. Then:

$$f'(x) = \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h\to 0} \frac{2xh + h^2}{h} = 2x. \tag{3}$$

This says that the slope of f at x is $2x$, and the graph becomes steeper as x increases.

Takeaway: The derivative is the bridge between algebra and geometry, between a function's formula and its shape.

We now move to a formal treatment of differentiability, including theorems, counterexamples, and structural interpretations.

I.2 Derivatives from the Viewpoint of Sequences

To motivate the formal definition of the derivative, we now introduce an approach based on sequences. This builds intuition around limits and prepares the ground for rigorous analysis.

Goal: Understand how a function behaves infinitesimally close to a point by analyzing nearby values.

Definition (Difference Quotient): Given a function $f : \mathbb{R} \to \mathbb{R}$ and a point $x \in \mathbb{R}$, the difference quotient at x with increment $h \neq 0$ is defined as

$$Q(h) := \frac{f(x+h) - f(x)}{h}.$$

This is the average rate of change of f over the interval $[x, x+h]$.

Definition (Derivative via Limit of Difference Quotients): We say f is *differentiable at* x if the limit

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

exists. This limit, if it exists, is denoted $f'(x)$.

Sequential Definition of the Derivative: An equivalent definition uses sequences. We say f is differentiable at x if for every sequence (h_n) with $h_n \to 0$ and $h_n \neq 0$, the sequence

$$\left(\frac{f(x+h_n) - f(x)}{h_n} \right)$$

converges to the same limit L, independent of the choice of sequence. Then $f'(x) = L$.

Example I.2 (Sequence-Based Computation of Derivative)**.** Let $f(x) = x^2$. Choose $x = 2$ and $h_n = 1/n$. Then:

$$\frac{f(2+h_n) - f(2)}{h_n} = \frac{(2+\frac{1}{n})^2 - 4}{1/n} = \frac{4 + \frac{4}{n} + \frac{1}{n^2} - 4}{1/n} = \frac{\frac{4}{n} + \frac{1}{n^2}}{1/n} = 4 + \tfrac{1}{n}.$$

Thus, the sequence tends to 4 as $n \to \infty$, and we conclude $f'(2) = 4$.

Remark 1. This sequential characterization emphasizes that differentiability requires the difference quotient to converge *uniformly across all sequences* tending to zero. If even one sequence yields a different limit, the function is not differentiable at that point.

Example I.3 (Failure of Sequential Convergence)**.** Let $f(x) = |x|$ and consider $x = 0$. Define two sequences:

$$h_n^{(+)} = \tfrac{1}{n}, \quad h_n^{(-)} = -\tfrac{1}{n}.$$

Then:

$$\frac{f(h_n^{(+)}) - f(0)}{h_n^{(+)}} = \frac{\frac{1}{n}}{\frac{1}{n}} = 1, \qquad \frac{f(h_n^{(-)}) - f(0)}{h_n^{(-)}} = \frac{\frac{1}{n}}{-\frac{1}{n}} = -1.$$

Since the left- and right-hand limits differ, $f'(0)$ does not exist. Hence, f is not differentiable at 0.

Conclusion: The derivative measures the limiting behavior of difference quotients. The sequential perspective helps clarify the distinction between continuity (convergence of outputs) and differentiability (convergence of slopes). In the next section, we formalize this intuition into precise linear approximation.

I.3 Derivatives and Local Linearization

We begin with the fundamental concept of the derivative as a local linear approximation.

Theorem I.1 (Differentiability Implies Local Linearity)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable at x_0. Then there exists a linear map $L(h) = f'(x_0)h$ such that*

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - L(h)}{h} = 0. \tag{4}$$

Proof. By the definition of the derivative at x_0:

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}, \tag{5}$$

which implies that

$$f(x_0 + h) = f(x_0) + f'(x_0)h + o(h), \tag{6}$$

where $o(h)$ denotes a function such that $\lim_{h \to 0} o(h)/h = 0$. □

Theorem I.2 (Chain Rule)**.** *Let $f = g \circ h$, where $g : \mathbb{R} \to \mathbb{R}$ is differentiable at $h(x)$ and $h : \mathbb{R} \to \mathbb{R}$ is differentiable at x. Then*

$$\frac{\mathrm{d}f}{\mathrm{d}x} = g'(h(x)) \cdot h'(x). \tag{7}$$

Proof. Let $\epsilon(h) \to 0$ and $\delta(x) \to 0$ be error terms:

$$g(h(x) + \Delta h) = g(h(x)) + g'(h(x))\Delta h + \epsilon(\Delta h)\Delta h, \tag{8}$$

$$h(x + \Delta x) = h(x) + h'(x)\Delta x + \delta(\Delta x)\Delta x. \tag{9}$$

Substitute and simplify to obtain the limit formula for $f(x)$. □

Theorem I.3 (Differentiability $\not\Rightarrow$ Continuity of Derivative)**.** *There exists a function $f : \mathbb{R} \to \mathbb{R}$ such that f is differentiable at every point, yet the derivative f' is not continuous on $\mathbb{R}$.*

Proof. Define a piecewise function $f : \mathbb{R} \to \mathbb{R}$ by

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

We show that:

1. f is differentiable at all $x \in \mathbb{R}$,
2. f' exists everywhere but is discontinuous at $x = 0$.

Step 1: Differentiability on $\mathbb{R} \setminus \{0\}$.

For $x \neq 0$, f is clearly differentiable as the product of smooth functions. Applying the product and chain rules:

$$\begin{aligned} f'(x) &= \frac{d}{dx}\left(x^2 \sin(1/x)\right) \\ &= 2x \sin(1/x) - \cos(1/x), \end{aligned} \tag{10}$$

valid for $x \neq 0$.

Step 2: Differentiability at $x = 0$.

We compute the derivative at $x = 0$ using the definition:

$$\begin{aligned} f'(0) &= \lim_{x \to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0} \frac{x^2 \sin(1/x)}{x} \\ &= \lim_{x \to 0} x \sin(1/x). \end{aligned}$$

Since $|\sin(1/x)| \leq 1$, we have:

$$|x \sin(1/x)| \leq |x| \xrightarrow{x \to 0} 0,$$

so the limit exists and equals 0. Hence,

$$f'(0) = 0.$$

Step 3: Discontinuity of f' at $x = 0$.

Let us now analyze the behavior of $f'(x)$ near zero. From equation (10), we have:

$$f'(x) = 2x \sin(1/x) - \cos(1/x), \quad x \neq 0.$$

As $x \to 0$, the first term $2x \sin(1/x)$ tends to zero:

$$|2x \sin(1/x)| \leq 2|x| \xrightarrow{x \to 0} 0.$$

However, the second term $\cos(1/x)$ does *not* converge:

$$\cos(1/x) \text{ oscillates between } -1 \text{ and } 1 \text{ as } x \to 0.$$

Therefore, $f'(x)$ does not converge to $f'(0) = 0$ as $x \to 0$.

In fact, the full expression $f'(x) = 2x \sin(1/x) - \cos(1/x)$ satisfies:

$$\limsup_{x \to 0} f'(x) = 1, \quad \liminf_{x \to 0} f'(x) = -1,$$

so f' has a non-removable discontinuity at $x = 0$.

Conclusion: f is differentiable everywhere, but f' is not continuous at $x = 0$, completing the counterexample. □

Theorem I.4 (Non-Differentiability from Non-Uniform Limit)**.** *Let $f(x) = |x|$. Then f is continuous everywhere but not differentiable at $x = 0$.*

Proof. Compute the limit of the difference quotient:

$$\lim_{h\to 0^+} \frac{|h| - 0}{h} = \lim_{h\to 0^+} \frac{h}{h} = 1, \tag{11}$$

$$\lim_{h\to 0^-} \frac{|-h| - 0}{h} = \lim_{h\to 0^-} \frac{-h}{h} = -1. \tag{12}$$

The one-sided limits disagree; hence, the derivative does not exist at $x = 0$. □

Theorem I.5 (Characterization of Continuity via ε–δ and Sequential Convergence)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ and let $x_0 \in \mathbb{R}$. The following statements are equivalent:*

(i) f is continuous at x_0;

(ii) For every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in \mathbb{R}$, if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$;

(iii) For every sequence $(x_n) \subset \mathbb{R}$ with $\lim_{n\to\infty} x_n = x_0$, it holds that $\lim_{n\to\infty} f(x_n) = f(x_0)$.

Proof. **(i) ⇒ (ii):**
This is immediate, as (ii) restates the standard ε–δ definition of continuity at a point.

(ii) ⇒ (iii):
Assume the ε–δ condition holds. Let (x_n) be a sequence in $\mathbb{R}$ such that $x_n \to x_0$. We must show that $f(x_n) \to f(x_0)$.

Let $\varepsilon > 0$ be arbitrary. By assumption, there exists $\delta > 0$ such that for all $x \in \mathbb{R}$ with $|x - x_0| < \delta$, we have $|f(x) - f(x_0)| < \varepsilon$. Since $x_n \to x_0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $|x_n - x_0| < \delta$. Therefore, for all $n \geq N$, $|f(x_n) - f(x_0)| < \varepsilon$, which proves that $f(x_n) \to f(x_0)$.

(iii) ⇒ (i):
Assume that for every sequence $(x_n) \subset \mathbb{R}$ with $x_n \to x_0$, we have $f(x_n) \to f(x_0)$. We must show that f is continuous at x_0.

Suppose for contradiction that f is not continuous at x_0. Then there exists $\varepsilon_0 > 0$ such that for every $\delta > 0$, there exists $x \in \mathbb{R}$ with $|x - x_0| < \delta$ and $|f(x) - f(x_0)| \geq \varepsilon_0$.

We now construct a sequence (x_n) that converges to x_0 but violates sequential continuity. For each $n \in \mathbb{N}$, set $\delta = \frac{1}{n}$ and choose $x_n \in \mathbb{R}$ such that $|x_n - x_0| < \frac{1}{n}$ and $|f(x_n) - f(x_0)| \geq \varepsilon_0$. Then $x_n \to x_0$, but $|f(x_n) - f(x_0)| \geq \varepsilon_0$ for all n, so $f(x_n) \not\to f(x_0)$, which contradicts our assumption in (iii). Hence, f must be continuous at x_0. □

Theorem I.6 (Uniform Continuity and Sequential Criterion)**.** *Let $f : \mathbb{R} \to \mathbb{R}$. Then the following are equivalent:*

(i) f is uniformly continuous on $\mathbb{R}$;

(ii) For every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in \mathbb{R}$, if $|x - y| < \delta$, then $|f(x) - f(y)| < \varepsilon$;

(iii) For all sequences $(x_n), (y_n) \subset \mathbb{R}$, if $|x_n - y_n| \to 0$, then $|f(x_n) - f(y_n)| \to 0$.

Proof. **(i) $\Rightarrow$ (ii):** This is the definition of uniform continuity.

(ii) $\Rightarrow$ (iii): Assume the uniform ε–δ condition holds. Let sequences (x_n), (y_n) be such that $|x_n - y_n| \to 0$. Fix $\varepsilon > 0$. Then there exists $\delta > 0$ such that for all $x, y \in \mathbb{R}$,

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon.$$

Since $|x_n - y_n| \to 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $|x_n - y_n| < \delta$. Hence, for all $n \geq N$, we have $|f(x_n) - f(y_n)| < \varepsilon$, so $|f(x_n) - f(y_n)| \to 0$, as desired.

(iii) $\Rightarrow$ (i): We prove the contrapositive: assume f is not uniformly continuous. Then there exists $\varepsilon_0 > 0$ such that for all $\delta > 0$, there exist $x, y \in \mathbb{R}$ with $|x-y| < \delta$ and $|f(x) - f(y)| \geq \varepsilon_0$.

Construct sequences $(x_n), (y_n)$ such that for each $n \in \mathbb{N}$, $|x_n - y_n| < \frac{1}{n}$ and $|f(x_n) - f(y_n)| \geq \varepsilon_0$. Then $|x_n - y_n| \to 0$, but $|f(x_n) - f(y_n)| \geq \varepsilon_0$ for all n, so the difference does not tend to zero. This contradicts (iii), hence f must be uniformly continuous.

□

Theorem I.7 (Continuity and Open Sets)**.** *Let $f : \mathbb{R} \to \mathbb{R}$. Then f is continuous (in the ε–δ sense) if and only if for every open set $U \subseteq \mathbb{R}$, the preimage $f^{-1}(U) \subseteq \mathbb{R}$ is open.*

Proof. **($\Rightarrow$)** Suppose f is continuous in the ε–δ sense. Let $U \subseteq \mathbb{R}$ be open and let $x_0 \in f^{-1}(U)$, so $f(x_0) \in U$. Since U is open, there exists $\varepsilon > 0$ such that the open ball $B_\varepsilon(f(x_0)) \subseteq U$.

By continuity at x_0, there exists $\delta > 0$ such that for all $x \in \mathbb{R}$,

$$|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon \Rightarrow f(x) \in B_\varepsilon(f(x_0)) \subseteq U.$$

Hence, $x \in f^{-1}(U)$, so $B_\delta(x_0) \subseteq f^{-1}(U)$. Therefore, $f^{-1}(U)$ is open.

($\Leftarrow$) Suppose $f^{-1}(U)$ is open for every open set $U \subseteq \mathbb{R}$. To show f is continuous at x_0, take arbitrary $\varepsilon > 0$. Consider the open interval $U = (f(x_0) - \varepsilon, f(x_0) + \varepsilon)$, which is open. Then $f^{-1}(U)$ is open and contains x_0, so there exists $\delta > 0$ such that $B_\delta(x_0) \subseteq f^{-1}(U)$. Hence, for all $x \in \mathbb{R}$ with $|x - x_0| < \delta$, we have $f(x) \in U$, i.e., $|f(x) - f(x_0)| < \varepsilon$. Thus f is continuous at x_0.

□

Remark 2 (Topological and Measure-Theoretic Significance of the Cantor–Lebesgue Function). Let $C([0,1])$ denote the Banach space of real-valued continuous functions on the closed interval $[0,1]$, equipped with the uniform norm

$$\|f\|_\infty = \sup_{x \in [0,1]} |f(x)|.$$

Then:

- $C([0,1])$ is complete $\Rightarrow$ every Cauchy sequence in $C([0,1])$ converges uniformly.
- $C([0,1])$ is separable $\Rightarrow$ it contains a countable dense subset (e.g., piecewise linear functions with rational breakpoints and values).
- $C([0,1])$ is infinite-dimensional $\Rightarrow$ there exists no finite basis spanning the space.

Chain of inclusions among function spaces:

$$\begin{aligned} f \in C^1([0,1]) &\Rightarrow f \in \mathrm{AC}([0,1]) \\ &\Rightarrow f \in C([0,1]). \end{aligned}$$

Each inclusion is strict. In particular, the Cantor–Lebesgue function $C(x)$ satisfies:

$$\begin{aligned} C(x) &\in C([0,1]) \\ C(x) &\notin \mathrm{AC}([0,1]) \\ C'(x) &= 0 \quad \text{for almost every } x \in [0,1]. \end{aligned}$$

Apparent paradox:

$$\begin{aligned} f \in \mathrm{AC}([0,1]), &\quad f'(x) = 0 \text{ a.e.} \Rightarrow f(x) = f(0) \quad \text{(constant)} \\ C'(x) = 0 \text{ a.e.}, &\quad C \notin \mathrm{AC}([0,1]) \Rightarrow C(x) \text{ is not constant.} \end{aligned}$$

This demonstrates that differentiability almost everywhere does not imply recoverability by integration unless f is absolutely continuous.

Measure-theoretic structure of the Cantor function:

$$\begin{aligned} \mu_C &\Rightarrow \text{singular measure supported on the Cantor set} \\ \text{For } A \subset [0,1], \quad \mu_C(A) &= \lambda(C^{-1}(A)) \text{ is singular w.r.t. Lebesgue measure.} \end{aligned}$$

Here λ is Lebesgue measure and μ_C is the distributional derivative of $C(x)$:

$$\frac{dC}{dx} = \mu_C, \quad \mu_C \perp \lambda.$$

Interpretation:

- The Cantor function $C(x)$ is continuous and non-constant,
- It increases on a set of Lebesgue measure zero,
- It has derivative zero almost everywhere,
- Yet it is *not* representable as an integral of its derivative.

Conclusion: The space $C([0,1])$ contains functions exhibiting topological regularity (continuity), yet singular with respect to classical calculus. The Cantor function exemplifies the subtle interplay between:

- Topology: continuity $\Rightarrow$ limits are preserved;
- Measure theory: derivative exists almost everywhere, but integration fails to recover the function;
- Functional analysis: norm convergence does not imply differentiability or absolute continuity.

Thus, while the smooth and absolutely continuous functions form a dense and analytically tractable subspace, they occupy only a small sliver of the full richness of $C([0,1])$. The Cantor function sits outside this analytic subspace, showing that continuity alone permits extreme measure-theoretic singularity.

Some Illustrative Examples

Example I.4 (Chain Rule in Action)**.** Let $f(x) = \sin(x^2)$.

Solution: Apply the chain rule with outer function $g(u) = \sin u$ and inner function $h(x) = x^2$:

$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\sin(x^2) = \cos(x^2)\cdot 2x. \tag{13}$$

Example I.5 (Corner Point and Directional Derivatives)**.** Let $f(x) = |x|$.

Solution:

$$f'(x) = \begin{cases} 1 & x > 0, \\ -1 & x < 0, \end{cases}$$

but $f'(0)$ is undefined, even though the function is continuous.

This demonstrates that continuous functions need not be differentiable, and the derivative, when it exists, encodes slope information that may be ill-defined at sharp points.

Example I.6 (Differentiable but Not C^1)**.** Let

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

Solution: $g(x)$ is differentiable everywhere:

$$g'(x) = 2x\sin(1/x) - \cos(1/x), \quad x \neq 0, \tag{14}$$

$$g'(0) = \lim_{x\to 0}\frac{x^2\sin(1/x)}{x} = \lim_{x\to 0} x\sin(1/x) = 0. \tag{15}$$

However, g' is not continuous at $x = 0$. The function is C^0 and C^1 on $\mathbb{R}\setminus\{0\}$ but fails to be C^1 globally.

Example I.7 (Higher-Order Derivatives)**.** Let $f(x) = \exp(-x^2)$. Then:

$$f'(x) = -2x\exp(-x^2), \tag{16}$$

$$f''(x) = (-2 + 4x^2)\exp(-x^2), \tag{17}$$

$$f^{(n)}(x) = P_n(x)\exp(-x^2), \tag{18}$$

where $P_n(x)$ are Hermite polynomials up to scaling.

This structure is key in the theory of orthogonal polynomials and spectral decomposition of linear operators.

Example I.8 (Uniform Continuity Without Differentiability)**.** Define

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

1. Continuity at $x = 0$ (via ε–δ definition): We claim that for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all x with $|x| < \delta$,

$$|f(x) - f(0)| = |x\sin(1/x)| < \varepsilon.$$

Since $|\sin(1/x)| \leq 1$, we choose $\delta = \varepsilon$, so that $|x\sin(1/x)| \leq |x| < \varepsilon$.

Thus, f is continuous at $x = 0$. Continuity elsewhere follows from smoothness of $x \sin(1/x)$ on $\mathbb{R} \setminus \{0\}$.

2. Uniform Continuity on $\mathbb{R}$: We use the Heine–Cantor theorem: continuous functions on compact intervals are uniformly continuous. For $x \in [-a, -\delta] \cup [\delta, a]$, f is continuous on a compact domain and hence uniformly continuous. To cover $\mathbb{R}$, we glue these intervals with the central neighborhood $(-\delta, \delta)$ where we already showed ε–δ uniform continuity directly. Therefore, f is uniformly continuous on all of $\mathbb{R}$.

3. Non-Differentiability at $x = 0$ (via ε–δ failure): Let us examine the difference quotient:

$$\frac{f(h) - f(0)}{h} = \frac{h \sin(1/h)}{h} = \sin(1/h),$$

which oscillates between $[-1, 1]$ as $h \to 0$ and has no limit. Thus,

$$\lim_{h \to 0} \frac{f(h) - f(0)}{h}$$

does not exist. So f fails the formal definition of differentiability:

$$\forall \varepsilon > 0,\ \exists \delta > 0 \text{ such that } 0 < |h| < \delta \Rightarrow \left| \frac{f(x+h) - f(x)}{h} - L \right| < \varepsilon.$$

No such L exists, so f is not differentiable at $x = 0$.

4. Classification:

$$f \in C^0(\mathbb{R}) \quad \text{but} \quad f \notin C^1(\mathbb{R}).$$

Moreover, f is *Lipschitz* on any compact $[\delta, 1]$ but not globally due to unbounded derivative near 0.

Conclusion: Uniform continuity does not imply differentiability. This function is a classic counterexample illustrating the strictness of the differentiability condition.

Example I.9 (Oscillatory Derivative and the Mean Value Theorem)**.** Let

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

1. Continuity at $x = 0$:

$$|f(x)| = |x^2 \cos(1/x)| \leq x^2 \to 0 \quad \text{as } x \to 0.$$

Hence f is continuous at 0. By ε–δ argument, given $\varepsilon > 0$, let $\delta = \sqrt{\varepsilon}$, then

$$|x| < \delta \Rightarrow |f(x) - f(0)| = |x^2 \cos(1/x)| < \varepsilon.$$

2. Differentiability at $x = 0$: Compute the limit of the difference quotient:

$$\frac{f(h) - f(0)}{h} = \frac{h^2 \cos(1/h)}{h} = h \cos(1/h) \to 0.$$

Thus $f'(0) = 0$.

3. Derivative for $x \neq 0$:

$$f'(x) = 2x \cos(1/x) + \sin(1/x).$$

This derivative is *not continuous* at 0, since $\sin(1/x)$ oscillates without bound. Thus, $f \in C^0(\mathbb{R})$, but $f \notin C^1(\mathbb{R})$.

4. Failure of Uniform Convergence: Let $f_n(x) = x^2 \cos(nx)$ on $[-1, 1]$. Then $f_n \to 0$ pointwise but not uniformly. This mimics the behavior of $f(x)$ near 0: the limiting function exists and is continuous, but the convergence of derivatives is not uniform.

5. Mean Value Theorem Insight: Let f be as above and consider $[-h, h]$ for small h. Then by MVT, $\exists\, c_h \in (-h, h)$ such that:

$$f'(c_h) = \frac{f(h) - f(-h)}{2h} = 0.$$

So although $f'(x)$ oscillates between -1 and 1, the average slope is zero — yet this does not imply pointwise convergence of the derivative.

6. Classification:

$$f \in C^0(\mathbb{R}), \quad f \text{ differentiable everywhere}, \quad f' \notin C^0(\mathbb{R}).$$

Takeaway: The derivative can exist at every point while still being discontinuous. Differentiability does not imply the derivative behaves "nicely." This requires additional structure such as uniform continuity or bounded variation.

The derivative encodes local linear behavior and supports approximation of nonlinear functions via linear structure. While differentiability implies local approximation, the converse does not hold without additional regularity. The hierarchy of smoothness—from C^0 to C^k to analytic—distinguishes classes of functions by how tightly they can be locally approximated by polynomials.

We now proceed to multivariable generalizations and the geometric structure of the Jacobian matrix.

I.4 Multivariable Differentiation and the Jacobian

Theorem I.8 (Total Derivative as Best Linear Approximation)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at a point $x_0 \in \mathbb{R}^n$. Then there exists a unique linear transformation $Df(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{\|h\| \to 0} \frac{\|f(x_0 + h) - f(x_0) - Df(x_0)[h]\|}{\|h\|} = 0.$$

Proof. By the definition of differentiability at x_0, there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ and a remainder function $r : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(x_0 + h) = f(x_0) + L(h) + r(h),$$

where $\lim_{\|h\| \to 0} \frac{\|r(h)\|}{\|h\|} = 0$. This linear map L is uniquely determined and defines the total derivative at x_0, denoted $Df(x_0)$. □

Remark 3. The total derivative generalizes the notion of the tangent line to higher-dimensional settings, where it defines a tangent plane or hyperplane. From the perspective of smooth manifolds, it corresponds to the differential or pushforward between tangent spaces.

Definition I.1 (Jacobian Matrix)**.** Let $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $x \in \mathbb{R}^n$. The Jacobian matrix of f at x, denoted $J_f(x)$, is the $m \times n$ matrix whose (i, j)-th entry is given by

$$(J_f(x))_{ij} = \frac{\partial f_i}{\partial x_j}(x).$$

Theorem I.9 (Jacobian Matrix Represents the Total Derivative)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $x_0 \in \mathbb{R}^n$ in the sense of Fréchet. Then the total derivative*

$$Df(x_0) \in \mathsf{Lin}(\mathbb{R}^n, \mathbb{R}^m)$$

is the unique linear map approximating f near x_0, and is represented in the standard basis by the Jacobian matrix $J_f(x_0) \in \mathbb{R}^{m\times n}$. That is, for all $h \in \mathbb{R}^n$,

$$Df(x_0)[h] = J_f(x_0) \cdot h.$$

Proof. **Step 1: Fréchet Differentiability.** The function f is Fréchet differentiable at x_0 if there exists a bounded linear map $Df(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\|h\|\to 0} \frac{\|f(x_0 + h) - f(x_0) - Df(x_0)[h]\|}{\|h\|} = 0.$$

This $Df(x_0)$ is unique when it exists, and defines the best linear approximation of f near x_0.

Step 2: Basis Representation. Let $\{e_1, \ldots, e_n\}$ denote the standard basis of $\mathbb{R}^n$, and $\{u_1, \ldots, u_m\}$ the standard basis of $\mathbb{R}^m$. For each $j = 1, \ldots, n$, define the partial derivatives of f at x_0 by:

$$\frac{\partial f}{\partial x_j}(x_0) := \lim_{t\to 0} \frac{f(x_0 + te_j) - f(x_0)}{t} \in \mathbb{R}^m.$$

Then the Jacobian matrix $J_f(x_0) \in \mathbb{R}^{m\times n}$ is formed by assembling these vectors as columns:

$$J_f(x_0) = \begin{bmatrix} | & & | \\ \frac{\partial f}{\partial x_1}(x_0) & \cdots & \frac{\partial f}{\partial x_n}(x_0) \\ | & & | \end{bmatrix}.$$

Step 3: Linearity of the Differential. Given any vector $h \in \mathbb{R}^n$, express it as a linear combination of basis vectors:

$$h = \sum_{j=1}^{n} h_j e_j.$$

Then by linearity of the total derivative:

$$Df(x_0)[h] = \sum_{j=1}^{n} h_j Df(x_0)[e_j] = \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(x_0).$$

This expression corresponds exactly to the matrix-vector product:

$$Df(x_0)[h] = J_f(x_0) \cdot h,$$

where h is treated as a column vector.

Step 4: Construction of the Coordinates of the Jacobian Matrix. Thus, the abstract linear map $Df(x_0) \in \mathsf{Lin}(\mathbb{R}^n, \mathbb{R}^m)$ is concretely realized in coordinates by $J_f(x_0)$:

$$\begin{aligned} Df(x_0) &\in \mathsf{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m) \\ &\cong \mathbb{R}^{m\times n}, \quad \text{via } L \mapsto \text{matrix representation w.r.t. standard bases.} \end{aligned}$$

Therefore, the Jacobian *is* the matrix of the total derivative with respect to the standard bases. The local linear approximation becomes

$$f(x_0 + h) = f(x_0) + J_f(x_0)h + o(\|h\|), \quad \text{as } h \to 0.$$

This concludes the identification. □

Remark 4 (Jacobian as the Pushforward in Differential Geometry). Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth map. In the language of differential geometry, the Jacobian matrix of f at a point $x \in \mathbb{R}^n$ is the matrix representation (in standard bases) of the *pushforward* map

$$df_x : T_x\mathbb{R}^n \longrightarrow T_{f(x)}\mathbb{R}^m,$$

which is the derivative of f at x viewed as a linear map between tangent spaces.

Interpretation:

- The tangent space $T_x\mathbb{R}^n$ is canonically isomorphic to $\mathbb{R}^n$, and similarly $T_{f(x)}\mathbb{R}^m \cong \mathbb{R}^m$.
- The pushforward df_x maps a tangent vector $v \in T_x\mathbb{R}^n$ to the velocity vector of the pushed-forward curve $f \circ \gamma$, where $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ is any smooth curve with $\gamma(0) = x$, $\dot{\gamma}(0) = v$.
- That is,
$$df_x(v) = \frac{d}{dt} f(\gamma(t))\bigg|_{t=0}.$$
- When expressed in local coordinates (i.e., standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$), the pushforward map df_x is given by the Jacobian matrix $J_f(x)$, with entries
$$[J_f(x)]_{ij} = \frac{\partial f^i}{\partial x^j}(x),$$
where $f = (f^1, \ldots, f^m)$ is the coordinate representation of f.

Summary: The Jacobian is not merely a matrix of partial derivatives—it encodes the differential df_x, which is a bundle map between tangent spaces:

$$df : T\mathbb{R}^n \to T\mathbb{R}^m, \quad \text{with } df|_x = df_x.$$

This formalism generalizes to maps between arbitrary smooth manifolds $f : M \to N$, where the differential at each point

$$df_p : T_pM \to T_{f(p)}N$$

defines a smooth bundle morphism between the total spaces of the tangent bundles:

$$df : TM \to TN.$$

In this broader setting, the Jacobian arises only when one trivializes the tangent bundles using coordinate charts—i.e., when one passes to local vector space isomorphisms $T_pM \cong \mathbb{R}^n$, $T_{f(p)}N \cong \mathbb{R}^m$. The Jacobian then becomes the matrix representative of df_p with respect to these coordinate frames.

Thus, the Jacobian encapsulates the full geometric action of the map f on tangent vectors—it is the local linearization of f and expresses how infinitesimal displacements transform under the mapping.

Theorem I.10 (Inverse Function Theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on an open set $U \subset \mathbb{R}^n$, and suppose $x_0 \in U$ satisfies $\det J_f(x_0) \neq 0$. Then there exists an open neighborhood $V \subset U$ of x_0 such that $f|_V : V \to f(V)$ is a diffeomorphism.*

Proof. Let $A \equiv Df(x_0) = J_f(x_0) \in \mathbb{R}^{n\times n}$, with A invertible. Set $y_0 \equiv f(x_0)$.

We construct a local inverse via fixed point iteration. For y near y_0, define the map

$$\Phi_y(x) \equiv x - A^{-1}(f(x) - y).$$

Then $\Phi_y(x) = x$ if and only if $f(x) = y$.

We show that Φ_y is a contraction on a neighborhood of x_0. Since f is C^1, for x near x_0 we write

$$f(x) = f(x_0) + A(x - x_0) + R(x),$$

where

$$\lim_{x \to x_0} \frac{\|R(x)\|}{\|x - x_0\|} = 0.$$

Fix $\varepsilon > 0$. Then for sufficiently small $\rho > 0$, we have

$$\|R(x)\| \leq \varepsilon \|x - x_0\| \quad \text{for all } x \in B_\rho(x_0).$$

Let $y \in B_\delta(y_0)$ for $\delta > 0$ small. Then

$$\begin{aligned} \Phi_y(x) - x_0 &= x - A^{-1}(f(x) - y) - x_0 \\ &= (x - x_0) - A^{-1}(f(x) - y - A(x - x_0)) \\ &= -A^{-1}(R(x) + f(x_0) - y). \end{aligned} \tag{19}$$

Taking norms:

$$\begin{aligned} \|\Phi_y(x) - x_0\| &\leq \|A^{-1}\| \cdot (\|R(x)\| + \|f(x_0) - y\|) \\ &\leq \|A^{-1}\| \cdot (\varepsilon \|x - x_0\| + \|f(x_0) - y\|). \end{aligned} \tag{20}$$

To ensure Φ_y maps $B_\rho(x_0)$ into itself, we impose

$$\|A^{-1}\|(\varepsilon\rho + \delta) \leq \rho.$$

Solving for δ:

$$\delta \leq \rho \left(\frac{1}{\|A^{-1}\|} - \varepsilon \right).$$

This is possible provided $\varepsilon < \frac{1}{\|A^{-1}\|}$.

Next, we verify that Φ_y is a contraction. Compute the derivative:

$$D\Phi_y(x) = I - A^{-1} Df(x).$$

Hence,

$$\|D\Phi_y(x)\| = \|A^{-1}(A - Df(x))\| \leq \|A^{-1}\| \cdot \|Df(x) - A\|.$$

Choose $\rho > 0$ small enough so that

$$\|Df(x) - A\| < \frac{1}{2\|A^{-1}\|} \quad \text{for all } x \in B_\rho(x_0).$$

Then

$$\|D\Phi_y(x)\| < \frac{1}{2},$$

so Φ_y is a contraction on $B_\rho(x_0)$.

By the Banach fixed point theorem, for each $y \in B_\delta(y_0)$, the map Φ_y has a unique fixed point $x_y \in B_\rho(x_0)$ satisfying $\Phi_y(x_y) = x_y$, hence $f(x_y) = y$.

Define $f^{-1}(y) \equiv x_y$ for $y \in B_\delta(y_0)$. Then f^{-1} is well-defined, continuous, and maps a neighborhood of y_0 into a neighborhood of x_0.

To prove differentiability, note that all steps depend on the differentiability of f and continuity of Df. Since the fixed point is obtained via smooth contraction, f^{-1} is C^1, and the chain rule gives

$$D(f^{-1})(f(x)) = (Df(x))^{-1}.$$

Hence f is a local diffeomorphism near x_0. $\square$

Remark 5. This theorem gives a topological characterization of the Jacobian determinant: a nonzero determinant implies local invertibility and the preservation of dimension and orientation. In manifold theory, it justifies local chart maps and smooth coordinate transitions.

Theorem I.11 (Change of Variables in Integration)**.** *Let $f : U \to \mathbb{R}^n$ be a C^1-diffeomorphism from an open set $U \subset \mathbb{R}^n$ to its image $f(U) \subset \mathbb{R}^n$. Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a nonnegative measurable function with compact support contained in $f(U)$. Then*

$$\int_{f(U)} \phi(y)\,dy = \int_U \phi(f(x))\,|\det J_f(x)|\;dx,$$

where $J_f(x) \in \mathbb{R}^{n\times n}$ denotes the Jacobian matrix of f at x.

Proof. Let us assume without loss of generality that ϕ is continuous with compact support contained in $f(U)$. The result will then follow for general nonnegative measurable functions by standard approximation (e.g., via monotone convergence).

Step 1: Local linear approximation. Since $f \in C^1$ and is a diffeomorphism, for each point $x_0 \in U$, the map f is locally well-approximated by its differential:

$$f(x) = f(x_0) + J_f(x_0)(x - x_0) + r(x),$$

where $\lim_{x\to x_0} \dfrac{\|r(x)\|}{\|x - x_0\|} = 0$. Thus, over sufficiently small neighborhoods $V \ni x_0$, we may regard f as almost affine.

Step 2: Partition of unity. Since $\operatorname{supp}(\phi) \subset f(U)$ is compact, and f is a diffeomorphism, its preimage under f is compact in U. Choose a finite open cover $\{V_i\}_{i=1}^N$ of $\operatorname{supp}(\phi \circ f) \subset U$ such that each $f|_{V_i}$ is a diffeomorphism onto its image. Let $\{\psi_i\}_{i=1}^N$ be a smooth partition of unity subordinate to this cover. Then,

$$\int_U \phi(f(x))\,|\det J_f(x)|\;dx = \sum_{i=1}^N \int_{V_i} \psi_i(x)\phi(f(x))\,|\det J_f(x)|\;dx.$$

Step 3: Local change of variables. Fix i. Define $y = f(x)$ and note that the change of variables theorem for compactly supported smooth functions in a single chart gives:

$$\int_{f(V_i)} \phi(y)\psi_i(f^{-1}(y))\,dy = \int_{V_i} \psi_i(x)\phi(f(x))\,|\det J_f(x)|\;dx.$$

Step 4: Summation. Summing over all i, we get:

$$\int_{f(U)} \phi(y)\,dy = \sum_{i=1}^N \int_{f(V_i)} \phi(y)\psi_i(f^{-1}(y))\,dy = \sum_{i=1}^N \int_{V_i} \psi_i(x)\phi(f(x))\,|\det J_f(x)|\;dx.$$

But the sum of ψ_i is identically 1 on $\operatorname{supp}(\phi \circ f)$, so the total integral becomes:

$$\int_U \phi(f(x))\,|\det J_f(x)|\;dx.$$

Step 5: Extension to general nonnegative measurable ϕ. Let $\phi_k \nearrow \phi$ be a sequence of continuous compactly supported functions. Then by the Monotone Convergence Theorem:

$$\int_{f(U)} \phi(y)\,dy = \lim_{k\to\infty} \int_{f(U)} \phi_k(y)\,dy = \lim_{k\to\infty} \int_U \phi_k(f(x))\,|\det J_f(x)|\;dx = \int_U \phi(f(x))\,|\det J_f(x)|\;dx.$$

$\square$

Remark 6. This establishes the measure-theoretic interpretation of the Jacobian determinant. It encodes how volume elements are scaled under the transformation f, and appears as a multiplicative factor in integrals, analogous to the determinant of a linear change of basis.

Theorem I.12 (Manifold Interpretation of Differentiability)**.** *Let M and N be smooth manifolds, and let $f : M \to N$ be a continuous map. Then f is differentiable at a point $p \in M$ if for every choice of coordinate charts (U, φ) around p and (V, ψ) around $f(p)$, the local representative*

$$\psi \circ f \circ \varphi^{-1} : \varphi(U \cap f^{-1}(V)) \subset \mathbb{R}^n \to \mathbb{R}^m$$

is differentiable in the classical sense at $\varphi(p)$.

Proof. The definition of a smooth map between manifolds is local. That is, differentiability of f at a point $p \in M$ must be verified via charts of M and N near p and $f(p)$, respectively. Given charts

$$\begin{aligned} &(U, \varphi), \quad \varphi : U \to \varphi(U) \subset \mathbb{R}^n, \\ &(V, \psi), \quad \psi : V \to \psi(V) \subset \mathbb{R}^m, \end{aligned}$$

with $p \in U$ and $f(p) \in V$, we consider the map

$$\psi \circ f \circ \varphi^{-1} : \varphi(U \cap f^{-1}(V)) \to \mathbb{R}^m.$$

The set $\varphi(U \cap f^{-1}(V))$ is open in $\mathbb{R}^n$ because charts are diffeomorphisms onto open subsets and f is continuous. The composition is well-defined on this open subset, and differentiability of f at p is defined to mean that this composition is differentiable at $\varphi(p) \in \mathbb{R}^n$.

Thus, differentiability of f at p reduces to differentiability of $\psi \circ f \circ \varphi^{-1}$ at the Euclidean point $\varphi(p)$. This aligns the manifold notion of differentiability with the classical multivariable calculus framework.

□

Remark 7 (Role of Open Sets and Coordinate Charts). The use of charts and open sets is essential in the differential topology of manifolds:

- Each chart (U, φ) maps an open neighborhood $U \subset M$ diffeomorphically to an open subset $\varphi(U) \subset \mathbb{R}^n$. This openness ensures that standard calculus tools (e.g., partial derivatives, limits) apply without boundary pathologies.
- The composition $\psi \circ f \circ \varphi^{-1}$ exists only when φ^{-1} maps into $\mathrm{dom}(f)$ and f maps into $\mathrm{dom}(\psi)$. Thus we require $U \cap f^{-1}(V)$ to be nonempty and open in M to preserve smooth structure under preimage and composition.
- Without open domains, differentiability cannot be defined in a neighborhood, and the derivative loses meaning as a local linear approximation.

Differentiability on manifolds is always a local Euclidean question, but must be posed on open sets to guarantee compatibility with the smooth atlas and the existence of valid coordinate representations.

Remark 8. This theorem emphasizes that differentiability is a local property defined relative to charts. The Jacobian matrix in coordinates expresses the derivative as a linear map between tangent spaces $df_p : T_pM \to T_{f(p)}N$, and global smoothness is obtained via the compatibility of these local expressions across overlapping charts.

Example I.10. Let $f(x, y) = (x^2 + y^2,\ xy)$. Then the Jacobian matrix is

$$J_f(x, y) = \begin{bmatrix} 2x & 2y \\ y & x \end{bmatrix}. \tag{21}$$

The rank of J_f at a point governs the local dimensionality of the image and injectivity behavior.

Example I.11. Let $f(x, y, z) = (xyz,\ e^{x+y+z})$. Then

$$J_f = \begin{bmatrix} yz & xz & xy \\ e^{x+y+z} & e^{x+y+z} & e^{x+y+z} \end{bmatrix}. \tag{22}$$

This illustrates nonlinear entanglement of variables across components.

I.5 Change of Coordinates and the Chain Rule

Theorem I.13 (Jacobian Determinant as Local Volume Distortion)**.** *Let $F : U \subset \mathbb{R}^n \to \mathbb{R}^n$ be a C^1 diffeomorphism onto its image $F(U)$. Let $\phi : F(U) \to \mathbb{R}$ be Lebesgue integrable. Then for every measurable subset $A \subset U$,*

$$\int_{F(A)} \phi(y)\, dy = \int_A \phi(F(x))\, |\det DF(x)|\ dx.$$

In particular, $|\det DF(x)|$ measures the infinitesimal volume scaling factor under the map F at the point x.

Proof. Let $x \in U$, and consider the first-order approximation of F near x. Since F is C^1, the Taylor expansion gives

$$F(x + h) = F(x) + DF(x)h + r(x, h),$$

with

$$\lim_{\|h\| \to 0} \frac{\|r(x, h)\|}{\|h\|} = 0.$$

Let $Q_\varepsilon(x)$ denote a cube of side length $\varepsilon > 0$ centered at x and aligned with the coordinate axes. For ε small, $F(Q_\varepsilon(x))$ is close to the image of $Q_\varepsilon(x)$ under the linear map $DF(x)$. That is,

$$F(Q_\varepsilon(x)) \approx DF(x)(Q_\varepsilon(x)),$$

and hence the volume satisfies

$$\mathrm{Vol}(F(Q_\varepsilon(x))) \approx |\det DF(x)| \cdot \varepsilon^n.$$

To pass from infinitesimal geometry to integration, we consider a finite measurable region $A \subset U$ and partition it into a countable collection of disjoint cubes $\{Q_i\}$ of small diameter, with $x_i \in Q_i$ representative points. Using the approximation above and assuming ϕ is bounded and continuous, we form the Riemann sum:

$$\begin{aligned}
\int_{F(A)} \phi(y)\, dy &\approx \sum_i \phi(F(x_i))\, \mathrm{Vol}(F(Q_i)) \\
&\approx \sum_i \phi(F(x_i)) |\det DF(x_i)|\, \mathrm{Vol}(Q_i) \\
&= \sum_i \phi(F(x_i)) |\det DF(x_i)| \int_{Q_i} 1\, dx \\
&\approx \int_A \phi(F(x)) |\det DF(x)|\, dx.
\end{aligned} \tag{23}$$

To justify taking the limit, observe:

- The approximation becomes exact as the mesh size of the partition goes to zero;
- The map F is injective and continuously differentiable, so it maps null sets to null sets and preserves measurability of images;
- The error in replacing F by its linearization decays uniformly on compact subsets;
- By the area formula and standard measure-theoretic arguments, the convergence of Riemann sums gives the identity.

This proves the formula for bounded, compactly supported continuous ϕ. By density of such functions in $L^1(F(U))$ and dominated convergence, the identity extends to all $\phi \in L^1(F(U))$.

The interpretation of $|\det DF(x)|$ as a volume distortion factor follows directly: the image of an infinitesimal cube at x under F has volume approximately scaled by $|\det DF(x)|$. □

Example I.12 (Composition via Chain Rule)**.** Let $f(u, v) = (u + v, uv)$ and $u = e^x$, $v = \ln y$. Then:

$$J_f(u, v) = \begin{bmatrix} 1 & 1 \\ v & u \end{bmatrix}, \qquad J_{(u,v)}(x, y) = \begin{bmatrix} e^x & 0 \\ 0 & \frac{1}{y} \end{bmatrix}, \tag{24}$$

$$J_g(x, y) = J_f \cdot J_{(u,v)} = \begin{bmatrix} e^x & \frac{1}{y} \\ ve^x & \frac{u}{y} \end{bmatrix}. \tag{25}$$

Example I.13 (Polar Coordinates)**.** Let $x = r \cos\theta$, $y = r \sin\theta$. Then

$$J_{(x,y)}(r, \theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}, \qquad \det J = r. \tag{26}$$

This determinant measures infinitesimal area distortion under the mapping.

I.6 Integration and Measure

Example I.14 (Riemann Midpoint Sum)**.** Let $f(x) = x$ on $[0, 1]$, partitioned into 4 intervals. Then:

$$\text{Midpoints: } x_i^* = \frac{2i - 1}{8}, \tag{27}$$

$$\text{Approximation: } \sum_{i=1}^{4} f(x_i^*) \cdot \frac{1}{4} = \frac{1}{4} \cdot \frac{16}{8} = \frac{1}{2}. \tag{28}$$

Example I.15 (Step Function)**.** Define

$$f(x) = \begin{cases} 1 & x \in [0, 1/2), \\ 2 & x \in [1/2, 1]. \end{cases}$$

Then

$$\int_0^1 f(x)\,\mathrm{d}x = \frac{1}{2} + \frac{1}{2} \cdot 2 = \frac{3}{2}. \tag{29}$$

Example I.16 (Improper Integral)**.**

$$\int_{-\infty}^{\infty} \frac{1}{1 + x^2}\,\mathrm{d}x = \pi.$$

This convergent improper integral highlights the need for infinite limits and decay conditions.

Theorem I.14 (Unification of Integration Notions via the Lebesgue Integral)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function. Then:*

1. *If f is bounded and continuous on $[a, b]$, then the Riemann integral exists and*

$$\int_a^b f(x)\,dx = \int_a^b f\,d\lambda.$$

2. *If f is a step function on $[a, b]$, then*

$$\int_a^b f\,d\lambda = \sum_{i=1}^n c_i \cdot \lambda(E_i),$$

where $f(x) = \sum_{i=1}^n c_i \chi_{E_i}(x)$ with disjoint measurable $E_i \subset [a, b]$.

3. *If $f \geq 0$ and $\int_{-\infty}^{\infty} f(x)\,dx$ exists as an improper Riemann integral, then*

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{\mathbb{R}} f\,d\lambda.$$

Proof. (1) Riemann ⇒ Lebesgue:

Let f be bounded and continuous on $[a, b]$. Then f is uniformly continuous, and for any $\varepsilon > 0$, there exists a tagged partition $\{[x_{i-1}, x_i], x_i^*\}$ such that

$$\left| \int_a^b f(x)\,dx - \sum_i f(x_i^*)(x_i - x_{i-1}) \right| < \varepsilon.$$

Let ϕ_n be a sequence of simple functions adapted to dyadic partitions converging uniformly to f. Then

$$\lim_{n\to\infty} \int_a^b \phi_n(x)\,dx = \int_a^b f(x)\,dx, \tag{30}$$

$$\lim_{n\to\infty} \int_a^b \phi_n(x)\,d\lambda = \int_a^b f\,d\lambda. \tag{31}$$

But Riemann and Lebesgue integrals agree on simple functions. Hence, both limits exist and are equal, and

$$\int_a^b f(x)\,dx = \int_a^b f\,d\lambda.$$

(2) Step function:

Let $f = \sum_{i=1}^n c_i \chi_{E_i}$ where $E_i \subset [a, b]$ are disjoint and measurable. Then the Lebesgue integral is defined as

$$\int_a^b f\,d\lambda = \sum_{i=1}^n c_i \int_{E_i} d\lambda = \sum_{i=1}^n c_i \cdot \lambda(E_i). \tag{32}$$

This is compatible with Riemann integration when E_i are intervals and c_i are constant values of f on those intervals.

(3) Improper ⇒ Lebesgue:

Let $f : \mathbb{R} \to [0, \infty)$ be measurable and suppose

$$\int_{-\infty}^{\infty} f(x)\,dx = \lim_{R\to\infty} \int_{-R}^{R} f(x)\,dx < \infty.$$

Define the truncated functions $f_R(x) \equiv f(x) \cdot \chi_{[-R,R]}(x)$. Each f_R is bounded and measurable with compact support, hence Lebesgue integrable:

$$\int_{-R}^{R} f(x)\,dx = \int_{\mathbb{R}} f_R\,d\lambda. \tag{33}$$

Moreover, $f_R(x) \nearrow f(x)$ pointwise as $R \to \infty$. Then by the monotone convergence theorem:

$$\int_{\mathbb{R}} f\,d\lambda = \lim_{R\to\infty} \int_{\mathbb{R}} f_R\,d\lambda = \lim_{R\to\infty} \int_{-R}^{R} f(x)\,dx. \tag{34}$$

This recovers the improper integral as a Lebesgue integral.

□

Theorem I.15 (Dominated Convergence Theorem)**.** *Let $\{f_n\}$ be a sequence of measurable functions $f_n : X \to \mathbb{R}$ such that:*

$$f_n(x) \to f(x) \quad \textit{pointwise a.e.}, \tag{35}$$

$$|f_n(x)| \le g(x) \quad \textit{for all } n, \ \textit{and some } g \in L^1(X). \tag{36}$$

Then:

$$f \in L^1(X), \quad \textit{and} \quad \lim_{n\to\infty} \int_X f_n\,d\lambda = \int_X f\,d\lambda. \tag{37}$$

Proof. Since $|f_n(x)| \le g(x)$ for all n and $g \in L^1(X)$, we have:

$$\int_X |f_n(x)|\,d\lambda \le \int_X g(x)\,d\lambda < \infty,$$

so each f_n is integrable, and the sequence $\{f_n\}$ is uniformly L^1-bounded.

By pointwise convergence and domination, we have:

$$|f(x)| = \lim_{n\to\infty} |f_n(x)| \le g(x),$$

so f is also integrable:

$$\int_X |f(x)|\,d\lambda \le \int_X g(x)\,d\lambda < \infty.$$

Let $\varepsilon > 0$ be arbitrary. Define the difference function:

$$h_n(x) \equiv |f_n(x) - f(x)|.$$

Then $h_n(x) \to 0$ pointwise almost everywhere, and $h_n(x) \le 2g(x)$, which is integrable.

Apply the Lebesgue Dominated Convergence Theorem to the nonnegative sequence $\{h_n\}$:

$$\lim_{n\to\infty} \int_X |f_n(x) - f(x)|\,d\lambda = 0.$$

Hence $f_n \to f$ in L^1, and in particular:

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

□

I.7 Linear Maps and Eigenstructure

Theorem I.16 (Matrices and Linear Maps are Equivalent)**.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a function. Then:*

1. *T is linear if and only if there exists a matrix $A \in \mathbb{R}^{m\times n}$ such that*
$$T(x) = Ax \quad \text{for all } x \in \mathbb{R}^n.$$
2. *Conversely, every matrix $A \in \mathbb{R}^{m\times n}$ defines a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(x) = Ax$.*

Proof. (1) Suppose T is linear. Then for all $x, y \in \mathbb{R}^n$ and scalars $\alpha, \beta \in \mathbb{R}$,
$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y).$$
Let $\{e_1, \dots, e_n\}$ be the standard basis of $\mathbb{R}^n$. Define:
$$a_j \equiv T(e_j) \in \mathbb{R}^m.$$
Then each $x \in \mathbb{R}^n$ can be written as $x = \sum_{j=1}^n x_j e_j$, so:
$$T(x) = T\left(\sum_{j=1}^n x_j e_j\right) = \sum_{j=1}^n x_j T(e_j) = \sum_{j=1}^n x_j a_j.$$
Define the matrix $A = [a_1 \ \cdots \ a_n] \in \mathbb{R}^{m\times n}$ with j-th column equal to a_j. Then:
$$T(x) = Ax.$$
Hence every linear map corresponds to matrix multiplication.
(2) Conversely, let $A \in \mathbb{R}^{m\times n}$ and define $T(x) = Ax$. Then for any $x, y \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$,
$$T(\alpha x + \beta y) = A(\alpha x + \beta y) = \alpha Ax + \beta Ay = \alpha T(x) + \beta T(y),$$
so T is linear. □

Theorem I.17 (Eigenvectors Characterize Invariant Directions)**.** *Let $A \in \mathbb{R}^{n\times n}$. A nonzero vector $v \in \mathbb{R}^n$ satisfies*
$$Av = \lambda v$$
for some $\lambda \in \mathbb{C}$ if and only if the line $\ell = \{cv : c \in \mathbb{R}\}$ is invariant under A, meaning:
$$A(\ell) \subseteq \ell.$$

Proof. Suppose $Av = \lambda v$ for some $\lambda \in \mathbb{C}$ and $v \neq 0$. Then for any $c \in \mathbb{R}$,
$$A(cv) = cAv = c\lambda v = \lambda(cv),$$
so $A(cv) \in \ell$. Thus ℓ is invariant under A.

Conversely, suppose $\ell = \{cv : c \in \mathbb{R}\}$ is A-invariant for some nonzero v. Then $Av \in \ell$, so there exists $\lambda \in \mathbb{R}$ such that
$$Av = \lambda v.$$
Hence v is an eigenvector of A with eigenvalue λ.

This proves that the eigenvectors of A are precisely the directions in which A acts as scalar multiplication, i.e., directions preserved by the action of A up to scaling. □

Example I.17 (Diagonalizable Matrix)**.** Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$. Eigenvalues are $\lambda = 2, 3$; eigenvectors are canonical basis.

Conclusion: Eigenbasis diagonalizes the operator.

Example I.18 (Defective Matrix)**.** $A = \begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ has a repeated eigenvalue $\lambda = 4$ but only one eigenvector.

Conclusion: Not diagonalizable; needs generalized eigenvectors (Jordan form).

Example I.19 (Symmetric Matrix)**.** Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$. Then:

$$\det(\lambda I - A) = \lambda^2 - 4\lambda - 1. \tag{38}$$

Real symmetric matrices always admit an orthonormal eigenbasis.

Theorem I.18 (Diagonalization and Spectral Structure of Real Matrices)**.** *Let $A \in \mathbb{R}^{n \times n}$. Then:*

1. *A is diagonalizable over $\mathbb{C}$ if and only if the geometric multiplicity of each eigenvalue equals its algebraic multiplicity.*
2. *A is diagonalizable over $\mathbb{R}$ if and only if it has n linearly independent real eigenvectors.*
3. *If A is symmetric, then:*
$$A = Q\Lambda Q^\top,$$
where Q is orthogonal and Λ is diagonal with real entries.

Proof. (1) Let $\lambda_1, \ldots, \lambda_k$ be the eigenvalues of A (possibly complex), with algebraic multiplicities m_i and geometric multiplicities $g_i = \dim\ker(A - \lambda_i I)$. Then A is diagonalizable over $\mathbb{C}$ if and only if

$$\sum_{i=1}^{k} g_i = n,$$

which occurs precisely when $g_i = m_i$ for all i. In that case, the union of eigenbases forms a basis of $\mathbb{C}^n$, and A is similar to a diagonal matrix:

$$A = P\Lambda P^{-1}, \quad \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

(2) For diagonalizability over $\mathbb{R}$, it is necessary that:

- All eigenvalues are real;
- The geometric multiplicity equals the algebraic multiplicity for each eigenvalue;
- The corresponding eigenvectors span $\mathbb{R}^n$.

If A has complex eigenvalues or insufficiently many real eigenvectors, it may still be diagonalizable over $\mathbb{C}$ but not over $\mathbb{R}$. The defective case arises when $g_i < m_i$ for some i, which prevents diagonalizability entirely.

(3) If A is symmetric, then by the Spectral Theorem for real symmetric matrices, A has real eigenvalues and admits an orthonormal eigenbasis. That is, there exists an orthogonal matrix Q such that:

$$Q^\top AQ = \Lambda,$$

with Λ diagonal. Equivalently,

$$A = Q\Lambda Q^\top.$$

Here, Q is composed of orthonormal eigenvectors, and Λ contains the corresponding real eigenvalues. This guarantees real diagonalizability and removes the possibility of defectiveness.

□

I.8 Matrix Exponential and Linear Dynamics

Example I.20 (Skew-Symmetric Generator)**.** Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Then:

$$e^{At} = \sum_{n=0}^{\infty} \frac{(At)^n}{n!} = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}. \tag{39}$$

Example I.21 (Decoupled Growth System)**.** $A = \mathrm{diag}(2, 3)$. Then:

$$e^{At} = \begin{bmatrix} e^{2t} & 0 \\ 0 & e^{3t} \end{bmatrix}, \quad x(t) = e^{At}x(0). \tag{40}$$

Example I.22 (Upper-Triangular Coupling)**.** $A = \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix}$. Solutions reflect coupling of components through non-diagonal structure.

Remark 9. The matrix exponential provides a unifying framework for solving all linear ODEs of the form $\frac{\mathrm{d}x}{\mathrm{d}t} = Ax$, including both stable, oscillatory, and growing dynamics.

Theorem I.19 (Matrix Exponential and Structure of Linear Solutions)**.** *Let $A \in \mathbb{R}^{n\times n}$. Then:*

1. *The unique solution to the initial value problem $\frac{\mathrm{d}x}{\mathrm{d}t} = Ax$, $x(0) = x_0$, is $x(t) = e^{At}x_0$.*
2. *If A is diagonalizable, i.e., $A = PDP^{-1}$ with D diagonal, then*

$$e^{At} = Pe^{Dt}P^{-1},$$

where e^{Dt} is diagonal with entries $e^{\lambda_i t}$.

Proof. Define the matrix exponential via its power series:

$$e^{At} = \sum_{k=0}^{\infty} \frac{(At)^k}{k!}.$$

This series converges absolutely for all $t \in \mathbb{R}$ since the norm of each term is bounded by a geometric series. The map $t \mapsto e^{At}$ is infinitely differentiable with:

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{At} = \sum_{k=1}^{\infty} \frac{A^k t^{k-1}}{(k-1)!} = A\sum_{k=0}^{\infty} \frac{(At)^k}{k!} = Ae^{At}.$$

Let $x(t) \equiv e^{At}x_0$. Then:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\big(e^{At}x_0\big) = Ae^{At}x_0 = Ax(t), \quad x(0) = e^{A\cdot 0}x_0 = Ix_0 = x_0.$$

This verifies that $x(t) = e^{At}x_0$ is a solution. Uniqueness follows from the general existence and uniqueness theorem for linear ODEs (e.g., Picard–Lindelöf).

Now suppose A is diagonalizable. Then there exists $P \in \mathsf{GL}_n(\mathbb{C})$ and a diagonal matrix $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ such that

$$A = PDP^{-1}.$$

We compute:

$$e^{At} = e^{PDP^{-1}t} = \sum_{k=0}^{\infty} \frac{(PDP^{-1}t)^k}{k!}.$$

Using associativity and $PP^{-1} = I$, we have:

$$(PDP^{-1})^k = PD^kP^{-1} \quad \text{for all } k \in \mathbb{N}.$$

Therefore:

$$e^{At} = \sum_{k=0}^{\infty} \frac{t^k}{k!}(PD^kP^{-1}) = P\left(\sum_{k=0}^{\infty} \frac{(Dt)^k}{k!}\right)P^{-1} = Pe^{Dt}P^{-1}. \tag{41}$$

Since D is diagonal, computing e^{Dt} reduces to exponentiating the diagonal entries:

$$e^{Dt} = \operatorname{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t}).$$

Therefore:

$$e^{At} = P \cdot \operatorname{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t}) \cdot P^{-1}.$$

Substituting into the solution formula:

$$x(t) = e^{At}x_0 = Pe^{Dt}P^{-1}x_0.$$

This gives an explicit expression for the trajectory $x(t)$ in terms of the eigenstructure of A.

This representation shows:

- If A has real eigenvalues, $x(t)$ grows or decays exponentially along the eigendirections.
- If A has complex eigenvalues with nonzero imaginary parts, $x(t)$ oscillates with exponential scaling.
- If A is defective (not diagonalizable), the above derivation fails, and e^{At} must be computed via Jordan form or directly from the series.

Thus, diagonalizability allows reduction of e^{At} to scalar exponentials. In all cases, e^{At} uniquely solves $\frac{\mathrm{d}x}{\mathrm{d}t} = Ax$. □

The following theorem is more advanced and will be discussed in more detail later:

Theorem I.20 (Linear Eigenfunction Decomposition of e^{At})**.** *Let $A \in \mathbb{R}^{n\times n}$ be diagonalizable over $\mathbb{C}$ with eigenpairs $\{(\lambda_i, v_i)\}_{i=1}^n$. Then any initial condition $x_0 \in \mathbb{C}^n$ decomposes as*

$$x_0 = \sum_{i=1}^{n} c_i v_i, \quad \text{with } c_i = \langle w_i, x_0\rangle,$$

where $\{w_i\}$ is the dual (left eigenvector) basis satisfying $\langle w_i, v_j\rangle = \delta_{ij}$. Then the solution to $\frac{\mathrm{d}x}{\mathrm{d}t} = Ax$ is:

$$x(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t} v_i.$$

Proof. Since A is diagonalizable over $\mathbb{C}$, there exists a complete basis of eigenvectors $\{v_i\}$ with $Av_i = \lambda_i v_i$, and a corresponding dual basis $\{w_i\}$ of left eigenvectors with $w_i^\top A = \lambda_i w_i^\top$, normalized such that $\langle w_i, v_j \rangle = \delta_{ij}$.

Any $x_0 \in \mathbb{C}^n$ decomposes as

$$x_0 = \sum_{i=1}^{n} c_i v_i, \quad \text{with } c_i = \langle w_i, x_0 \rangle.$$

The solution to the system $\frac{\mathrm{d}x}{\mathrm{d}t} = Ax$ is

$$x(t) = e^{At} x_0 = \sum_{i=1}^{n} c_i e^{At} v_i.$$

Since v_i is an eigenvector, we have:

$$e^{At} v_i = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k v_i = \sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_i^k v_i = e^{\lambda_i t} v_i.$$

Therefore:

$$x(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t} v_i.$$

Each mode v_i evolves independently with exponential scaling $e^{\lambda_i t}$, governed by its eigenvalue.

This decomposition shows that e^{At} acts diagonally in the eigenbasis and that the solution is the sum of dynamically decoupled modes.

□

II Inner Product Spaces and Spectral Geometry

This lecture develops the structure of vector spaces equipped with an inner product, leading to orthogonality, projections, and the spectral theorem for symmetric matrices. These form the linear geometric backbone of much of modern analysis and applied mathematics.

II.1 Normed and Inner Product Spaces

Let V be a vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. The structures of norms and inner products on V introduce geometric notions of length, angle, and orthogonality, essential for analysis and geometry.

Definition II.1 (Normed Vector Space)**.** A norm on V is a function $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ such that, for all $v, w \in V$ and scalars $\lambda \in \mathbb{F}$,

1. (Positive-definiteness) $\|v\| = 0 \iff v = 0$,
2. (Homogeneity) $\|\lambda v\| = |\lambda| \cdot \|v\|$,
3. (Triangle inequality) $\|v + w\| \leq \|v\| + \|w\|$.

The pair $(V, \|\cdot\|)$ is called a normed vector space.

Example II.1 (Measure-Theoretic and Geometric View of Classical Norms on $\mathbb{R}^n$)**.** Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and fix $p \in [1, \infty]$. The p-norm is defined by

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} & \text{if } 1 \leq p < \infty, \\ \max_{1 \leq i \leq n} |x_i| & \text{if } p = \infty. \end{cases}$$

These norms may be interpreted as L^p-norms on the finite set $\{1, \ldots, n\}$ with counting measure μ, i.e.,

$$\|x\|_p = \left(\int_{\{1,\ldots,n\}} |x(i)|^p \, d\mu(i) \right)^{1/p}, \quad \text{or} \quad \|x\|_\infty = \operatorname{ess\,sup}_i |x(i)|.$$

This connects finite-dimensional linear algebra to the theory of L^p-spaces in measure theory. Under this lens, $\ell^p(\mathbb{R}^n)$ is a discrete L^p-space and inherits duality, completeness, and convexity from general L^p-theory.

Geometry of the Unit Balls: The unit ball in the ℓ^p-norm is defined by

$$B_p = \{x \in \mathbb{R}^n : \|x\|_p \leq 1\}.$$

These balls exhibit rich geometric diversity:

- $p = 1$: The unit ball is a cross-polytope—sharp-cornered and faceted (octahedral in $\mathbb{R}^3$).
- $p = 2$: The Euclidean ball—perfectly round and smooth; uniquely rotation-invariant.
- $p = \infty$: The unit ball is a cube—flat faces and sharp corners.
- $1 < p < \infty$: The unit ball is strictly convex and smooth, with differentiable boundary and positive curvature at all points.

Inner Product Structure:

- For $p = 2$, the norm is induced by the standard dot product:

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i,$$

satisfying the parallelogram law and leading to a Euclidean (Hilbert) geometry.

- For $p \neq 2$, no inner product exists such that $\|x\|_p = \sqrt{\langle x, x\rangle}$. This failure is fundamental: the parallelogram identity is not satisfied unless $p = 2$. Consequently, concepts like orthogonality must be generalized:

  A vector $x \in \mathbb{R}^n$ is said to be *Birkhoff orthogonal* to $y \in \mathbb{R}^n$ if

$$\|x + \lambda y\| \geq \|x\| \quad \text{for all } \lambda \in \mathbb{R}.$$

  This captures the idea of directional minimality without relying on angle or projection.

Functional Analytic Continuation: These norms are not limited to finite dimension: for a measure space (X, μ), define

$$\|f\|_{L^p(X)} = \left(\int_X |f(x)|^p \, d\mu(x)\right)^{1/p}, \quad 1 \leq p < \infty.$$

The finite-dimensional ℓ^p norms are precisely the norms on step functions over a finite measure space with uniform weight. As $n \to \infty$, $\mathbb{R}^n$ becomes $\ell^p(\mathbb{N})$, and the continuous analog becomes the L^p-spaces of real analysis.

In summary, the classical ℓ^p-norms on $\mathbb{R}^n$ encode both algebraic and geometric structure, whose measure-theoretic generalizations lead naturally to infinite-dimensional function spaces. The special case $p = 2$ is the unique point of intersection where the geometry becomes Hilbertian, spectral theory applies, and familiar projections and orthogonality re-emerge.

mak

Definition II.2 (Inner Product Space)**.** An inner product is a map $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{F}$ satisfying:

1. (Linearity) $\langle \lambda u + v, w\rangle = \lambda\langle u, w\rangle + \langle v, w\rangle$,
2. (Conjugate symmetry) $\langle v, u\rangle = \overline{\langle u, v\rangle}$,
3. (Positive-definiteness) $\langle v, v\rangle \geq 0$, with equality iff $v = 0$.

The induced norm is $\|v\| := \sqrt{\langle v, v\rangle}$, and $(V, \langle\cdot,\cdot\rangle)$ becomes a normed space.

Example II.2. The standard inner product on $\mathbb{R}^n$ is

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i,$$

and on $\mathbb{C}^n$, it becomes

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i \overline{y_i}.$$

Theorem II.1 (Cauchy–Schwarz Inequality)**.** *Let V be an inner product space. For all $u, v \in V$,*

$$|\langle u, v\rangle| \leq \|u\| \cdot \|v\|,$$

with equality iff u and v are linearly dependent.

Proof. Let V be an inner product space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and let $u, v \in V$. If $v = 0$, then $\langle u, v\rangle = 0$ and the inequality becomes $0 \le \|u\| \cdot 0 = 0$, which holds trivially with equality. So assume $v \neq 0$.

Consider the function

$$f(\lambda) := \|u - \lambda v\|^2 = \langle u - \lambda v, u - \lambda v\rangle \ge 0 \quad \text{for all } \lambda \in \mathbb{F}.$$

Expanding this using the sesquilinearity of the inner product:

$$\begin{aligned} f(\lambda) &= \langle u, u\rangle - \overline{\lambda}\langle u, v\rangle - \lambda\langle v, u\rangle + |\lambda|^2\langle v, v\rangle \\ &= \|u\|^2 - 2\operatorname{Re}(\overline{\lambda}\langle u, v\rangle) + |\lambda|^2\|v\|^2. \end{aligned}$$

We now minimize this quadratic function over $\lambda \in \mathbb{F}$. Let us choose

$$\lambda := \frac{\langle u, v\rangle}{\|v\|^2},$$

which is well-defined since $v \neq 0$. Substituting into the expression:

$$\begin{aligned} f(\lambda) &= \|u\|^2 - \frac{|\langle u, v\rangle|^2}{\|v\|^2} - \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \frac{|\langle u, v\rangle|^2}{\|v\|^2} \\ &= \|u\|^2 - \frac{|\langle u, v\rangle|^2}{\|v\|^2}. \end{aligned}$$

Since $f(\lambda) \ge 0$, it follows that

$$\|u\|^2 - \frac{|\langle u, v\rangle|^2}{\|v\|^2} \ge 0,$$

which rearranges to

$$|\langle u, v\rangle|^2 \le \|u\|^2\|v\|^2.$$

Taking square roots on both sides (noting all terms are nonnegative),

$$|\langle u, v\rangle| \le \|u\| \cdot \|v\|.$$

Equality case: Equality holds if and only if $f(\lambda) = 0$, i.e., $u = \lambda v$ for some $\lambda \in \mathbb{F}$, which means u and v are linearly dependent.

□

Theorem II.2 (Parallelogram Law)**.** *Let V be a normed space. Then the norm is induced by an inner product if and only if the following identity holds:*

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

Proof. We prove both directions of the equivalence.

(⇒) **If the norm is induced by an inner product, then the parallelogram law holds.** Assume $\|x\|^2 = \langle x, x\rangle$ for some inner product $\langle\cdot,\cdot\rangle$ on V. Then for any $u, v \in V$,

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v\rangle = \langle u, u\rangle + 2\operatorname{Re}\langle u, v\rangle + \langle v, v\rangle, \\ \|u - v\|^2 &= \langle u - v, u - v\rangle = \langle u, u\rangle - 2\operatorname{Re}\langle u, v\rangle + \langle v, v\rangle. \end{aligned}$$

Adding these:

$$\|u + v\|^2 + \|u - v\|^2 = 2\langle u, u\rangle + 2\langle v, v\rangle = 2\|u\|^2 + 2\|v\|^2.$$

Thus, the parallelogram identity holds.

($\Leftarrow$) **If the norm satisfies the parallelogram law, then it is induced by an inner product.**

Assume $\|\cdot\|$ satisfies the identity

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2 \quad \text{for all } u, v \in V.$$

Define a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ via the polarization identity:

$$\langle u, v \rangle := \frac{1}{4}\left(\|u+v\|^2 - \|u-v\|^2\right).$$

We verify that this defines an inner product:

- *Bilinearity:* $\langle \cdot, \cdot \rangle$ is linear in each argument since norms are quadratic and the identity distributes over linear combinations.
- *Symmetry:* $\langle u, v \rangle = \langle v, u \rangle$, since the definition is symmetric in u and v.
- *Positive-definiteness:* $\langle u, u \rangle = \frac{1}{4}\left(\|2u\|^2 - 0\right) = \|u\|^2 \geq 0$, and equality only when $u = 0$.

Therefore, $\langle \cdot, \cdot \rangle$ is a valid inner product, and the norm satisfies

$$\|u\|^2 = \langle u, u \rangle,$$

so the norm is induced by it.

$\square$

Remark 10. This is both a test and a characterization. The identity is automatic in inner product spaces and fails in ℓ^p for $p \neq 2$. It reflects the linearity of energy distribution in Euclidean geometry.

Example II.3 (Failure of Inner Product Structure)**.** Let $V = \mathbb{R}^2$ with the ℓ^1 norm:

$$\|x\|_1 = |x_1| + |x_2|.$$

Then,

$$\|(1,0) + (0,1)\|^2 + \|(1,0) - (0,1)\|^2 = 2^2 + 2^2 = 8,$$

but

$$2\|(1,0)\|^2 + 2\|(0,1)\|^2 = 2 + 2 = 4.$$

Hence, the parallelogram identity fails, and no inner product induces this norm.

Remark 11. The distinction between general normed spaces and inner product spaces is fundamental in functional analysis. The geometry of Hilbert spaces arises only when the inner product structure is available. The failure of Pythagoras is a sharp boundary between the two regimes.

II.2 Completeness and Banach Spaces

Theorem II.3 (Completeness of ℓ^p Spaces)**.** *Let $1 \leq p \leq \infty$. The sequence space*

$$\ell^p := \left\{ x = (x_n)_{n=1}^\infty \in \mathbb{R}^\mathbb{N} \;\middle|\; \|x\|_p := \left(\sum_{n=1}^\infty |x_n|^p\right)^{1/p} < \infty \right\}$$

(for $p < \infty$) and

$$\ell^\infty := \left\{ x = (x_n)_{n=1}^\infty \in \mathbb{R}^\mathbb{N} \,\middle|\, \|x\|_\infty := \sup_{n\in\mathbb{N}} |x_n| < \infty \right\}$$

are Banach spaces under their respective norms.

Proof. We show that ℓ^p is complete for $1 \le p \le \infty$.

Let $(x^{(k)})_{k\in\mathbb{N}} \subset \ell^p$ be a Cauchy sequence. Then for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $j, k \ge N$,

$$\|x^{(j)} - x^{(k)}\|_p < \varepsilon.$$

Fix $n \in \mathbb{N}$. Then $(x_n^{(k)})_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, hence converges to some limit $x_n \in \mathbb{R}$. Define $x := (x_n)$.

We must show $x \in \ell^p$ and $\|x^{(k)} - x\|_p \to 0$. For any $\varepsilon > 0$, choose $k \ge N$, then by Fatou's lemma (or monotone convergence for non-negative series):

$$\|x^{(k)} - x\|_p^p = \sum_{n=1}^\infty |x_n^{(k)} - x_n|^p \le \liminf_{j\to\infty} \|x^{(k)} - x^{(j)}\|_p^p < \varepsilon^p.$$

Hence, $x \in \ell^p$ and the convergence holds in norm. Thus ℓ^p is complete. □

Theorem II.4 (Closed Subspaces of Banach Spaces are Banach)**.** *Let $(X, \|\cdot\|)$ be a Banach space and let $Y \subseteq X$ be a linear subspace. Then Y is a Banach space (i.e., complete) under the induced norm if and only if Y is closed in X.*

Proof. **(⇒)** Suppose Y is a Banach space. Then every Cauchy sequence $(y_n) \subset Y$ converges to some $y \in Y$. But since convergence in X implies convergence in X's topology, the limit lies in X. Hence, Y contains all its limit points: it is closed.

(⇐) Suppose $Y \subseteq X$ is closed. Let $(y_n) \subset Y$ be a Cauchy sequence. Since X is complete, $y_n \to y \in X$. But Y is closed, so $y \in Y$. Therefore, every Cauchy sequence in Y converges to a point in Y. Thus Y is complete.

□

II.3 Subspaces of Normed and Banach Spaces

Subspaces play a central role in functional analysis, particularly in studying restrictions of operators, constructing approximations, and analyzing functional decompositions. In the context of normed or Banach spaces, the key structural question is whether a subspace inherits completeness from the ambient space.

Definition II.3 (Normed Subspace)**.** Let $(X, \|\cdot\|)$ be a normed vector space. A subset $Y \subseteq X$ is called a *normed subspace* if:

1. Y is a linear subspace of X,
2. Y is equipped with the norm induced from X, i.e., $\|y\|_Y := \|y\|_X$ for all $y \in Y$.

Example II.4. Let $X = C([0,1])$ with the supremum norm $\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|$. Then the subspace

$$Y := \{f \in C([0,1]) \mid f(0) = 0\}$$

is a normed subspace. It is closed in X, and hence a Banach space under the induced norm.

Theorem II.5 (Closed Subspaces of Banach Spaces are Banach)**.** *Let X be a Banach space, and let $Y \subset X$ be a linear subspace. Then Y is a Banach space under the induced norm if and only if Y is closed in X.*

Proof. See previous section for detailed argument. □

Example II.5 (Non-closed Subspace)**.** Let $X = C([0,1])$ with the supremum norm. Define

$$Y := \{f \in C([0,1]) \mid f \text{ is a polynomial}\}.$$

Then Y is a normed subspace, but not complete. For example, the sequence $f_n(x) := \sum_{k=0}^{n} \frac{x^k}{k!}$ converges uniformly to $\exp(x) \in C([0,1])$, but $\exp(x) \notin Y$. Thus, Y is not closed, and hence not Banach.

Remark 12. This distinction is crucial in approximation theory. The space of polynomials is dense in many function spaces, but not closed. Understanding whether a subspace is closed allows us to determine whether projection methods or series expansions converge within the space.

Corollary 1. *Every finite-dimensional subspace of a normed space is closed and hence complete.*

Proof. In normed spaces, finite-dimensional subspaces are topologically isomorphic to $\mathbb{R}^n$, which is complete. Hence they are closed and complete. □

Remark 13. Infinite-dimensional subspaces need not be closed. A dense, non-closed subspace can fail to be complete even when the ambient space is Banach.

II.4 Orthonormal Bases and Projections

In inner product spaces, orthonormal systems encode an idealized coordinate system where notions of angle and length are preserved. They play a central role in decomposing vectors, simplifying linear operators, and extracting spectral content from geometric or analytic data.

Definition II.4 (Orthonormal Set)**.** Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. A finite set $\{v_i\}_{i=1}^k \subset V$ is said to be *orthonormal* if

$$\langle v_i, v_j\rangle = \delta_{ij},$$

i.e., the vectors are pairwise orthogonal and each has unit norm.

Remark 14. An orthonormal set forms a *partial isometry* from the standard basis of $\mathbb{R}^k$ into V. If the set is maximal and spans V, it defines a Hilbert basis, and the associated mapping is an isometric isomorphism.

Theorem II.6 (Orthogonal Projection)**.** *Let $W \subset V$ be a finite-dimensional subspace, and let $\{w_1, \dots, w_k\}$ be an orthonormal basis of W. Then for any $v \in V$, the unique vector in W minimizing $\|v - w\|$ is given by the orthogonal projection:*

$$\pi_W(v) = \sum_{i=1}^{k} \langle v, w_i\rangle w_i.$$

Moreover, the residual vector $v - \pi_W(v) \in V$ satisfies $\langle v - \pi_W(v), w_i\rangle = 0$ for all i, i.e.,

$$v - \pi_W(v) \perp W.$$

Proof. Let V be an inner product space and $W \subset V$ a finite-dimensional subspace with orthonormal basis $\{w_1, \dots, w_k\}$. Given any $v \in V$, we want to find the unique $w^* \in W$ that minimizes the distance $\|v - w\|$.

Any vector $w \in W$ can be written uniquely as

$$w = \sum_{i=1}^{k} a_i w_i,$$

for some coefficients $a_i \in \mathbb{F}$. Then the squared norm becomes

$$\begin{aligned}
\|v - w\|^2 &= \left\| v - \sum_{i=1}^{k} a_i w_i \right\|^2 \\
&= \left\langle v - \sum_{i=1}^{k} a_i w_i,\, v - \sum_{j=1}^{k} a_j w_j \right\rangle \\
&= \langle v, v\rangle - \sum_{j=1}^{k} \overline{a_j}\langle v, w_j\rangle - \sum_{i=1}^{k} a_i \langle w_i, v\rangle + \sum_{i,j=1}^{k} a_i \overline{a_j}\langle w_i, w_j\rangle.
\end{aligned}$$

Since $\{w_i\}$ is orthonormal, $\langle w_i, w_j\rangle = \delta_{ij}$, so the double sum collapses:

$$\sum_{i,j=1}^{k} a_i \overline{a_j}\langle w_i, w_j\rangle = \sum_{i=1}^{k} |a_i|^2.$$

Thus,

$$\|v - w\|^2 = \|v\|^2 - 2\,\mathrm{Re}\left(\sum_{i=1}^{k} a_i \langle w_i, v\rangle\right) + \sum_{i=1}^{k} |a_i|^2.$$

To minimize this quadratic expression in the a_i, observe that the unique minimum occurs when

$$a_i = \langle v, w_i\rangle.$$

Substituting these values yields the projection

$$\pi_W(v) = \sum_{i=1}^{k} \langle v, w_i\rangle w_i.$$

Orthogonality of the residual. Define the residual vector:

$$r := v - \pi_W(v).$$

Then for each $j = 1, \dots, k$,

$$\langle r, w_j\rangle = \left\langle v - \sum_{i=1}^{k} \langle v, w_i\rangle w_i,\, w_j \right\rangle = \langle v, w_j\rangle - \sum_{i=1}^{k} \langle v, w_i\rangle\langle w_i, w_j\rangle = \langle v, w_j\rangle - \langle v, w_j\rangle = 0.$$

Hence $r \perp w_j$ for all j, so $r \perp W$.

Uniqueness. If there were another $w' \in W$ such that $\|v - w'\| \le \|v - \pi_W(v)\|$, then $w' - \pi_W(v) \in W$ and

$$\|v - w'\|^2 = \|r + \pi_W(v) - w'\|^2 = \|r\|^2 + \|\pi_W(v) - w'\|^2 + 2\,\mathrm{Re}\langle r, \pi_W(v) - w'\rangle.$$

But since $r \perp W$ and $\pi_W(v) - w' \in W$, the cross term vanishes. Therefore,

$$\|v - w'\|^2 = \|r\|^2 + \|\pi_W(v) - w'\|^2 \geq \|r\|^2,$$

with equality iff $\pi_W(v) = w'$. So the minimizer is unique.

□

Example II.6 (Projection onto a Line)**.** Let $v = (3, 4) \in \mathbb{R}^2$, and let $w = (1, 0)$ be a unit vector in the x-direction. Then:

$$\mathrm{proj}_w(v) = \langle v, w\rangle w = 3 \cdot (1, 0) = (3, 0),$$

and the residual $v - \mathrm{proj}_w(v) = (0, 4)$ lies along the y-axis.

The projection splits v into orthogonal components along and perpendicular to w.

Definition II.5 (Orthogonal Decomposition)**.** Given an inner product space V and a subspace $W \subset V$, every $v \in V$ admits a unique decomposition

$$v = \pi_W(v) + \pi_{W^\perp}(v),$$

where $\pi_W(v) \in W$ and $\pi_{W^\perp}(v) \in W^\perp$. This decomposition is a direct sum:

$$V = W \oplus W^\perp.$$

Remark 15. The decomposition is functorial under isometries and natural in the category of finite-dimensional inner product spaces. The projection map π_W is a self-adjoint idempotent operator with image W and kernel $W^\perp$.

Theorem II.7 (Gram–Schmidt Orthonormalization)**.** *Let V be a finite-dimensional inner product space over $\mathbb{R}$ or $\mathbb{C}$, and let $\{v_1, \ldots, v_n\} \subset V$ be linearly independent. Then there exists an orthonormal set $\{u_1, \ldots, u_n\}$ such that for each k,*

$$\mathrm{span}(u_1, \ldots, u_k) = \mathrm{span}(v_1, \ldots, v_k).$$

Proof. The proof proceeds by induction.

Base case: Define

$$u_1 = \frac{v_1}{\|v_1\|},$$

which is well-defined since $v_1 \neq 0$. Then $\|u_1\| = 1$, and the span condition is satisfied.

Inductive step: Suppose $\{u_1, \ldots, u_{k-1}\}$ has been constructed such that:

1. $\langle u_i, u_j\rangle = \delta_{ij}$ for all i, j,
2. $\mathrm{span}(u_1, \ldots, u_{k-1}) = \mathrm{span}(v_1, \ldots, v_{k-1})$.

Let $W_{k-1} := \mathrm{span}(u_1, \ldots, u_{k-1})$, and define the orthogonal projection of v_k onto W_{k-1}:

$$\pi_{W_{k-1}}(v_k) = \sum_{j=1}^{k-1} \langle v_k, u_j\rangle u_j.$$

Define the orthogonal residual:

$$w_k := v_k - \pi_{W_{k-1}}(v_k).$$

Then $w_k \perp u_j$ for all $j < k$. Since $v_1, \ldots, v_n$ are linearly independent, $w_k \neq 0$. Finally, define

$$u_k := \frac{w_k}{\|w_k\|}.$$

The set $\{u_1, \ldots, u_n\}$ is orthonormal and yields a nested flag of subspaces respecting the original spans. □

Remark 16 (Coordinate-free Reformulation). The Gram–Schmidt process defines a canonical morphism from the poset of linearly independent tuples in V to the orthonormal frame bundle of V. It is equivariant under orthogonal transformations and yields an isometry $\mathbb{R}^n \to V$ via the map:

$$(x_1, \ldots, x_n) \mapsto \sum_{i=1}^{n} x_i u_i.$$

This provides a concrete instantiation of abstract Euclidean structure on V.

Remark 17 (QR Factorization). Given a full-rank matrix $A \in \mathbb{R}^{m\times n}$ with linearly independent columns, Gram–Schmidt yields an orthogonal matrix $Q \in \mathbb{R}^{m\times n}$ and upper triangular $R \in \mathbb{R}^{n\times n}$ such that

$$A = QR.$$

This is the analytic shadow of the orthonormalization process viewed through matrix coordinates.

II.5 Symmetric Matrices and Spectral Theorem

Remark 18. Symmetric matrices are the finite-dimensional shadows of self-adjoint operators. They represent the class of linear maps whose action preserves inner product geometry, making them central to the geometry of quadratic forms, variational calculus, and orthogonal decompositions. Their spectral behavior reflects deep mathematical phenomena—real eigenvalues, orthogonal eigenspaces, and full diagonalizability—which do not hold in general linear algebra but are guaranteed in the symmetric case due to the rich structure imposed by symmetry.

Definition II.6 (Symmetric and Hermitian Matrices)**.** Let $A \in \mathbb{F}^{n\times n}$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

- If $\mathbb{F} = \mathbb{R}$, then A is called *symmetric* if $A^\top = A$.
- If $\mathbb{F} = \mathbb{C}$, then A is called *Hermitian* if $A^* = \overline{A}^\top = A$.

Definition II.7 (Self-Adjoint Operator)**.** Let V be a finite-dimensional inner product space over $\mathbb{F}$, and let $T : V \to V$ be a linear operator. The adjoint $T^* : V \to V$ is defined by

$$\langle Tv, w\rangle = \langle v, T^*w\rangle \quad \text{for all } v, w \in V.$$

The operator T is called *self-adjoint* if $T = T^*$. In coordinates, relative to an orthonormal basis, T is represented by a symmetric (real case) or Hermitian (complex case) matrix.

Remark 19. Self-adjointness is a coordinate-free generalization of matrix symmetry. It ensures that the operator acts compatibly with the inner product structure, and its spectrum is real. In infinite dimensions, this leads to the theory of unbounded operators, essential in quantum mechanics and PDEs. In the finite-dimensional case, self-adjointness guarantees full diagonalizability and orthonormal eigenspaces.

Example II.7 (Diagonalization of a Symmetric Matrix)**.** Let

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}.$$

We verify that $A^\top = A$, so A is symmetric. We now compute its eigenvalues and an orthonormal basis of eigenvectors to diagonalize A.

Step 1: Compute the characteristic polynomial.

$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda - 2 & 1 \\ 1 & \lambda - 3 \end{bmatrix} = (\lambda - 2)(\lambda - 3) - 1 = \lambda^2 - 5\lambda + 5.$$

The eigenvalues are the roots of $\lambda^2 - 5\lambda + 5 = 0$, which are

$$\lambda_{1,2} = \frac{5 \pm \sqrt{5^2 - 4 \cdot 1 \cdot 5}}{2} = \frac{5 \pm \sqrt{25 - 20}}{2} = \frac{5 \pm \sqrt{5}}{2}.$$

Step 2: Compute eigenvectors.

Let $\lambda_1 = \frac{5+\sqrt{5}}{2}$. We solve $(A - \lambda_1 I)v = 0$. Define

$$A - \lambda_1 I = \begin{bmatrix} 2 - \lambda_1 & -1 \\ -1 & 3 - \lambda_1 \end{bmatrix}.$$

We can choose one row and solve:

$$(2 - \lambda_1)x - y = 0 \quad \Rightarrow \quad y = (2 - \lambda_1)x.$$

So an eigenvector is

$$v_1 = \begin{bmatrix} 1 \\ 2 - \lambda_1 \end{bmatrix}.$$

Similarly, for $\lambda_2 = \frac{5-\sqrt{5}}{2}$, we obtain

$$v_2 = \begin{bmatrix} 1 \\ 2 - \lambda_2 \end{bmatrix}.$$

Step 3: Normalize the eigenvectors.

We compute the norms:

$$\|v_1\|^2 = 1^2 + (2 - \lambda_1)^2,$$
$$\|v_2\|^2 = 1^2 + (2 - \lambda_2)^2.$$

Note:

$$2 - \lambda_1 = 2 - \frac{5 + \sqrt{5}}{2} = \frac{-1 - \sqrt{5}}{2}, \quad (2 - \lambda_1)^2 = \frac{(1 + \sqrt{5})^2}{4} = \frac{6 + 2\sqrt{5}}{4},$$

so

$$\|v_1\|^2 = 1 + \frac{6 + 2\sqrt{5}}{4} = \frac{4 + 6 + 2\sqrt{5}}{4} = \frac{10 + 2\sqrt{5}}{4}.$$

Thus, the normalized eigenvector is:

$$u_1 = \frac{1}{\sqrt{\frac{10+2\sqrt{5}}{4}}} \begin{bmatrix} 1 \\ \frac{-1-\sqrt{5}}{2} \end{bmatrix}.$$

Similarly, compute u_2 as the normalized v_2.

Step 4: Diagonalize A.

Let $Q = [u_1 \ u_2] \in \mathbb{R}^{2\times 2}$ be the orthogonal matrix with columns the normalized eigenvectors, and let

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Then

$$A = Q\Lambda Q^\top.$$

Conclusion: This validates the spectral theorem for this explicit case. The symmetric matrix A is orthogonally diagonalizable, with real eigenvalues and orthonormal eigenvectors.

Theorem II.8 (Spectral Theorem for Real Symmetric Matrices)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. Then:*

1. *All eigenvalues of A are real.*
2. *Eigenvectors corresponding to distinct eigenvalues are orthogonal.*
3. *There exists an orthonormal basis $\{v_1, \dots, v_n\} \subset \mathbb{R}^n$ consisting of eigenvectors of A.*

Equivalently, there exists an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ and a real diagonal matrix $\Lambda \in \mathbb{R}^{n\times n}$ such that

$$A = Q\Lambda Q^\top.$$

Proof. We proceed in four steps.

Step 1: Existence of a real eigenvalue.

Since A is symmetric and real, it is diagonalizable over $\mathbb{C}$. By the fundamental theorem of algebra, the characteristic polynomial $p_A(\lambda) = \det(A - \lambda I)$ has n (possibly complex) roots.

Let $\lambda \in \mathbb{C}$ be an eigenvalue, and $v \in \mathbb{C}^n$ be a corresponding eigenvector, $v \neq 0$, so that $Av = \lambda v$.

Consider the scalar $\langle Av, v\rangle$. Since A is real and symmetric, we have:

$$\langle Av, v\rangle = \langle v, Av\rangle = \langle v, \lambda v\rangle = \lambda\langle v, v\rangle.$$

But also:

$$\langle Av, v\rangle = \overline{\langle Av, v\rangle} = \overline{\langle \lambda v, v\rangle} = \overline{\lambda}\langle v, v\rangle.$$

Since $\langle v, v\rangle \neq 0$, we conclude $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$.

Thus, every eigenvalue of a real symmetric matrix is real.

Step 2: Orthogonality of eigenvectors.

Let $\lambda_1 \neq \lambda_2$ be distinct real eigenvalues of A, with corresponding eigenvectors $v_1, v_2 \in \mathbb{R}^n$. Then:

$$Av_1 = \lambda_1 v_1, \qquad Av_2 = \lambda_2 v_2.$$

Compute:

$$\langle Av_1, v_2\rangle = \lambda_1\langle v_1, v_2\rangle, \qquad \langle v_1, Av_2\rangle = \lambda_2\langle v_1, v_2\rangle.$$

But A is symmetric, so $\langle Av_1, v_2\rangle = \langle v_1, Av_2\rangle$, hence

$$\lambda_1\langle v_1, v_2\rangle = \lambda_2\langle v_1, v_2\rangle.$$

Subtracting gives

$$(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0.$$

Since $\lambda_1 \neq \lambda_2$, it follows that $\langle v_1, v_2\rangle = 0$.

Step 3: Inductive diagonalization.

We prove the result by induction on n.

Base case: For $n = 1$, any 1×1 real symmetric matrix is diagonal and the claim is trivial.

Inductive step: Let $A \in \mathbb{R}^{n\times n}$ be symmetric. From Step 1, A has a real eigenvalue λ, with a real unit eigenvector $v \in \mathbb{R}^n$, $\|v\| = 1$.

Let $V := \operatorname{span}\{v\}$, and let $V^\perp \subset \mathbb{R}^n$ be the orthogonal complement. Then $\dim V^\perp = n - 1$, and we claim A preserves $V^\perp$, i.e., $Aw \in V^\perp$ for all $w \in V^\perp$.

Indeed, for $w \in V^\perp$, we compute

$$\langle Aw, v\rangle = \langle w, Av\rangle = \langle w, \lambda v\rangle = \lambda\langle w, v\rangle = 0.$$

Thus $Aw \perp v$, so A maps $V^\perp$ into itself.

Let A' denote the restriction of A to $V^\perp$, i.e., $A' = A|_{V^\perp} \in \mathbb{R}^{(n-1)\times(n-1)}$. Since A is symmetric, A' is also symmetric with respect to the induced inner product.

By the induction hypothesis, A' admits an orthonormal eigenbasis $\{v_2, \dots, v_n\} \subset V^\perp$. Then $\{v, v_2, \dots, v_n\}$ is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A.

Step 4: Matrix formulation.

Let $Q = [v_1\, v_2\, \dots\, v_n] \in \mathbb{R}^{n\times n}$ be the orthogonal matrix whose columns are the orthonormal eigenvectors of A. Let $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ be the diagonal matrix of corresponding eigenvalues.

Then

$$AQ = Q\Lambda \quad \Rightarrow \quad Q^\top AQ = \Lambda,$$

and thus

$$A = Q\Lambda Q^\top.$$

This completes the proof. □

Example II.8 (Spectral Decomposition of a Symmetric Matrix)**.** Let

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

We aim to diagonalize A by computing its spectral decomposition:

$$A = Q\Lambda Q^\top,$$

where Q is orthogonal and Λ is diagonal with the eigenvalues of A.

Step 1: Verify symmetry. Note:

$$A^\top = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = A,$$

so A is symmetric. Thus, the Spectral Theorem applies: all eigenvalues are real and there exists an orthonormal basis of eigenvectors.

Step 2: Compute eigenvalues. Solve the characteristic equation:

$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda - 2 & -1 \\ -1 & \lambda - 2 \end{bmatrix} = (\lambda - 2)^2 - 1 = \lambda^2 - 4\lambda + 3.$$

Thus, the eigenvalues are

$$\lambda_1 = 1, \quad \lambda_2 = 3.$$

Step 3: Compute eigenvectors.

<u>For $\lambda = 1$</u>:

$$(A - I)v = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 0 \Rightarrow x + y = 0.$$

So a basis for this eigenspace is:

$$v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

For $\lambda = 3$:

$$(A - 3I)v = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \Rightarrow x - y = 0.$$

So a basis for this eigenspace is:

$$v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Step 4: Normalize eigenvectors.

We compute:

$$\|v_1\| = \sqrt{1^2 + (-1)^2} = \sqrt{2}, \quad \|v_2\| = \sqrt{1^2 + 1^2} = \sqrt{2}.$$

So the normalized eigenvectors are:

$$u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Step 5: Construct Q and Λ.

Let

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.$$

Then

$$Q^\top A Q = \Lambda.$$

Alternatively, verify:

$$A = Q\Lambda Q^\top.$$

Explicitly compute:

$$\begin{aligned} Q\Lambda &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{3}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{3}{\sqrt{2}} \end{bmatrix}, \\ Q\Lambda Q^\top &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{3}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{3}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = A. \end{aligned}$$

Thus, A admits an orthonormal diagonalization with real eigenvalues and mutually orthogonal eigenvectors, confirming the spectral theorem.

II.6 Quadratic Forms and Geometry

Quadratic forms provide a canonical way to study curvature, energy, and second-order variation in multivariable systems. They naturally arise in linear algebra through symmetric matrices, in calculus through Hessians, and in geometry through metric tensors.

Definition II.8 (Quadratic Form)**.** Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. A function $Q : \mathbb{R}^n \to \mathbb{R}$ is called a *quadratic form* if it has the expression

$$Q(x) = x^\top A x,$$

where $x \in \mathbb{R}^n$ is viewed as a column vector and $x^\top$ is its transpose.

Remark 20. The symmetry of A ensures that $x^\top Ax$ is always a scalar. Moreover, it guarantees that $Q(x)$ depends only on the symmetric part of any bilinear form, as

$$x^\top Ax = x^\top \left(\frac{A + A^\top}{2}\right) x.$$

Thus, only symmetric matrices correspond to genuine quadratic forms.

Example II.9 (Matrix Form and Expansion)**.** Let $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix} \in \mathbb{R}^{2\times 2}$, and let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Then:

$$Q(x) = x^\top Ax = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = ax_1^2 + 2bx_1x_2 + dx_2^2.$$

Quadratic forms generalize scalar-valued second-degree polynomials in multiple variables.

Definition II.9 (Level Set of a Quadratic Form)**.** Given a quadratic form $Q(x) = x^\top Ax$, the set

$$\{x \in \mathbb{R}^n : Q(x) = c\}$$

is called the *level set* at height c. The set $Q(x) = 1$ is particularly important in classifying the shape induced by A.

Example II.10 (Conic Sections via Quadratic Form)**.** Let $A \in \mathbb{R}^{2\times 2}$ be symmetric, and consider the level set $Q(x) = x^\top Ax = 1$. To classify the geometry of this curve, perform the following steps:

Step 1: Diagonalize A**.** Since A is symmetric, the spectral theorem ensures the existence of an orthogonal matrix $Q \in \mathbb{R}^{2\times 2}$ and a diagonal matrix $D = \operatorname{diag}(\lambda_1, \lambda_2)$ such that

$$A = QDQ^\top.$$

Step 2: Change of coordinates. Let $x = Qy$, so that $Q^\top x = y$. Then:

$$Q(x) = x^\top Ax = (Qy)^\top QDQ^\top Qy = y^\top Dy = \lambda_1 y_1^2 + \lambda_2 y_2^2.$$

Step 3: Analyze the sign of eigenvalues.

- If $\lambda_1 > 0$, $\lambda_2 > 0$, the level set is an ellipse.
- If $\lambda_1\lambda_2 < 0$, the level set is a hyperbola.
- If exactly one eigenvalue is zero, the level set degenerates to a pair of lines.

The eigenvalues of A determine the curvature type of the conic section.

Definition II.10 (Definiteness of a Quadratic Form)**.** Let $Q(x) = x^\top Ax$ be a quadratic form.

- Q is *positive definite* if $Q(x) > 0$ for all $x \neq 0$.
- Q is *negative definite* if $Q(x) < 0$ for all $x \neq 0$.
- Q is *indefinite* if it takes both positive and negative values.

Theorem II.9 (Sylvester's Criterion)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then A is positive definite if and only if all its leading principal minors are strictly positive:*

$$\det A_k > 0 \quad \textit{for all } k = 1, \ldots, n,$$

where A_k denotes the $k \times k$ leading principal submatrix of A, i.e.,

$$A_k = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}.$$

Remark 21 (Geometric Interpretation). The condition $\det A_k > 0$ ensures that the quadratic form $Q(x) = x^\top Ax$ restricts to a strictly convex function on every k-dimensional subspace generated by the first k standard basis vectors. This guarantees that the associated ellipsoids are properly oriented and non-degenerate in all coordinate-aligned subspaces, and hence that A induces a Riemannian inner product.

Example II.11 (Energy Form in Classical Mechanics). Let $m_1, m_2 > 0$ be masses of two decoupled particles moving along orthogonal spatial axes. Define:

$$A = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \quad x = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix},$$

where x is the velocity vector in $\mathbb{R}^2$. The kinetic energy K of the system is:

$$\begin{aligned} K(v) &= \frac{1}{2} x^\top Ax \\ &= \frac{1}{2} \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ &= \frac{1}{2}(m_1 v_1^2 + m_2 v_2^2). \end{aligned}$$

This function defines a strictly convex quadratic form:

- The matrix A is diagonal with $m_1, m_2 > 0$, so it is symmetric and positive definite.
- The energy functional $K(v)$ satisfies:

$$K(v) > 0 \quad \text{for all } v \neq 0, \qquad K(0) = 0.$$

- Its level sets $\{v \in \mathbb{R}^2 : K(v) = c\}$ are ellipses:

$$\frac{v_1^2}{2c/m_1} + \frac{v_2^2}{2c/m_2} = 1.$$

  These ellipses describe isoenergy contours, symmetric with respect to coordinate axes, representing equipartition in decoupled modes.

Connection to Sylvester's Criterion: We verify positive definiteness via principal minors:

$$\begin{aligned} \det A_1 &= m_1 > 0, \\ \det A_2 &= \det \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} = m_1 m_2 > 0. \end{aligned}$$

Hence, Sylvester's criterion confirms A is positive definite, and the kinetic energy is strictly positive except at rest.

Conclusion: This quadratic form encodes the stored motion energy in a mechanical system with two orthogonal degrees of freedom. The positivity of the energy functional reflects the convex geometry of kinetic energy and underlies stability and variational principles in classical mechanics.

Remark 22. Quadratic forms encode the second-order behavior of scalar-valued functions. In multivariable calculus, the Hessian matrix of a function f at a point x_0 is symmetric, and the sign of the quadratic form $h^\top H(x_0)h$ determines whether f has a local minimum, maximum, or saddle at x_0. This gives geometric interpretation to the second derivative test.

Remark 23. In Riemannian geometry, the metric tensor assigns to each point $p \in M$ a positive definite quadratic form on T_pM. The curvature and geodesic flow of the manifold are governed by the structure of these quadratic forms across the manifold.

II.7 Fourier and Spectral Basis Expansion

The decomposition of functions into orthogonal modes is one of the deepest ideas in analysis. In finite dimensions, the spectral theorem provides a basis of eigenvectors for symmetric matrices. In infinite dimensions, the appropriate generalization is a complete orthonormal set of eigenfunctions of a self-adjoint operator, such as the Laplacian. This leads naturally to the theory of Fourier series, orthogonal expansions, and spectral geometry.

Definition II.11 (Orthonormal Basis in Hilbert Space)**.** Let $\mathcal{H}$ be a Hilbert space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A countable set $\{e_n\}_{n=1}^{\infty} \subset \mathcal{H}$ is called an *orthonormal basis* if:

1. $\langle e_n, e_m \rangle = \delta_{nm}$ for all $n, m \in \mathbb{N}$,
2. Every $f \in \mathcal{H}$ can be written as

$$f = \sum_{n=1}^{\infty} \langle f, e_n \rangle e_n,$$

where the convergence is in the norm topology of $\mathcal{H}$.

Example II.12 (One type of Fourier basis)**.** Let $\mathcal{H} = L^2([0, 2\pi]; \mathbb{R})$ with inner product

$$\langle f, g \rangle = \int_0^{2\pi} f(x) g(x)\, dx.$$

The functions

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos(nx)}{\sqrt{\pi}}, \frac{\sin(nx)}{\sqrt{\pi}} \right\}_{n=1}^{\infty}$$

form a complete orthonormal basis for $L^2([0, 2\pi])$. Any $f \in L^2$ admits the expansion:

$$f(x) = \sum_{n=0}^{\infty} \left(a_n \cos(nx) + b_n \sin(nx) \right),$$

with convergence in the L^2-norm, where

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx)\, dx, \qquad b_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(nx)\, dx.$$

This expresses f as a linear combination of eigenfunctions of the negative Laplacian on the circle.

Theorem II.10 (Spectral Theorem for Compact Self-Adjoint Operators)**.** *Let $T : \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator on a separable Hilbert space. Then:*

1. *The spectrum of T consists of a countable set of real eigenvalues $\{\lambda_n\}$ with $\lambda_n \to 0$.*
2. *There exists an orthonormal basis $\{e_n\} \subset \mathcal{H}$ consisting of eigenvectors of T.*
3. *For every $f \in \mathcal{H}$,*

$$Tf = \sum_{n=1}^{\infty} \lambda_n \langle f, e_n \rangle e_n.$$

Remark 24. The classical Fourier basis arises from applying the above theorem to the operator $T = -\frac{d^2}{dx^2}$ with periodic boundary conditions. The functions $\{1, \cos(nx), \sin(nx)\}$ are eigenfunctions with eigenvalues $0, n^2, n^2$, respectively. Spectral expansions generalize Fourier analysis by encoding structure through the eigenfunctions of geometric or analytic operators.

Definition II.12 (Laplace Operator on the Circle)**.** Define $\Delta = \frac{d^2}{dx^2}$ acting on $C^\infty_{\text{per}}([0, 2\pi])$. Then $-\Delta$ is positive semi-definite and self-adjoint on $L^2([0, 2\pi])$, with eigenfunctions:

$$\phi_0(x) = \frac{1}{\sqrt{2\pi}}, \quad \phi_n^{\cos}(x) = \frac{\cos(nx)}{\sqrt{\pi}}, \quad \phi_n^{\sin}(x) = \frac{\sin(nx)}{\sqrt{\pi}}, \quad n \geq 1,$$

and corresponding eigenvalues $0, n^2, n^2$. The Fourier expansion is then a spectral decomposition of f in terms of $-\Delta$-eigenfunctions.

Remark 25. This interpretation gives a geometric reading: the oscillatory modes of a vibrating circular string are exactly the Laplacian eigenfunctions. The frequency squared corresponds to the eigenvalue. This bridges differential geometry, spectral theory, and classical mechanics.

Example II.13 (Heat Equation on the Circle)**.** Consider the PDE

$$\partial_t u = \Delta u, \quad u(x, 0) = f(x), \quad u(x + 2\pi, t) = u(x, t).$$

Expanding f in the Fourier basis:

$$f(x) = \sum_{n=0}^{\infty} \left(a_n \cos(nx) + b_n \sin(nx)\right),$$

the solution becomes:

$$u(x, t) = \sum_{n=0}^{\infty} \left(a_n e^{-n^2 t} \cos(nx) + b_n e^{-n^2 t} \sin(nx)\right).$$

Each frequency mode decays exponentially in time, with rate governed by its Laplacian eigenvalue.

Remark 26. This illustrates the power of spectral basis expansions. The infinite-dimensional operator is diagonalized by the Fourier basis, and the PDE reduces to a countable family of scalar ODEs. The structure of the solution is encoded entirely in the spectrum.

Remark 27. In geometric analysis, the spectrum of the Laplace–Beltrami operator on a compact Riemannian manifold encodes deep information about volume, curvature, and topology. The study of whether one can recover geometry from spectrum is captured in the question: "Can one hear the shape of a drum?"

II.8 Functional Calculus for the Laplacian on the Circle

Theorem II.11 (Functional Calculus for the Laplacian on the Circle)**.** *Let $\Delta = \frac{d^2}{dx^2}$ be the Laplacian on $L^2([0, 2\pi])$ with periodic boundary conditions. Then:*

1. *The operator $-\Delta$ is unbounded, densely defined, and self-adjoint on its domain*

$$\mathcal{D}(-\Delta) = \left\{ f \in L^2([0, 2\pi]) \,\middle|\, f \in C^\infty_{\text{per}}, \sum_{n=1}^{\infty} n^4 \left(|a_n|^2 + |b_n|^2\right) < \infty \right\},$$

where $f(x) = \sum_{n=0}^{\infty} \left(a_n \cos(nx) + b_n \sin(nx)\right)$ is the Fourier expansion of f.

2. *The Fourier basis $\{\phi_n\}$ given by*

$$\phi_0(x) = \frac{1}{\sqrt{2\pi}}, \quad \phi_n^{\cos}(x) = \frac{\cos(nx)}{\sqrt{\pi}}, \quad \phi_n^{\sin}(x) = \frac{\sin(nx)}{\sqrt{\pi}}, \quad n \geq 1,$$

forms a complete orthonormal set of eigenfunctions of $-\Delta$, with eigenvalues

$$-\Delta\phi_0 = 0, \quad -\Delta\phi_n^{\cos} = n^2\phi_n^{\cos}, \quad -\Delta\phi_n^{\sin} = n^2\phi_n^{\sin}.$$

3. *For any bounded Borel function $F : [0, \infty) \to \mathbb{C}$, the operator $F(-\Delta)$ is defined by*

$$F(-\Delta)f := \sum_{n=0}^{\infty} F(\lambda_n)\langle f, \phi_n\rangle \phi_n,$$

where the eigenvalues $\lambda_0 = 0$, $\lambda_n = n^2$ for $n \geq 1$, and the sum converges in the L^2-norm. This defines a bounded self-adjoint operator on $L^2([0, 2\pi])$, known as the spectral functional calculus of $-\Delta$.

4. *In particular, the heat semigroup is given by*

$$e^{t\Delta} f = \sum_{n=0}^{\infty} e^{-t\lambda_n}\langle f, \phi_n\rangle \phi_n,$$

which solves the initial value problem

$$\partial_t u = \Delta u, \quad u(x, 0) = f(x).$$

Similarly, the unitary Schrödinger propagator is

$$e^{it\Delta} f = \sum_{n=0}^{\infty} e^{-it\lambda_n}\langle f, \phi_n\rangle \phi_n,$$

which solves

$$i\partial_t u = -\Delta u, \quad u(x, 0) = f(x).$$

Proof. We prove each part in sequence.

(1) Self-adjointness and Domain. The operator $\Delta = \frac{d^2}{dx^2}$ with periodic boundary conditions is symmetric on the dense domain $C^{\infty}_{\text{per}}([0, 2\pi]) \subset L^2([0, 2\pi])$. For $f, g \in C^{\infty}_{\text{per}}$,

$$\langle \Delta f, g\rangle = \int_0^{2\pi} f''(x)g(x)\,dx = \int_0^{2\pi} f(x)g''(x)\,dx = \langle f, \Delta g\rangle,$$

using integration by parts and periodicity to cancel boundary terms. This symmetry, together with essential self-adjointness of Δ on this domain, implies that the closure $-\Delta$ is self-adjoint.

The Sobolev space H^2_{per} is the natural domain for $-\Delta$, but in terms of Fourier series, we require $f \in L^2$ with

$$\sum_{n=1}^{\infty} n^4(|a_n|^2 + |b_n|^2) < \infty,$$

which ensures $f'' \in L^2$, so $f \in \mathcal{D}(-\Delta)$.

(2) Orthonormal Basis and Eigenvalues. Each of the functions $\phi_n^{\cos}(x) = \frac{\cos(nx)}{\sqrt{\pi}}$, $\phi_n^{\sin}(x) = \frac{\sin(nx)}{\sqrt{\pi}}$, and $\phi_0 = \frac{1}{\sqrt{2\pi}}$ lies in $L^2([0, 2\pi])$, and they are mutually orthonormal. Indeed,

$$\langle \cos(nx), \cos(mx)\rangle = \begin{cases} 0 & n \neq m, \\ \pi & n = m, \end{cases} \qquad \langle \cos(nx), \sin(mx)\rangle = 0,$$

and similarly for sine terms.

Furthermore,

$$-\Delta \phi_n^{\cos}(x) = \frac{d^2}{dx^2}\left(\frac{\cos(nx)}{\sqrt{\pi}}\right) = n^2 \frac{\cos(nx)}{\sqrt{\pi}} = n^2 \phi_n^{\cos}(x),$$

and likewise for sine. Thus, these are eigenfunctions with eigenvalue n^2, and the constant function is the eigenfunction for eigenvalue zero.

(3) Spectral Representation. Since $-\Delta$ is self-adjoint and has compact resolvent (equivalent to the embedding $H^2 \hookrightarrow L^2$ being compact), the spectral theorem applies. We may write:

$$f = \sum_{n=0}^{\infty} \langle f, \phi_n \rangle \phi_n,$$

with convergence in L^2. For any bounded Borel function F, we define:

$$F(-\Delta) f := \sum_{n=0}^{\infty} F(\lambda_n) \langle f, \phi_n \rangle \phi_n,$$

which converges in L^2 by Parseval's identity:

$$\|F(-\Delta) f\|_{L^2}^2 = \sum_{n=0}^{\infty} |F(\lambda_n)|^2 |\langle f, \phi_n \rangle|^2 \leq \|F\|_\infty^2 \|f\|_{L^2}^2.$$

So $F(-\Delta)$ is a bounded self-adjoint operator.

(4) Heat and Schrödinger Equations. Taking $F(\lambda) = e^{-t\lambda}$ yields:

$$e^{t\Delta} f := \sum_{n=0}^{\infty} e^{-t\lambda_n} \langle f, \phi_n \rangle \phi_n,$$

and since each eigenfunction solves $\partial_t \phi_n(t) = -\lambda_n \phi_n(t)$, the full solution

$$u(x,t) = \sum_{n=0}^{\infty} e^{-t\lambda_n} \langle f, \phi_n \rangle \phi_n(x)$$

solves $\partial_t u = \Delta u$, with $u(x,0) = f(x)$. Similarly, for the Schrödinger equation, we take $F(\lambda) = e^{-it\lambda}$, and the same analysis applies:

$$u(x,t) = \sum_{n=0}^{\infty} e^{-it\lambda_n} \langle f, \phi_n \rangle \phi_n(x)$$

solves $i\partial_t u = -\Delta u$. In both cases, the differential equation reduces to scalar multiplication in the spectral basis.

□

II.9 Preview: Applications and Dynamics

- Gram matrix and least-squares: $A^\top A$ symmetric, positive semi-definite.
- Principal Component Analysis (PCA): eigenvectors of covariance matrix.
- Vibrations and dynamics: diagonalizing stiffness or inertia matrices.
- PDEs: solving via eigenfunction expansion.

Example II.14 (Linear ODE System)**.** Let $\frac{\mathrm{d}x}{\mathrm{d}t} = Ax$ with $A = Q\Lambda Q^\top$. Then:

$$x(t) = Q e^{\Lambda t} Q^\top x(0),$$

so dynamics decouple in eigenbasis coordinates.

Remark 28. This structure, the spectral decomposition of self-adjoint operators, will recur in every advanced mathematical context. From Schrödinger operators to kernel machines, from Riemannian geometry to signal processing, it provides the framework for mode decomposition and stability analysis.

III Multivariable Calculus and Tensor Fields

III.1 Smooth Vector Fields

Definition III.1 (Smooth Vector Field)**.** A smooth vector field on a manifold M is a smooth section X of the tangent bundle TM, i.e., a map

$$X : M \to TM$$

such that $X(p) \in T_pM$ for all $p \in M$, and in local coordinates $X = X^i(x)\frac{\partial}{\partial x^i}$ where the component functions X^i are smooth.

Definition III.2 (Smooth Covector Field (1-Form))**.** A smooth covector field, or 1-form, on M is a smooth section ω of the cotangent bundle T^*M, i.e., a map

$$\omega : M \to T^*M$$

such that $\omega(p) \in T_p^*M$ for all $p \in M$, and in local coordinates $\omega = \omega_i(x)dx^i$ with smooth component functions ω_i.

Definition III.3 (Tensor Field of Type (k, ℓ))**.** A (k, ℓ)-tensor field on M is a smooth section of the tensor bundle

$$T^{(k,\ell)}M = \underbrace{TM \otimes \cdots \otimes TM}_{k \text{ times}} \otimes \underbrace{T^*M \otimes \cdots \otimes T^*M}_{\ell \text{ times}}.$$

At each point $p \in M$, it assigns a tensor $T_p \in T_p^{(k,\ell)}M$, a multilinear map as in the previous definition.

Example III.1 (Volume Form)**.** On an oriented n-dimensional manifold M, a volume form is a nowhere-vanishing smooth $(0, n)$-tensor field

$$\omega \in \Gamma(\Lambda^n T^*M),$$

where $\Lambda^n T^*M$ denotes the top exterior power of the cotangent bundle. Locally, $\omega = f(x)dx^1 \wedge \cdots \wedge dx^n$ with $f(x) > 0$.

Theorem III.1 (Tensor Contraction)**.** *Let $T \in \Gamma(T^{(k,\ell)}M)$ be a (k, ℓ)-tensor field. A contraction of T is a $(k-1, \ell-1)$-tensor field obtained by pairing one contravariant and one covariant index:*

$$\mathrm{contr}_{i,j}(T)(\cdots) = T^{a_1\cdots a_k}_{b_1\cdots b_\ell}\delta^{a_i}_{b_j}\cdots.$$

This operation is bilinear and commutes with smooth pullbacks under diffeomorphisms.

Proof. We aim to show that contraction over one contravariant and one covariant index of a smooth (k, ℓ)-tensor field $T \in \Gamma(T^{(k,\ell)}M)$ yields a well-defined smooth $(k-1, \ell-1)$-tensor field. The argument proceeds in several steps.

1. **Pointwise definition.** Let $p \in M$ be fixed. A tensor $T_p \in T_p^{(k,\ell)}M$ is a multilinear map

$$T_p : \underbrace{T_p^*M \times \cdots \times T_p^*M}_{k \text{ times}} \times \underbrace{T_pM \times \cdots \times T_pM}_{\ell \text{ times}} \to \mathbb{R}.$$

Choose positions $i \in \{1, \dots, k\}$ and $j \in \{1, \dots, \ell\}$. Define the contraction $\mathrm{contr}_{i,j}(T)_p \in T_p^{(k-1,\ell-1)}M$ by inserting a dual basis pair (e^a, e_a) into those positions and summing over a:

$$\mathrm{contr}_{i,j}(T)_p = \sum_{a=1}^{n} T_p(\dots, e^a, \dots; \dots, e_a, \dots),$$

where e^a is placed at slot i, and e_a at slot j. This expression is independent of basis since it uses the natural dual pairing.

2. **Coordinate expression.** In a coordinate chart $(x^1, \dots, x^n)$, a (k, ℓ)-tensor field has components
$$T = T^{i_1 \dots i_k}_{j_1 \dots j_\ell}(x)\, \partial_{i_1} \otimes \cdots \otimes \partial_{i_k} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_\ell}.$$
Contraction over the i-th contravariant and j-th covariant slot yields
$$(\mathrm{contr}_{i,j}(T))^{\hat{i}_1 \dots \hat{i}_{k-1}}_{\hat{j}_1 \dots \hat{j}_{\ell-1}} = \sum_{a=1}^{n} T^{i_1 \dots a \dots i_k}_{j_1 \dots a \dots j_\ell},$$
where hats indicate omission of contracted indices. The resulting components are smooth because they are obtained via smooth operations on the original components.

3. $C^\infty(M)$**-linearity.** Let $f \in C^\infty(M)$. Then
$$\mathrm{contr}_{i,j}(fT) = f \cdot \mathrm{contr}_{i,j}(T),$$
because contraction is a pointwise linear operation. This ensures that $\mathrm{contr}_{i,j}(T)$ is again a smooth tensor field.

4. **Coordinate invariance.** Under a coordinate change $x^i \mapsto x^{i'}$, tensor components transform according to
$$T^{i_1 \dots i_k}_{j_1 \dots j_\ell} \mapsto \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial x^{i'_k}}{\partial x^{i_k}} \frac{\partial x^{j_1}}{\partial x^{j'_1}} \cdots \frac{\partial x^{j_\ell}}{\partial x^{j'_\ell}} T^{i_1 \dots i_k}_{j_1 \dots j_\ell}.$$
In the contraction sum $\sum_a T^{\dots a \dots}_{\dots a \dots}$, the transformation Jacobians cancel due to one upper and one lower index being contracted. Therefore, the result transforms as a $(k-1, \ell-1)$-tensor.

5. **Conclusion.** All properties of a smooth tensor field are preserved: multilinearity, smoothness, and coordinate transformation behavior. Thus, $\mathrm{contr}_{i,j}(T) \in \Gamma(T^{(k-1,\ell-1)}M)$ is well-defined.

□

III.2 Multilinear Maps and Tensors

Definition III.4 (Multilinear Map)**.** Let $V_1, \dots, V_k, W$ be vector spaces over a field $\mathbb{F}$. A function
$$T : V_1 \times \cdots \times V_k \to W$$
is called a *k-linear map* (or *tensor of type* $(k, 0)$) if T is linear in each argument separately.

Example III.2 (Bilinear Form)**.** The inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is a symmetric bilinear map.

Definition III.5 (Tensor Product)**.** Given $T_1 \in \mathsf{Lin}(V, \mathbb{R})$ and $T_2 \in \mathsf{Lin}(W, \mathbb{R})$, the tensor product
$$T_1 \otimes T_2 \in \mathsf{Bilin}(V \times W, \mathbb{R})$$
is defined by
$$(T_1 \otimes T_2)(v, w) = T_1(v) \cdot T_2(w).$$

Theorem III.2 (Construction of General Tensors and Index Manipulation)**.** *Let V be a finite-dimensional real vector space with basis $\{e_i\}_{i=1}^n$ and dual basis $\{e^i\}_{i=1}^n$. Let $g : V \times V \to \mathbb{R}$ be a symmetric non-degenerate bilinear form (metric). Then:*

- *Every (k,ℓ)-tensor $T \in \mathsf{T}^{(k,\ell)}(V)$ can be expressed as a finite linear combination:*

$$T = T^{i_1 \dots i_k}_{j_1 \dots j_\ell}\, e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e^{j_1} \otimes \cdots \otimes e^{j_\ell}.$$

- *The metric g induces canonical isomorphisms:*

$$\flat : V \to V^*, \quad v \mapsto g(v, -), \quad \textit{and its inverse} \quad \sharp : V^* \to V.$$

- *These isomorphisms allow for the* raising and lowering of indices*, converting covariant to contravariant components and vice versa.*

Proof. Let $\dim V = n$. The tensor algebra $T^{(k,\ell)}(V)$ consists of multilinear maps:

$$T : \underbrace{V^* \times \cdots \times V^*}_{k \text{ times}} \times \underbrace{V \times \cdots \times V}_{\ell \text{ times}} \to \mathbb{R}.$$

Let $\{e_i\}$ be a basis for V and $\{e^i\}$ the dual basis for V^*. Then any $T \in T^{(k,\ell)}(V)$ can be expanded in the basis:

$$\left\{ e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e^{j_1} \otimes \cdots \otimes e^{j_\ell} \right\}.$$

By multilinearity, we write:

$$T = \sum_{i_1,\dots,i_k,j_1,\dots,j_\ell} T^{i_1 \dots i_k}_{j_1 \dots j_\ell}\, e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e^{j_1} \otimes \cdots \otimes e^{j_\ell},$$

where the components $T^{i_1 \dots i_k}_{j_1 \dots j_\ell}$ are real scalars.

Now let $g : V \times V \to \mathbb{R}$ be a symmetric bilinear form with matrix representation $g_{ij} = g(e_i, e_j)$, assumed invertible. The induced map $\flat : V \to V^*$ is defined by:

$$v = v^i e_i \Rightarrow v^\flat = g_{ij} v^i e^j.$$

The inverse isomorphism $\sharp : V^* \to V$ satisfies:

$$\alpha = \alpha_i e^i \Rightarrow \alpha^\sharp = g^{ij} \alpha_i e_j,$$

where (g^{ij}) is the inverse of (g_{ij}).

These maps allow us to convert between covariant and contravariant components:

$$T_i \Rightarrow T^j = g^{ij} T_i \quad \text{(raise index)}, \tag{42}$$

$$T^i \Rightarrow T_j = g_{ij} T^i \quad \text{(lower index)}. \tag{43}$$

For higher tensors, the index manipulation is performed by contracting with g_{ij} or g^{ij} in the appropriate slots. For example, for a $(1,1)$ tensor $T = T^i_j\, e_i \otimes e^j$, we can define:

$$(T^\flat)^i_k = g_{kj} T^i_j, \qquad (T^\sharp)^j_\ell = g^{ij} T_{i\ell}.$$

Since g is non-degenerate, these operations are isomorphisms, preserving the tensor structure while shifting index types.

Hence, the tensor space is generated as claimed, and index conversion is canonically defined via the metric structure. $\square$

Theorem III.3 (Canonical Isomorphism between Vectors and Covectors)**.** *Let V be a finite-dimensional real inner product space with inner product $\langle \cdot, \cdot \rangle$. Then the mapping*

$$\flat : V \to V^*, \quad v \mapsto \langle v, \cdot \rangle$$

is a linear isomorphism. Its inverse $\sharp : V^* \to V$ is defined by the Riesz Representation Theorem and satisfies

$$\forall \omega \in V^*, \quad \langle \omega^\sharp, w \rangle = \omega(w) \quad \textit{for all } w \in V.$$

Proof. Let V be an n-dimensional real inner product space with inner product $\langle \cdot, \cdot \rangle$. Define the *musical isomorphism*

$$\flat : V \to V^*, \quad \mathbf{v} \mapsto \langle \mathbf{v}, \cdot \rangle.$$

This map assigns to each vector $\mathbf{v} \in V$ a covector $\mathbf{v}^\flat \in V^*$, defined pointwise by

$$\mathbf{v}^\flat(\mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle \quad \text{for all } \mathbf{w} \in V.$$

Linearity: For all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and scalars $a, b \in \mathbb{R}$, we compute

$$(a\mathbf{v}_1 + b\mathbf{v}_2)^\flat(\mathbf{w}) = \langle a\mathbf{v}_1 + b\mathbf{v}_2, \mathbf{w} \rangle = a \langle \mathbf{v}_1, \mathbf{w} \rangle + b \langle \mathbf{v}_2, \mathbf{w} \rangle = a\, \mathbf{v}_1^\flat(\mathbf{w}) + b\, \mathbf{v}_2^\flat(\mathbf{w}),$$

so the map $\flat$ is linear.

Injectivity: Suppose $\mathbf{v}^\flat = 0$, i.e., $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{w} \in V$. In particular, taking $\mathbf{w} = \mathbf{v}$, we obtain

$$\langle \mathbf{v}, \mathbf{v} \rangle = 0.$$

Since the inner product is positive definite, this implies $\mathbf{v} = \mathbf{0}$. Hence $\ker \flat = \{0\}$, and $\flat$ is injective.

Dimension Count: The vector space V has $\dim V = n$, and its dual space V^* also satisfies $\dim V^* = n$. Since $\flat$ is a linear map from V to V^*, and is injective with equal dimensions, it is an isomorphism:

$$\flat : V \xrightarrow{\sim} V^*.$$

Coordinate Expression: Let $\{\mathbf{e}_i\}_{i=1}^n$ be an orthonormal basis for V, so $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$. Then any vector $\mathbf{v} \in V$ has coordinates

$$\mathbf{v} = \sum_{i=1}^n v^i \mathbf{e}_i,$$

and we compute the action of $\mathbf{v}^\flat \in V^*$ on $\mathbf{w} = \sum_{j=1}^n w^j \mathbf{e}_j$ as

$$\mathbf{v}^\flat(\mathbf{w}) = \left\langle \sum_i v^i \mathbf{e}_i, \sum_j w^j \mathbf{e}_j \right\rangle = \sum_{i,j} v^i w^j \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \sum_i v^i w^i.$$

Thus, $\mathbf{v}^\flat$ has the coordinate representation $(v^1, \ldots, v^n) \in V^*$, i.e., the same components in this orthonormal basis.

Inverse Map: The inverse map $\sharp : V^* \to V$ satisfies, for any $\boldsymbol{\omega} \in V^*$ and $\mathbf{v} \in V$,

$$\langle \boldsymbol{\omega}^\sharp, \mathbf{v} \rangle = \boldsymbol{\omega}(\mathbf{v}).$$

This uniquely defines $\boldsymbol{\omega}^\sharp \in V$ by Riesz representation. In coordinates, if $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n) \in V^*$, then

$$\boldsymbol{\omega}^\sharp = \sum_{i=1}^n \omega_i \mathbf{e}_i \in V.$$

Conclusion: The flat map $\flat : V \to V^*$ is a linear isomorphism, with inverse $\sharp : V^* \to V$, satisfying the dual identities:

$$(\mathbf{v}^\flat)^\sharp = \mathbf{v}, \qquad (\boldsymbol{\omega}^\sharp)^\flat = \boldsymbol{\omega}.$$

This completes the construction of the musical isomorphisms between vectors and covectors on an inner product space. □

III.3 Jacobian and Multivariable Derivatives

Definition III.6 (Total Derivative)**.** Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function, and let $x_0 \in \mathbb{R}^n$. We say that f is *differentiable at* x_0 if there exists a linear map

$$Df(x_0) : \mathbb{R}^n \to \mathbb{R}^m$$

such that

$$\lim_{h\to 0} \frac{\|f(x_0+h) - f(x_0) - Df(x_0)(h)\|}{\|h\|} = 0.$$

That is, the error term

$$\varepsilon(h) := f(x_0+h) - f(x_0) - Df(x_0)(h)$$

satisfies $\|\varepsilon(h)\| = o(\|h\|)$ as $h \to 0$. In this case, the linear map $Df(x_0)$ is called the *total derivative* (or *Fréchet derivative*) of f at x_0.

Coordinate Representation: If $f = (f^1, \dots, f^m)$, then the total derivative is represented by the Jacobian matrix:

$$Df(x_0) = \begin{bmatrix} \frac{\partial f^1}{\partial x^1} & \cdots & \frac{\partial f^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x^1} & \cdots & \frac{\partial f^m}{\partial x^n} \end{bmatrix}_{x=x_0},$$

and for any $h \in \mathbb{R}^n$,

$$Df(x_0)(h) = \frac{d}{dt} f(x_0 + th)\Big|_{t=0}.$$

Remark 29. The total derivative $Df(x_0)$ provides the best linear approximation to f near x_0. That is, in local coordinates:

$$f(x_0+h) \approx f(x_0) + Df(x_0)[h].$$

Theorem III.4 (Jacobian as the Matrix Representation of the Total Derivative)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at x_0. Then the total derivative $Df(x_0)$ is given in coordinates by the Jacobian matrix $J_f(x_0) \in \mathbb{R}^{m\times n}$:*

$$Df(x_0)[h] = J_f(x_0) \cdot h,$$

where $h \in \mathbb{R}^n$ and the Jacobian entries are

$$(J_f(x_0))_{ij} = \frac{\partial f^i}{\partial x^j}(x_0).$$

Proof. Let $f = (f^1, \dots, f^m)$ and fix $x_0 \in \mathbb{R}^n$. For small $h \in \mathbb{R}^n$, by the definition of differentiability,

$$f(x_0+h) = f(x_0) + Df(x_0)[h] + o(\|h\|).$$

We now compute $Df(x_0)$ explicitly using the directional derivatives.

Let e_j denote the j-th standard basis vector in $\mathbb{R}^n$. Then the j-th column of the Jacobian is

$$Df(x_0)[e_j] = \left(\frac{\partial f^1}{\partial x^j}(x_0), \ldots, \frac{\partial f^m}{\partial x^j}(x_0) \right)^\top .$$

Linearity then implies that for any $h = \sum_{j=1}^n h^j e_j \in \mathbb{R}^n$,

$$Df(x_0)[h] = \sum_{j=1}^n h^j Df(x_0)[e_j] = J_f(x_0) \cdot h.$$

Thus the Jacobian $J_f(x_0)$ gives the coordinate representation of the linear map $Df(x_0)$ with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. □

Remark 30. The Jacobian plays the role of the *pushforward* of tangent vectors under f:

$$df_{x_0} : T_{x_0}\mathbb{R}^n \to T_{f(x_0)}\mathbb{R}^m, \quad v \mapsto J_f(x_0)v.$$

This interpretation becomes crucial when generalizing to differentiable maps between manifolds.

Example III.3 (Linear Map Case)**.** If $f(x) = Ax$ for some $A \in \mathbb{R}^{m\times n}$, then $Df(x_0) = A$ for all x_0, and the Jacobian is constant:

$$J_f(x) = A.$$

This highlights that linear maps are their own total derivatives.

III.4 Differential Forms and Pullbacks

Definition III.7 (Differential 1-Form)**.** Let $U \subset \mathbb{R}^n$ be open. A differential 1-form on U is a smooth assignment

$$\omega : x \mapsto \omega_x \in \mathsf{Lin}(\mathbb{R}^n, \mathbb{R}),$$

often written as $\omega = \sum_{i=1}^n a_i(x)\, dx^i$.

Definition III.8 (Pullback of a 1-Form)**.** Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be smooth and let ω be a 1-form on $\mathbb{R}^m$. Then the pullback $f^*\omega$ is the 1-form on $\mathbb{R}^n$ defined by

$$(f^*\omega)_x(v) = \omega_{f(x)}(Df(x)[v]).$$

III.5 Tangent and Cotangent Spaces

Definition III.9 (Tangent Space at a Point)**.** Let M be a smooth manifold and $p \in M$. The tangent space T_pM is the set of all derivations at p, i.e., linear maps $D : C^\infty(M) \to \mathbb{R}$ satisfying Leibniz's rule:

$$D(fg) = f(p)D(g) + g(p)D(f).$$

Definition III.10 (Cotangent Space)**.** The cotangent space T_p^*M is the dual space of T_pM, consisting of all linear functionals $\alpha : T_pM \to \mathbb{R}$.

III.6 Tensor Fields and Covariant Structure

Definition III.11 (Tensor Field)**.** A (k, ℓ)-tensor field on a manifold M assigns to each point $p \in M$ a multilinear map

$$T_p : \underbrace{T_p^*M \times \cdots \times T_p^*M}_{k \text{ times}} \times \underbrace{T_pM \times \cdots \times T_pM}_{\ell \text{ times}} \to \mathbb{R}$$

that varies smoothly with p.

Example III.4 (Metric Tensor)**.** A Riemannian metric is a smooth $(0, 2)$-tensor field g such that g_p is a positive-definite inner product on T_pM.

Theorem III.5 (Coordinate Transformation Law for Tensors)**.** *Let T be a tensor field. Under a change of coordinates $x^i \mapsto x^{i'}$, the components transform by*

$$T^{i'_1\cdots i'_k}_{j'_1\cdots j'_\ell} = \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial x^{i'_k}}{\partial x^{i_k}} \frac{\partial x^{j_1}}{\partial x^{j'_1}} \cdots \frac{\partial x^{j_\ell}}{\partial x^{j'_\ell}} T^{i_1\cdots i_k}_{j_1\cdots j_\ell}.$$

Proof. Let $T \in \Gamma(T^{(k,\ell)}M)$ be a smooth (k, ℓ)-tensor field on a manifold M, and let (x^i) and $(x^{i'})$ be two overlapping coordinate charts on M. We aim to determine how the components of T transform under the change of coordinates $x^i \mapsto x^{i'}$.

1. **Local basis transformation.** The tangent and cotangent basis vectors transform under a change of coordinates as follows:

$$\frac{\partial}{\partial x^i} = \frac{\partial x^{j'}}{\partial x^i}\frac{\partial}{\partial x^{j'}}, \qquad dx^i = \frac{\partial x^i}{\partial x^{j'}} dx^{j'}.$$

   These follow from the chain rule and define the Jacobian matrices for the transformation.

2. **Tensor decomposition in local coordinates.** In the (x^i) chart, the tensor field can be expressed locally as

$$T = T^{i_1\cdots i_k}_{j_1\cdots j_\ell}(x)\, \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_k}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_\ell}.$$

   Similarly, in the $(x^{i'})$ chart, the tensor field must be expressible as

$$T = T^{i'_1\cdots i'_k}_{j'_1\cdots j'_\ell}(x')\, \frac{\partial}{\partial x^{i'_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i'_k}} \otimes dx^{j'_1} \otimes \cdots \otimes dx^{j'_\ell}.$$

3. **Apply basis transformation to the expression for T.** Substitute the coordinate transformation of basis vectors and 1-forms into the original expression for T:

$$\frac{\partial}{\partial x^{i_r}} = \frac{\partial x^{i'_r}}{\partial x^{i_r}}\frac{\partial}{\partial x^{i'_r}}, \qquad dx^{j_s} = \frac{\partial x^{j_s}}{\partial x^{j'_s}} dx^{j'_s}.$$

   Applying this for each of the k vector indices and ℓ covector indices:

$$T = T^{i_1\cdots i_k}_{j_1\cdots j_\ell}\left(\prod_{r=1}^{k} \frac{\partial x^{i'_r}}{\partial x^{i_r}}\frac{\partial}{\partial x^{i'_r}}\right) \otimes \left(\prod_{s=1}^{\ell} \frac{\partial x^{j_s}}{\partial x^{j'_s}} dx^{j'_s}\right).$$

   By multilinearity of the tensor product:

$$T = \left(T^{i_1\cdots i_k}_{j_1\cdots j_\ell} \prod_{r=1}^{k} \frac{\partial x^{i'_r}}{\partial x^{i_r}} \prod_{s=1}^{\ell} \frac{\partial x^{j_s}}{\partial x^{j'_s}}\right) \cdot \left(\frac{\partial}{\partial x^{i'_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i'_k}} \otimes dx^{j'_1} \otimes \cdots \otimes dx^{j'_\ell}\right).$$

4. **Extract transformed components.** Comparing with the coordinate representation of T in the primed chart, we identify:

$$T^{i'_1\cdots i'_k}_{j'_1\cdots j'_\ell} = \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial x^{i'_k}}{\partial x^{i_k}} \cdot \frac{\partial x^{j_1}}{\partial x^{j'_1}} \cdots \frac{\partial x^{j_\ell}}{\partial x^{j'_\ell}} \cdot T^{i_1\cdots i_k}_{j_1\cdots j_\ell}.$$

5. **Conclusion.** The coordinate components of a (k, ℓ)-tensor transform under a change of chart by:

$$T^{i'_1\cdots i'_k}_{j'_1\cdots j'_\ell} = \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial x^{i'_k}}{\partial x^{i_k}} \cdot \frac{\partial x^{j_1}}{\partial x^{j'_1}} \cdots \frac{\partial x^{j_\ell}}{\partial x^{j'_\ell}} \cdot T^{i_1\cdots i_k}_{j_1\cdots j_\ell},$$

   as claimed.

□