# MAT237 Notes

Author

January 7, 2026

## Contents

# §1 Day 1: Administrative Stuff (Sept 2, 2025)

Everything is in the syllabus, but we did play with some blocks! There's many different ways to visualize the same thing. Went over classroom norms and whatnot, and then looked at syllabus, no math content today.

About these notes: I keep track of things mentioned in lecture. You can likely find more detail in the textbook.

# §2 Day 2: Speed and Velocity (Sept 4, 2025)

We want to differentiate and integrate functions $A \to B$, where $A \subseteq \mathbb{R}^m$, $B \subseteq \mathbb{R}^n$. Today we study the case where $m = 1$, so functions $A \subseteq \mathbb{R} \to \mathbb{R}^n$. Today, will mainly look at functions $\mathbb{R} \to \mathbb{R}^n$, as having a single parameter makes them much easier to work with.

Recall that distance is a scalar quantity, while velocity is a vector, meaning it has both magnitude and direction. Have $f : \mathbb{R} \to \mathbb{R}^n$ model some particle's position, and $|| \cdot ||$ be the euclidean norm. The average speed[1] over the time interval $t_1$ and $t_2$ ($t_1 < t_2$) is given by

$$\frac{||f(t_2) - f(t_1)||}{t_2 - t_1}$$

The instantaneous speed at time $t$ is given by

$$\lim_{h \to 0} \left|\left| \frac{f(t+h) - f(t)}{h} \right|\right|$$

The average velocity between $t_1$ and $t_2$ is given by

$$\frac{f(t_2) - f(t_1)}{t_2 - t_1}$$

and the instantaneous velocity at $t$ by

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

**Theorem 2.1** (Absolute Homogeneity of Euclidean Norm)**.** The euclidean norm has the *absolute homogeneity* property: For all $v = (v_1, \cdots, v_n) \in \mathbb{R}^n$, a scalar $\lambda$, have $||\lambda v|| = |\lambda| \, ||v||$

2.1 and various properties of norms were not mentioned in class.[2]

*Proof.*

$$\begin{aligned} ||\lambda v|| &= \left|\left| \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{pmatrix} \right|\right| \\ &= \sqrt{\lambda^2 {v_1}^2 + \cdots + \lambda^2 {v_n}^2} \\ &= \sqrt{\lambda^2({v_1}^2 + \cdots + {v_n}^2)} \\ &= |\lambda| \sqrt{{v_1}^2 + \cdots + {v_n}^2)} \\ &= |\lambda| \, ||v|| \end{aligned}$$

□

[1] I don't believe we are concerned with the 'actual' speed over time in this course (as in the physics definition), since that would involve finding the arc length and even more trouble. Just assume it means 'magnitude of the displacement vector'.

[2] I included this for completeness because people didn't believe that $\frac{||\gamma(6+h) - \gamma(6)||}{|h|}$ and $\left|\left| \frac{\gamma(6+h) - \gamma(6)}{h} \right|\right|$ were the same quantity.

## §3 Day 3: Graphs, Level sets, and Slices (Sept 8, 2025)

Office hours starting next week! Check Quercus for more details.

Last class, we looked at functions $\mathbb{R} \to \mathbb{R}^n$. Today we look at functions $\mathbb{R}^n \to \mathbb{R}$. We develop graphs, level sets, and slices because they offer new ways to analyze and study properties of such functions, that cannot be easily captured otherwise. For example, it is difficult to visualize a more than 3 dimensional vector.

Have $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}^n$.

**Definition 3.1** (Graph)**.** The graph of $f$ is $\{(x, f(x)) : x \in A\}$

Note that when you plot a function $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$, the graph would exist in $\mathbb{R}^{n+1}$.

**Definition 3.2** (Level Set)**.** The level set of $f$ at $k$ is $\{x \in \mathbb{R}^n : f(x) = k\}$[3]

To produce the **slice** of a graph, we need to hold a coordinate $x_i$ constant, which we set to $a$. We call the following a $x_i$-slice, where the $c$ is at the $i$-th position in the tuple:

$$\{(x_1, \cdots, x_{n+1}) \in \mathbb{R}^n : (x_1, \cdots, a, \cdots, x_n) \in A, x_{n+1} = f(x_1, \cdots, a, \cdots, x_n)\}$$

where in the first tuple, $x_i$ is omitted, with $x_i$ is replaced by $a$ in the second and third tuples. Note that the slice lives in $\mathbb{R}^n$, because we already have the information that $x_i = a$.

In this course, you always want to specify the domain and codomain of your function, to avoid confusion.

**Problem 3.3**

Give a function $f : \mathbb{R}^2 \to \mathbb{R}$ whose level set at $0$ is the set

$$\{(x, y) \in \mathbb{R}^2 : |x| = |y|\}$$

To solve these kinds of problems, you set $0 = |y| - |x|$ and you would get a candidate function $f(x, y) = |y| - |x|$. To show that two sets $A$ and $B$ are equal, you would typically prove that $A \subseteq B$, and $B \subseteq A$.

Alternatively you could define a function

$$g(x, y) = \begin{cases} 0 & \text{if } |x| = |y| \\ 1 & \text{otherwise} \end{cases}$$

which satisfies the requirements by construction.

[3] If $A \subseteq \mathbb{R}^2$, the level set is called a *contour*.

# §4 Day 4: Vector Fields and Transformations (Sept 9, 2025)

So far, we have seen parametric curves ($\mathbb{R} \to \mathbb{R}^n$), real valued functions ($\mathbb{R}^n \to \mathbb{R}$). Today we look at functions $\mathbb{R}^n \to \mathbb{R}^n$, which are called vector fields or transformations[4].

**Definition 4.1** (Vector Field)**.** A $n$ dimensional vector field is a function $F : A \to B$ with $A, B \subseteq \mathbb{R}^n$.

Note that vector fields are capitalized as per convention.

(Not testable) Newton's law of gravity states that the force exerted by an object at the origin with mass $m_1$ on an object at $(x, y, z)$ with mass $m_2$ is given by

$$F(x,y,z) = \frac{-Gm_1m_2}{||(x,y,z)||^2} \cdot \frac{(x,y,z)}{||(x,y,z)||}$$

The magnitude is controlled by the first part of the product (note that there are only scalars), and the direction is controlled by the unit vector (note that it is scaled down to have a magnitude of 1).

[4]very uncommon to call such functions transformations

# §5 Day 5: Coordinate Transformations (Sept 11, 2025)

**Definition 5.1** (Coordinate Transformation)**.** A **coordinate transformation** $f : A \to B$ is a continuous transformation that is usually bijective. $A$ and the map $f$ form a **coordinate system** for the codomain $B$.

We want to plot subsets of $B$ and describe them using the coordinate system defined by $f$ and $A$. We use $(u, v)$ to describe the elements in $A$, and $(x, y)$ for $B$, and would write

$$g(u, v) = (u^2 + v^2, v)$$

or simply

$$(x, y) = (u^2 + v^2, v)$$

**Definition 5.2** (Polar Coordinate Transformation)**.** A map $T : \mathbb{R}^2 \to \mathbb{R}^2$, with

$$T(r, \theta) = (r \cos\theta, r \sin\theta)$$

$T$ describes a map from polar coordinates to cartesian coordinates. The radius can be negative. Two notable properties are $T(r, \theta) = T(-r, \theta + \pi)$ and $T(r, \theta) = T(r, \theta + 2\pi)$, following from trigonometry.

The set $\{(r, \theta) \in \mathbb{R}^2 : r = 2\}$ would describe the set $\{(2\cos\theta, 2\sin\theta) \in \mathbb{R}^2 : \theta \in \mathbb{R}\}$, which corresponds to a circle of radius 2 centered at the origin. Restricting both sets to $\theta \in \mathbb{R}^+$ or $\theta \in \mathbb{R}^-$ would still correspond to the same set, meaning that the polar coordinate transformation is not injective. Another way to show is to consider the case $r = 0$.

What follows was not covered in lecture:

**Definition 5.3** (Cylindrical Coordinate Transformation)**.** A map $T : \mathbb{R}^3 \to \mathbb{R}^3$, with

$$T(r, \theta, z) = (r \cos\theta, r \sin\theta, z)$$

This is similar to the polar coordinate transformation except we add an additional $z$ field which remains unchanged.

**Definition 5.4** (Spherical Coordinate Transformation)**.** A map $T : \mathbb{R}^3 \to \mathbb{R}^3$ with

$$T(\rho, \theta, \phi) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi)$$

If this gets covered in class I'll make a writeup deriving the formula.

TODO: derive/motivate the formula

# §6 Day 6: Parametric, Implicit, and Explicit Form (Sep 15, 2025)

Have $n < m$ be positive integers.

**Definition 6.1** (Parametric Form)**.** A set $S \subseteq \mathbb{R}^m$ can be written in parametric form (with $n$-variables) if there exists a set $A \subseteq \mathbb{R}^n$, and a continuous map $g : A \to \mathbb{R}^m$ such that

$$S = \{g(x) : x \in A\}$$

**Definition 6.2** (Higher Dimensional Graphs)**.** Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function.

$$\text{graph}(f) = \{(x, f(x)) \in \mathbb{R}^n \times \mathbb{R}^m \mid x \in A\}$$

In this course $\mathbb{R}^a \times \mathbb{R}^b = \mathbb{R}^{a+b}$, and we don't care whether the nesting of ordered pairs is done in the first or second entry.

**Definition 6.3** (Explicit Form)**.** A set $S$ can be written in explicit form in $n$ variables if $S \subseteq \mathbb{R}^m$ is graph($f$) for some continuous function $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$.

**Definition 6.4** (Implicit Form)**.** A set $S$ can be written in implicit form in $n$ variables if there exists a constant $c \in \mathbb{R}^m$, and a continuous function $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ such that $S = f^{-1}(\{c\})$.

Implicit form is a generalized version of a level set. The level set is defined for functions $g : \mathbb{R}^n \to \mathbb{R}$, notice that $g^{-1}(\{c\}) = \{x \in \mathbb{R}^n : g(x) = c\}$ is the same as the level set of $g$ at $c$.

# §7 Day 7: Interior, Boundary, and Closure (Sep 16, 2025)

Let $A \subseteq \mathbb{R}^n$.

**Definition 7.1** (Interior Point)**.** $p \in \mathbb{R}^n$ is an interior point f $A$ if there exists an $\epsilon > 0$ such that $B_\epsilon(p) \subseteq A$.

**Definition 7.2** (Interior)**.** The interior denoted $A^o$ is the set of all interior points of $A$.

**Definition 7.3** (Boundary Point)**.** $p \in \mathbb{R}^n$ is a boundary point if for every $\epsilon > 0$, $B_\epsilon(p) \cap A$ and $B_\epsilon \cap A^c$ are non-empty.

The boundary point can be isolated, there may exist $\epsilon$-balls at $p$ such that $p$ is the only point in said ball.

**Definition 7.4** (Topological Boundary)**.** The topological boundary of $A$, written $\partial A$.

**Definition 7.5** (Limit Point)**.** $p \in \mathbb{R}^n$ is a limit point of $A$ if for every $\epsilon > 0$, $B_\epsilon \setminus \{p\}$ contains points in $A$. The set of all limit points is written $A^*$.

A set $B$ containing points in $A$ is the same as saying that $A \cap B \neq \emptyset$. Also to trip you up, sometimes questions will need you negate the definition, so do that if it seems unintuitive to explain.

**Definition 7.6** (Closure)**.** The closure of $A$ is written $\overline{A}$, defined as $\overline{A} = A^* \cup A$.

**Theorem 7.7.** Every interior point of $A$ is a limit point of $A$. That is $A^o \subseteq A^*$.

*Proof.* Let $p$ be an interior point of $A$. By definition, exist $\epsilon > 0$, such that $B_\epsilon(p) \subseteq A$. Let $\epsilon' > 0$ be arbitrary. If $\epsilon' > \epsilon$, note that $B_\epsilon(p) \subseteq B_{\epsilon'}(p)$e can take any point in the open ball $B_\epsilon(p) \setminus \{p\}$, which is a subset of $A$ following from the definition. If $\epsilon' \leq \epsilon$, we can take any point in $B_{\epsilon'}(p) \setminus \{p\} \subseteq B_\epsilon(p) \setminus \{p\} \subseteq A$. $\square$

Note that this proof relies on the fact for $\epsilon > 0$, there exist points in $B_\epsilon(p) \setminus \{p\}$ that are not $p$. This is not true from some topologies.

**Theorem 7.8.**

$$A^o \subseteq A \subseteq \overline{A}$$
$$A^o \cap \partial A = \emptyset$$
$$\overline{A} = A^o \cup \partial A$$
$$\partial A = \overline{A} \setminus A^o$$

You should be able to verify this by wrangling definitions.

# §8 Day 8: Sequences (Sep 18, 2025)

**Definition 8.1** (Sequence)**.** A sequence in in $\mathbb{R}^n$ is a function with domain $\{k \in \mathbb{Z} : k > k_0\}$ for some fixed $k_0 \in \mathbb{Z}$ and codomain $\mathbb{R}^n$.

**Definition 8.2** (Subsequence)**.** Let $x : \mathbb{N}^+ \to \mathbb{R}^n$ be a sequence and $m : \mathbb{N}^+ \to \mathbb{N}^+$ be a strictly increasing function. The sequences $\{x(m(k))\}_{k=1}^{\infty}$ is a subsequence of the sequences $\{x(k)\}_{k=1}^{\infty}$.

**Definition 8.3** (Convergence)**.** We fix a $p \in \mathbb{R}^n$. A sequence $\{x(k)\}_k$ in $\mathbb{R}^n$ converges to $p$ if for every $\epsilon > 0$, there exists a $K \in \mathbb{N}$ such that for every $k \in \mathbb{N}$, if $k \geq K$ then $||x(k) - p|| < \epsilon$.

**Theorem 8.4.** A sequence $\{x(k)\}$ converges to a point $p$ if and only if for every $\epsilon > 0$, the set of indices $\{k \in \mathbb{N}^+ : x(k) \notin B_\epsilon(p)\}$ is finite.

*Proof.* ( $\Longrightarrow$ ) Suppose $\{x(k)\}$ converges to $p$. Then we take an arbitrary $\epsilon > 0$. There exists a $K \in \mathbb{N}^+$ such that for $k \in \mathbb{N}^+$, $k \geq K$, $x(k)$ must belong to $B_\epsilon(p)$. Then points $x(j)$ not belonging to $B_\epsilon(p)$ must be of the form $x(j)$, where $j \in \mathbb{N}^+, j < k$. Since $j \in \{1, 2, \cdots, k-1\}$, we have finitely many such points.

( $\Longleftarrow$ ) Suppose that for any $\epsilon > 0$ there exists only finitely many indices such that their corresponding point is not in $B_\epsilon(p)$. Let $S$ denote the aforementioned set of indices. In the case where this is the empty set, we can take $K = 1$. If the set is non-empty, then take $K$ as the largest such index plus one.

In both cases, all indices $k \geq K$ have $x(k) \in B_\epsilon(p)$. We can prove this by contradiction: Suppose there exists some $j \geq K$, $x(j) \notin B_\epsilon(p)$. If $S = \emptyset$, then this is immediately a contradiction. Otherwise, then the existence of $j$ contradicts the maximality of $K$, since $K$ is defined as the largest such index plus one.

As such a $j$ does not exist, we have that for all $k \in \mathbb{N}^+$, if $k \geq K$ then $x(k) \in B_\epsilon(p)$. $\square$

We cannot say that $\{x(k)\}$ converges to $p$ is equivalent to for all $\epsilon > 0$, the set $\{x(k) : k \in \mathbb{N}^+, x(k) \notin B_\epsilon(p)\}$ is finite. Consider the sequence $k \in \mathbb{N}^+$, $x(k) = (-1)^k$. This sequence has only finitely many points in its image, being $\{1, -1\}$, but is clearly not convergent to any point $p \in \mathbb{R}$ (picking $\epsilon < \frac{1}{2}$ would work).

## §9 Day 9: Open and Closed Sets (Sep 22, 2025)

You should know that there exist sequential formulations equivalent to the definition of closed and open sets. We want sets that have nice properties: Let $A \subseteq \mathbb{R}^n$. We want a class of sets such that if a sequence in $\mathbb{R}^n$ converges to $a \in A$, then the *tail* of the sequence belongs in $A$.

**Definition 9.1** (Open Set)**.** $A$ is open if every point of $A$ is an interior point of $A$.

This is equivalent to saying $A = A^o$, and $A \cap \partial A = \emptyset$.

**Definition 9.2** (Closed Set)**.** $A$ is closed if every limit point of $A$ belongs to $A$.

That is, $A^* \subseteq A$.

**Theorem 9.3**

A set $A$ is open if and only if its complement $A^c = \mathbb{R}^n \setminus A$ is closed.

*Proof.* We prove both directions:

($\Rightarrow$) Suppose $A$ is open.
For contradiction, suppose that $A^c$ is not closed, meaning that there exists a limit point of $A^c$, being $p$, such that $p \notin A^c$. Thus $p \in A$. But $p$ is an interior point of $A$, hence there exists an $\epsilon > 0$, $B_\epsilon(p) \subseteq A$. But $B_\epsilon(p) \setminus \{p\} \subseteq A$, this contradicts our assumption that $p$ is a limit point of $A$.

($\Leftarrow$) Suppose $A^c$ is closed.
Suppose that $A$ is not open. Then there exists a point $p \in A$, such that $p$ is not an interior point. Let $k \in \mathbb{N}^+$. Then for each such $k$ there exists a point $b_k \in B_{\frac{1}{k}}(p)$, where $b_k \in A^c$. We check that this sequence $\{b_k\}_k$ converges to $p$:

> Take $\epsilon > 0$. By the archimedean property, exist some $N \in \mathbb{N}$, with $\frac{1}{N} < \epsilon$. For all $b_j$, $j \geq N$, have $b_j \in B_{\frac{1}{j}}(p)$.

But each $b_k \in A^c$. Since this sequence lies entirely in $A^c$, and $p \in A$, a limit point of $A^c$ doesn't belong to $A$. This contradicts $A^c$ being closed.

$\square$

## §10 Day 10: Compactness (Sep 23, 2025)

**Definition 10.1** (Compactness)**.** A set $A \subseteq \mathbb{R}^n$ is compact if every sequence of $A$ has a subsequence which converges to a point lying inside $A$.

Compactness here is called sequential compactness in other courses/textbooks. For this reason, our theorem relating compactness to being closed and bounded will be the Bolzano-Weierstrass theorem instead of the usual Heine-Borel.

**Definition 10.2** (Boundedness)**.** $A \subseteq \mathbb{R}^n$ is bounded if there exists a $R > 0$, $A \subseteq \{x \in \mathbb{R}^n : ||x|| < R\}$.

The complement of an unbounded set may not be bounded. The complement of a bounded set is always unbounded.

**Theorem 10.3** (Bolzano-Weierstrass)

A set is sequentially compact if and only if it is both closed and bounded.

*Proof.* We prove by showing both directions.

(⇒) Suppose that $S$ is compact.

- We show that $S$ is closed. Let $p$ be any limit point of $p$, meaning that there exists a sequence $\{x_k\}_k$ that converges to $p$.
- We show that $S$ is bounded.

(⇐) Suppose that $S$ is closed and bounded. WTS that it is compact.

□

TODO: finish this proof.

# §11 Day 11: Limits (Sep 25, 2025)

"You get a delta, you get a delta, everybody gets a delta" - Q.

**Definition 11.1** (Limit)**.** Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function. Let $a \in \mathbb{R}^n$ be a limit point of $A$, and let $b \in \mathbb{R}^m$. Define $b$ to the limit of $f$ at $a$ provided that:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A, 0 < ||x - a|| < \delta \implies ||f(x) - b|| < \epsilon$$

If the above is true, then we write $\lim_{x \to a} f(x) = b$ or say $f(x) \to b$ as $x \to a$.

**Remark 11.2**

In this course, if $x$ left unquantified, then implicitly $x$ is assumed to be an element of $\mathbb{R}^n$. Furthermore, we don't define limits at isolated points, because we want limits to be unique. Look at exercises 2.5.14, 2.5.15.

Having $a$ be a limit point of $A$ is important: without this condition, you could say that all points $b \in \mathbb{R}^m$ is the limit of $f$ at any limit point $k \in \mathbb{R}^n$ that is not a limit point of $A$, since there would exist some $\delta$-radius such that no points of $A$ exist in the $B_\delta(k)$, thus the

**Definition 11.3** (Isolated Point)**.** Have $A \subseteq \mathbb{R}^n$. A point $a \in \mathbb{R}^n$ is an isolated point of $A$ if $a \in A$ and $a$ is not a limit point of $A$.

This is equivalent to saying that $a$ is an isolated point iff $a \in A \cap (\mathbb{R}^n \setminus A^*)$

**Theorem 11.4** (Sequential Definition of Limits)

Let $A \subseteq \mathbb{R}^n$ be a set and let $f : A \to \mathbb{R}^m$ be a function. Let $a \in \mathbb{R}^n$ be a limit point of $A$ and let $b \in \mathbb{R}^m$. Then $\lim_{x \to a} f(x) = b$ if and only if every sequence of points $\{x(k)\}_k$ in $A \setminus \{a\}$ with $x(k) \to a$, the sequence of points $\{f(x(k))\}_k$ in $\mathbb{R}^m$ converges to $b$, that is $f(x(k)) \to b$.

Proof was left as an exercise.

**Theorem 11.5** (Component-wise Limits)

Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function. Let $a$ be a limit point of $A$ and $b = (b_1, \cdots, b_m) \in \mathbb{R}^m$. Let $f_i$ be the component functions of $f$, such that $f = (f_1, \cdots, f_m)$, then $\lim_{x \to a} f(x) = b$ if and only if for all $i \in \{1, \cdots, m\}$, $\lim_{x \to a} f_i(x) = b_i$.

*Proof.* We prove both directions:
( $\implies$ ): Let $\epsilon > 0$. By definition of $\lim_{x \to a} f(x) = b$, exist $\delta > 0$ (we use this same $\delta$) , for $x \in A$, if $||x - a|| < \delta$ then $||f(x) - b|| < \epsilon$. From the triangle inequality, have $|f_i(x) - b_i| \leq ||f(x) - b|| < \epsilon$, thus $|f_i(x) - b_i| < \epsilon$, which proves the claim.
( $\impliedby$ ): Let $\epsilon > 0$ be arbitrary. For $i \in \{1, \cdots, m\}$, have some $\delta_i > 0$, if $||x - a|| < \delta_i$ then

$|f_i(x) - b_i| < \frac{\epsilon}{\sqrt{m}}$. Take $\delta = \min\{\delta_1, \cdots, \delta_m\}$. Then we compute

$$||f(x) - b||^2 \leq \sum_{i=1}^{m} |f_i(x) - b_i|^2 < \sum_{i=1}^{m} \frac{\epsilon^2}{m} = \epsilon^2$$

Taking square roots, we get the desired result. □

# §12 Day 12: Continuity (Sep 29, 2025)

**Definition 12.1** (Continuity)**.** Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function, let $a \in A$. $f$ is continuous at $a$ if

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A, ||x - a|| < \delta \implies ||f(x) - f(a)|| < \epsilon$$

With this definition, $f$ is always continuous at isolated points in $A$, since there is a $\delta$ radius around said point that do not contain any other points in $A$. Yet we cannot write $\lim_{x \to a} f(x)$ here, since $a$ would not be a limit point (from last class, we only define limits at limit points).

If $a$ is indeed a limit point, then $\lim_{x \to a} f(x) = f(a)$ is equivalent to our definition of continuity, because in that case $\lim_{x \to a} f(x)$ is actually a defined symbol.

## 12.1 Studying Linear Transformations

**Theorem 12.2.** If $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, there exists a $M \in \mathbb{R}$, such that for all $x \in \mathbb{R}^n$, with $||x|| = 1$, have $||T(x)|| \leq M$.

*Proof.* Note that $x$ must have each component lesser than or equal to 1, otherwise the contradicts the norm of $x$ being 1.

$$\begin{aligned} ||T(x)|| &= ||T(x_1) + \cdots + T(x_n)|| \\ &\leq ||T(e_1) + \cdots + T(e_n)|| \\ &\leq \sum_{i=1}^{n} ||T(e_i)|| = M \end{aligned}$$

□

**Theorem 12.3.** If $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then there exists some $M \geq 0$, such that for all $x \in \mathbb{R}^n$, $||T(x)|| \leq M||x||$

*Proof.* Suppose $x \in \mathbb{R}^n$, $||x||$ may not be 1,

$$\begin{aligned} ||T(\frac{x}{||x||})|| &\leq M \\ \frac{||T(x)||}{||x||} &\leq M \\ ||T(x)|| &\leq M||x|| \end{aligned}$$

with the first line following by 12.2. Since the norm function is non-negative, $0 \leq M$. □

**Theorem 12.4.** Linear transformations are continuous.

*Proof.* We will show this from the definition of continuity. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation.

Let $\epsilon > 0$. Take $\delta = \frac{\epsilon}{M}$. Let $x \in A$ be arbitrary, and suppose $||x - a|| < \delta$. Want to show that $||L(x) - L(a)|| < \epsilon$. We proceed as follows:

$$\begin{aligned} ||L(x) - L(a)|| &= ||L(x - a)|| \\ &\leq M||x - a|| \\ &< M\delta = M\frac{1}{M} = \epsilon \end{aligned}$$

Note that such an $M$ exists by 12.3. □

## 12.2 Continuity and Topology

**Theorem 12.5** (Topological Definition of Continuity)

$f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if for any open $U \subseteq \mathbb{R}^m$, $f^{-1}(U)$ is open.

*Proof.* We will prove both directions.
( $\Longrightarrow$ ): Want to show $f^{-1}(U)$ be open. Suppose $f$ is continuous, and that $U$ is open. Let $x \in f^{-1}(U)$ be arbitrary. Then $f(y) \in U$ is an interior point of $U$ (every point in an open set is an interior point). Thus exists $\epsilon > 0$, $B_\epsilon(y) \subseteq U$. By continuity of $f$ at $x$, exists $\delta > 0$, $f(B_\delta(x)) \subseteq B_\epsilon(y)$, meaning that $B_\delta(x) \subseteq f^{-1}(B_\epsilon(y)) \subseteq f^{-1}(U)$, which is what we wanted to show.
( $\Longleftarrow$ ): Want to show that $f$ is continuous. Suppose for any open $U \subseteq \mathbb{R}^m$, $f^{-1}(U)$ is open. Then □

**Theorem 12.6** (Component Wise Continuity)**.** The map $f = (f_1, \cdots, f_m) : A \to \mathbb{R}^m$ is continuous at $a \in A$ if and only if for each $i \in \{1, \cdots, m\}$, $f_i$ is continuous at $a$.

The textbook says that the above follows by breaking into cases whether $a$ is an isolated or limit point of $A$. In the former case, continuity at $a$ 'vacuously' happens, and in the latter case we apply 11.5, breaking the limit into its components (or joining it back together using the other direction).

*Proof.* Consider $\pi_i \circ f$, $\pi_i$ is a linear map, and the composition of continuous functions is continuous, each $f_i$ is continuous. For the other direction, we verify using the definition. □

## §13 Day 13: Path Connected Sets (Oct 06, 2025)

**Definition 13.1** (Path-Connectedness)**.** A set $S \subseteq \mathbb{R}^n$ is path-connected if for all points $p, q \in S$, there exists a continuous function $\gamma : [a, b] \to S$ such that $\gamma(a) = p$, $\gamma(b) = q$, with $\mathrm{im}(\gamma) \subseteq S$.

**Definition 13.2** (Convex)**.** A set $S$ is convex if the line segment between any 2 points in $S$ are contained in $S$.

See that $f(S)$ for some continuous $f$ is also path connected, since we can use $f \circ \gamma$.

**Theorem 13.3** (IVT)**.** Let $f : [a, b] \to \mathbb{R}$ be a function. If $f$ is continuous on $[a, b]$, then $f([a, b])$ is path connected.

$[a, b]$ is path connected, and continuous images of path connected sets are path connected.

# §14 Day 14: Global Extrema (Oct 06, 2025)

**Definition 14.1.** Have $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$. Let $p \in A$.

- $p$ is a global maximum point of $f$ on $A$ if for all $x \in A$, $f(p) \geq f(x)$
- $p$ is a global minimum point of $f$ on $A$ if for all $x \in A$, $f(p) \leq f(x)$

We call $f(p)$ the global maximum/minimum value of $f$ on $A$, and say that $f$ attains a global maximum/minimum on $A$ respectively.

**Theorem 14.2** (Single Variable EVT)**.** Let $f : [a, b] \to \mathbb{R}$ be continuous. Then $f$ attains a maximum and a minimum on $[a, b]$.

**Theorem 14.3** (EVT)**.** Have $A \subseteq \mathbb{R}^n$ be non-empty and compact, and $f : A \to \mathbb{R}$ be continuous. $f$ attains a maximum and minimum on $A$.

# §15 Day 15: Derivatives of One Variable (Oct 06, 2025)

**Remark 15.1**

We did not learn differentiability until 20, but there are theorems that use this concept prior to its introduction. Also: "Professors say: 'Never differentiate in public. You will only embarrass yourself.'" -Q

Let $A \subseteq \mathbb{R}$ and let $f : A \to \mathbb{R}^m$ be a function. Let $a$ be an interior point of $A$.

**Definition 15.2** (Differential)**.** If $f$ is differentiable at $a$, define the linear map $df_a : \mathbb{R} \to \mathbb{R}^m$ given by $df_a(h) = f'(a)h$ called the differential of $f$ at $a$.

Vectors $v \in \mathbb{R}^m$ are identified with $m \times 1$ column vectors

$$\begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}$$

instead of $1 \times m$ row vectors, despite the fact that we write $v = (v_1, \cdots, v_m)$. This is because we can write for $v, w \in \mathbb{R}^m$, $v \cdot w = v^T w = w^T v$. We still distinguish between vectors and matrices.

From first year calculus, the linear approximation of $f : \mathbb{R} \to \mathbb{R}$ at a point $a$ is given by

$$g(x) = f(a) + f'(a)(x - a)$$

Substituting $x - a$ with $h$, we get that

$$f(a + h) \approx f(a) + \underbrace{hf'(a)}_{df_a(h)}$$

Rearranging, we get that $f(a + h) - f(a) - df_a(h) \approx 0$. In fact, we can strengthen this condition, such that

$$\lim_{h \to 0} \frac{f(a + h) - f(a) - L(h)}{h} = 0$$

where $L : \mathbb{R} \to \mathbb{R}^m$ is a linear transformation, such that $L(h) = f'(a)h = df_a(h)$.

# §16 Day 16: Partial Derivatives (Oct 07, 2025)

We now want to generalize 1 dimensional derivatives to $n$-dimensional derivatives. Unfortunately, we find that this is not the right generalization, for reasons you will see later.

**Definition 16.1** (Partial Derivative)**.** The $i$-th partial derivative of $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ at $a \in A^o$ is given by

$$\partial_j f(a) := \lim_{h \to 0} \frac{f(a + he_j) - f(a)}{h}$$

Equivalent notation for $\partial_j f$ are

$$\frac{\partial f}{\partial x_j}, \quad D_{e_j} f, \quad f_{x_j}, \quad D_j f, \quad \partial_{x_j} f, \quad \partial_j f$$

$\partial_j f(a)$ is relatively straightforward to compute (do it component-wise).

# §17 Day 17: Directional Derivatives (Oct 09, 2025)

**Definition 17.1** (Directional Derivative)**.** Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$, and $a \in A^o$, and fix $v \in \mathbb{R}^n$. The directional derivative of $f$ at $a$ in the direction $v$ is given by

$$D_v f(a) = \lim_{h \to 0} \frac{f(a + hv) - f(a)}{h}$$

As usual, $D_v f(a)$ exists if and only if $D_v f_i(a)$ (its components) exists for $i \in \{1, \cdots, m\}$.

More properties about the directional derivative comes later.

# §18 Day 18: Gradients (Oct 14, 2025)

Let $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}$, with $a \in A^o$.

**Definition 18.1** (Gradient)**.** The gradient of $f$ at $a$, denoted $\nabla f(a)$, is given by

$$\nabla f(a) = (\partial_1 f(a), \cdots, \partial_n f(a))$$

if each $\partial_i f(a)$ is defined for every $i \in \{1, \cdots, n\}$

The gradient is quite useful, for:

- Finding extrema (specifically 23.2)
- Finding the 'direction of steepest ascent'

The gradient greatly simplifies directional derivative computations.

**Theorem 18.2.** For $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$, if $f$ is differentiable at $a \in A$ then $D_v f(a) = \nabla f(a) \cdot v$.

*Proof.* Apply 20.3, giving us that $D_v f(a) = Df_a(v)$. Decompose $v = \sum_{i=1}^n v_i e_i$. Then

$$\begin{aligned} Df_a(v) &= Df_a(\sum_{i=1}^n v_i e_i) \\ &= \sum_{i=1}^n v_i Df_a(e_i) \\ &= \sum_{i=1}^n v_i \partial_i f(a) \end{aligned}$$

which is the exact same computation as $\nabla f(a) \cdot v$. □

We formalize the second point as follows:

**Theorem 18.3.** Suppose $f$ is differentiable at $a \in A^o$, $\nabla f(a) \neq 0$, then

$$\max\{D_u f(a) : u \in \mathbb{R}^n, ||u|| = 1\} = D_v f(a)$$

where $v = \frac{\nabla f(a)}{||\nabla f(a)||}$.

*Proof.* TODO: □

# §19 Day 19: Differentials and Jacobians (Oct 16, 2025)

Recall that in single var calculus, differentiability at a point implies continuity at a point. Our past few experiments with differentiability, not provide sufficient conditions for continuity at some point.

**Definition 19.1** (Differentiable)**.** $f$ is differentiable at $a$, if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$, such that

$$\lim_{||h|| \to 0} \frac{f(a+h) - f(a) - L(h)}{||h||} = 0$$

$L$ is called *the* differential of $f$ at $a$, written $df_a$.

The reason we say *the* differential, is because of the fact that differentials of a function at the same point are unique.

**Definition 19.2** (Jacobian)**.** The Jacobian of $f$ at $a$ is given by

$$Df(a) = [\partial_j f_i(a)]_{i,j} = \begin{bmatrix} | & & | \\ \partial_1 f(a) & \cdots & \partial_n f(a) \\ | & & | \end{bmatrix}$$

We have defined differentiability, the differential, as well as the jacobian. The notation $Df(a)$, $df_a$ suggests that they are closely related. We formalize their relationship as follows:

**Theorem 19.3.** If $f$ is differentiable at $A$, then all partial derivatives exist, and the matrix for $df_a$ is the jacobian $Df_a$, with $\partial_i f(a) = df_a(e_i)$.

*Proof.* Recall that a linear transformation is uniquely determined by its action on the basis vectors. As per the definition of differentiability,

$$0 = \lim_{h \to 0} \frac{f(a + he_i) - f(a) - hL(e_i)}{h} = \left[\lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h}\right] - L(e_i) = \partial_i f(a) - L(e_i)$$

so we can conclude that the partial derivative $\partial_i$ exists by limit law. $\square$

If we weren't given $df_a$, but somehow know that $f$ is differentiable at $a$ (hence $df_a$ exists) and given each $\partial_i f(a)$ instead, since we know the actions of $df_a$ on the basis vectors, we can retrieve the matrix of $df_a$.

# §20 Day 20: Differentiability (Oct 20, 2025)

**Definition 20.1** (Differentiable)**.** Let $f : a \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function, and $a \in A^o$ (interior point). $f$ is differentiable at $a$ if there exists a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$, such that

$$\lim_{||h|| \to 0} \frac{f(a+h) - f(a) - L(h)}{||h||} = 0$$

Note that the right hand side is the zero vector, not zero (the scalar). We omit this distinction for simplicity.

**Theorem 20.2.** Have $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$, $a \in A^o$. Then $f$ being differentiable at $a$, is equivalent to having a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ with

$$\lim_{||h|| \to 0} \frac{||f(a+h) - f(a) - L(h)||}{||h||} = 0$$

*Proof.* Consider the function $g : A \setminus \{\vec{0}\} \to \mathbb{R}^m$,

$$g(h) = \frac{f(a+h) - f(a) - L(h)}{||h||}$$

We see that

$$\lim_{||h|| \to 0} g(h) = 0 \iff \lim_{||h|| \to 0} ||g(h)|| = 0$$

□

**Theorem 20.3.** Have $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$, with $a \in A^o$, and $f$ differentiable at $a$. Then $D_v f(a) = Df_a(v)$.

*Proof.* Since $f$ is differentiable, substitute $h$ ($h$ vector) with $hv$ ($h$ is scalar, $v$ is given vector), giving

$$\begin{aligned} 0 = \lim_{h \to 0} \frac{f(a+hv) - f(a) - hdf_a(v)}{h} &= \lim_{h \to 0} \frac{f(a+hv) - f(a)}{h} - df_a(v) \\ &= D_v f(a) - df_a(v) \end{aligned}$$

Apply 19.3, since the matrix for $df_a$ is $Df_a$, we get $D_v f(a) = Df_a(v)$, as desired. □

**Definition 20.4** ($C^1$)**.** $f$ is continuously differentiable at $a$ if $\partial_1, \cdots, \partial_n f$ are defined on an open set containing $a$, with all being continuous at $a$.

**Theorem 20.5**

If $f$ is continuously differentiable ($C^1$) at $a$, then $f$ is differentiable at $a$.

**Theorem 20.6.** The composition of $C^1$ functions is $C^1$.

Proven using the chain rule.

# §21 Day 21: Chain Rule, MVT (Oct 21, 2025)

## 21.1 Chain Rule

The chain rule is a big theorem, that we can use to develop even more theory, such as the MVT, which has an unexpected generalization.

**Theorem 21.1** (Chain Rule)

Have $U \in \mathbb{R}^n$ be open, $V \in \mathbb{R}^m$ be open. Suppose $f : U \to V$ and $g : V \to \mathbb{R}^k$ are differentiable at $a \in U$ and $f(a) \in V$ respectively. Then $h = g \circ f$ is differentiable at $a$, with

$$dh_a = dg_{f(a)} \circ df_a$$

Recall the definition of differentiability. We use the equivalent formulation provided by 20.2, with some modifications:

$$\lim_{x \to a} \frac{||f(x) - f(a) - L(x-a)||}{||x-a||} = 0$$

We get this by taking $x - a = h$. This form is more useful in the proof of the chain rule

*Proof (Adapted from Spivak).* Let $b = f(a)$, $\lambda = df_a, \mu = dg_{f(a)}$. Define

1. For $x \in U$, define $\varphi(x) = f(x) - f(a) - \lambda(x-a)$
2. For $y \in V$, define $\psi(y) = g(y) - g(b) - \mu(y-b)$
3. For $x \in U$, define $\rho(x) = g \circ f(x) - g \circ f(a) - \mu \circ \lambda(x-a)$.

Then our givens are

$$\lim_{x \to a} \frac{||\varphi(x)||}{||x-a||} = 0, \quad \lim_{y \to b} \frac{||\psi(y)||}{||y-b||} = 0 \tag{$*$}$$

and we want to show that $\lim_{x \to a} \dfrac{||\rho(x)||}{||x-a||} = 0$.

$$\begin{aligned}
\rho(x) &= g(f(x)) - g(f(a)) - \mu(\lambda(x-a)) \\
&= g(f(x)) - g(f(a)) - \mu(f(x) - f(a) - \varphi(x)) && \text{def. of } \varphi \\
&= g(f(x)) - g(f(a)) - \mu(f(x) - f(a)) + \mu(\varphi(x)) && \mu \text{ is linear} \\
&= g(y) - g(b) - \mu(y-b) + \mu(\varphi(x)) && \text{def. of } y, b \\
&= \psi(f(x)) + \mu(\varphi(x))
\end{aligned}$$

So it suffices to show that each part of the sum has limit 0 as $x \to a$, over $||x - a||$. We do so first for $\mu(\varphi(x))$:

> By 12.3, since $\mu$ is linear, have some $M$ such that for all $v \in \mathbb{R}^m$, $||\mu(\varphi(x))|| \leq M||\varphi(x)||$
>
> $$\lim_{x \to a} \frac{||\mu(\varphi(x))||}{||x-a||} \leq \lim_{x \to a} \frac{M||\varphi(x)||}{||x-a||} = 0$$

with the last inequality following from ($*$).

Now we do the same for $\psi(f(x))$.

Let $\epsilon > 0$. Since $\lambda$ is linear, let $\epsilon' < \frac{\epsilon}{M}$, where $M$ is the scalar given by 12.3. From ($*$), we have some $\delta > 0$ such that if $||x - a|| < \delta$, then[5]

$$\begin{aligned} ||\psi(f(x))|| &< \epsilon'||f(x) - b|| \\ &= \epsilon'||\varphi(x) + \lambda(x - a)|| \\ &\leq \epsilon'||\varphi(x)|| + \epsilon' M||x - a|| \end{aligned}$$

Considering the limit expression again, have

$$\lim_{x\to a} \frac{||\psi(f(x))||}{||x - a||} \leq \lim_{x\to a} \frac{\epsilon'||\varphi(x)|| + \epsilon' M||x - a||}{||x - a||} = \epsilon' M < \epsilon$$

Finally, we have that

$$\lim_{x\to a} \frac{||\rho(x)||}{||x - a||} = \lim_{x\to a} \frac{||\psi(f(x)) + \mu(\varphi(x))||}{||x - a||} \leq \lim_{x\to a} \frac{||\psi(f(x))|| + ||\mu(\varphi(x))||}{||x - a||} = 0$$

$\square$

Have $f : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^m \to \mathbb{R}^p$. With the chain rule, we can now compute $\frac{\partial(g\circ f)}{\partial x_i}\Big|_{x=a}$ for some $a \in \mathbb{R}^n$ . We know that $d(g \circ f)_a = dg_{f(a)} \cdot df_a$, whose matrix is given by $Dg_{f(a)}Df_a$. Then via 19.3, have that $\frac{\partial(g\circ f)}{\partial x_i}\Big|_{x=a} = Dg_{f(a)}Df_a(e_i)$, which is the $i$-th column of that matrix. More explicitly, this is given by

$$Dg_{f(a)}\left(\sum_{j=1}^{m} \partial_i f^j(a) e_j\right) = \sum_{j=1}^{m} \partial_i f^j(a) \left(\sum_{k=1}^{p} \partial_j g^k(f(a)) e_k\right)$$

## 21.2 Leibniz Notation

**Example 21.2**

Have $x = u(s,t), y = v(s,t), z = w(x,y)$, where $s,t \in \mathbb{R}$ are the independent variables, and $u, v, w : \mathbb{R}^2 \to \mathbb{R}$ all being differentiable. What is $\frac{\partial z}{\partial s}$, at some $a \in \mathbb{R}^2$?

We can turn this into a much simpler form, by defining $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f(s,t) = (x,y) = (u(s,t), v(s,t))$, taking the function $g(x,y) = w(x,y) = z$. Then $z = w(u(s,t), v(s,t)) = g \circ f(s,t)$. We define $h = g \circ f$, so that we can express the final function $z$ in terms of $s$ and $t$. So the question is really asking about $\partial_1 h(a)$. Applying the chain rule to $h$, have

$$dh_a = dg_{f(a)} \circ df_a$$

[5] TODO: we implicitly handle the $y = b$ case, since $\psi(b) = 0$. Usually you define a new auxiliary function

To find $\partial_i h$, where $i \in \{1, 2\}$, it suffices to compute the $i$-th column of the Jacobian $Dh_a$ (by 19.3) . Also since $f, g$ are differentiable, we can represent the linear transformation using the Jacobian.

$$\begin{aligned}\partial_i h(a) = dh_a(e_i) &= dg_{f(a)} \circ df_a(e_i) \\ &= dg_{f(a)}(\partial_i f(a)) \\ &= \begin{pmatrix}\partial_1 g_{f(a)} & \partial_2 g_{f(a)}\end{pmatrix} \begin{pmatrix}\partial_i f^1(a) \\ \partial_i f^2(a)\end{pmatrix} \\ &= \partial_1 g_{f(a)} \partial_i f^1(a) + \partial_2 g_{f(a)} \partial_i f^2(a) \end{aligned} \tag{$*$}$$

Notice that here, $\partial_1 g_{f(a)}$ and $\partial_2 g_{f(a)}$ can be confusing for the reader, since $\partial_1$ could be w.r.t. $s, t$ (being $f$'s inputs), or $x, y$ (being $g$'s inputs). To properly distinguish between the two, we can use Leibniz notation, such that $\partial_1 g_{f(a)}$ is written $\frac{\partial g}{\partial x}\big|_{x=u(s,t),y=v(s,t)}$ and $\partial_2 g_{f(a)}$ is written $\frac{\partial g}{\partial y}\big|_{x=u(s,t),y=v(s,t)}$. Leibniz notation solve this ambiguity immediately by explicitly naming the 'independent variable'. But we took $g(x, y) = w(x, y) = z$, so we may replace $g$ with $z$ here. With that in mind, let's rewrite $(*)$ entirely in Leibniz notation. We first fix $i = 1$, so we're finding the partial derivative of $z$ with respect to $s$.

$$\left.\frac{\partial z}{\partial s}\right|_{(s,t)=a} = \left.\frac{\partial z}{\partial x}\right|_{x=u(s,t),y=v(s,t)} \left.\frac{\partial x}{\partial s}\right|_{(s,t)=a} + \left.\frac{\partial z}{\partial y}\right|_{x=u(s,t),y=v(s,t)} \frac{\partial y}{\partial s}\Big|_{(s,t)=a}$$

If we weren't evaluating at the point, we can drop the evaluation terms, and get the much simpler expression:

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$$

## 21.3 Mean Value Theorem

**Theorem 21.3** (MVT)**.** Have $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, with $U$ containing some line segment $L$ from $a$ to $b$, with $f$ being differentiable on $U$. Then there exists a $c \in L$, with $f(b) - f(a) = \nabla f(c) \cdot (b - a)$.

*Proof.* We first parameterize the line segment using $g(t) = (1 - t)a + tb$, for $t \in [0, 1]$. $g([0, 1]) = L$, and from our assumptions $L \subseteq U$. Take $h : [0, 1] \to \mathbb{R}$, $h(t) = f(g(t))$. Notice that $h$ is a single variable function, with $h(0) = f(a), h(1) = f(b)$. Since $h$ is differentiable on $[0, 1]$, there exists some $s \in (0, 1)$ with $h'(s) = \nabla h(s) = \frac{h(1)-h(0)}{1-0}$, and

$$\begin{aligned} f(b) - f(a) = h(1) - h(0) &= h'(s) && \text{single var MVT} \\ &= \nabla f(g(s)) \cdot g'(s) && \text{chain rule} \\ &= \nabla f(g(s)) \cdot (b - a) && g'(s) = b - a \\ &= \nabla f(c) \cdot (b - a) && \text{Take } c = g(s) \end{aligned}$$

since $s \in (0, 1)$, $c \in L$, and we are done. ☐

Recall that for functions $f : \mathbb{R} \to \mathbb{R}^n$, differentiable at $a$, have $\nabla f(a)^\top = Df(a)$

## §22 Day 22: Local Extrema (Oct 23, 2025)

**Definition 22.1** (Local Max)**.** Have $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$, with $a \in A$. $f$ has a local maximum at $a$ if there exists a $\delta > 0$, if $x \in A \cap B_\delta(a)$ then $f(x) \leq f(a)$.

The definition for local min is similar, except with $f(x) \geq f(a)$. A local extremum is a local max or min. A point $a$ can be both a local max and local min, such as when the function is constant on $B_\delta(a)$.

**Theorem 22.2** (Single Var Local EVT)**.** Let $I \subseteq \mathbb{R}$ be open, with $c \in I$. If $f$ has a local extremum at $c$ then $f'(c) = 0$, or $f$ is not differentiable at $c$.

*Proof.* Suppose $f$ has a local maximum at $c$. Then there exists a $\delta$, such that for all $x \in I \cap B_\delta(c)$, $f(x) \leq f(c)$. Take $a, b \in B_\delta(c)$, $a < c < b$. Then $\frac{f(a)-f(c)}{a-c}$ is some non-negative value. and $\frac{f(b)-f(c)}{b-c}$ is some non-positive value. Suppose $f$ is differentiable at $c$. Then taking the limit to $c$ from the left hand side, get

$$f'(c) = \lim_{x \to c^-} \frac{f(x) - f(c)}{x - c} \geq 0$$

and $f'(c) \leq 0$ for the right hand limit as well. Hence $f'(c) = 0$. □

**Theorem 22.3** (Local EVT)**.** Let $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}$ be a real-valued function. If $a$ is an interior point of $A$, and $f$ has a local maximum at $a$, then $\nabla f(a) = 0$ or $\nabla f(a)$ does not exist.

*Proof.* Suppose the preconditions are true. It suffices to show that if $\nabla f(a)$ exists, then $\nabla f(a) = 0$. We have

$$\nabla f(a) = \begin{pmatrix} \partial_1 f(a) \\ \cdots \\ \partial_n f(a) \end{pmatrix}$$

Let $\delta > 0$, such that for $x \in A \cap B_\delta(a)$, $f(x) \leq f(a)$. For $0 < h < \delta$, have

$$\frac{f(a + he_i) - f(a)}{h} \leq \frac{f(a) - f(a)}{h} = 0$$

similarly for $-\delta < h < 0$, have

$$\frac{f(a + he_i) - f(a)}{h} \geq \frac{f(a) - f(a)}{h} = 0$$

which proves the claim. □

# §23 Day 23: Optimization (Nov 03, 2025)

'Optimization is the best thing since sliced bread.' -Q

The generic optimization problem is of the following form:

**Problem 23.1**

Have $f$ be a real valued function on a set $S \subseteq \mathbb{R}^n$.
Find the global extrema of $f$ on $S$.

Knowing the properties of $f$ and $S$ is useful. $f$ being differentiable/$C^1$, $S$ being bounded/compact, lets you apply our many theorems to help you know if extrema exist, and how to find them.

**Theorem 23.2** (Optimization Lemma)**.** Let $S \subseteq \mathbb{R}^n$, and $f : S \to \mathbb{R}$. If $S$ is compact, $f$ is continuous on $S$, and $f$ is differentiable on $S^o$, then the global extrema of $f$ occur on the boundary $\partial S$ or at a critical point $p \in S^o$, where $\nabla f(p) = 0$.

*Proof.* Since $S$ is compact, $S = S^o \cup \partial S$, with $S^o \cap \partial S = \emptyset$. By 14.3, we know there is a global extrema at some point $a \in S$. We apply 22.3. Since $\nabla f$ exists at all points of $S$, $\nabla f(s) = 0$. Otherwise, $s$ must be on the boundary. □

## §24 Day 24: Tangent Spaces (Nov 04, 2025)

**Definition 24.1** (Tangent Vector)**.** Have $S \subseteq \mathbb{R}^n$, $p \in S$. For $v \in \mathbb{R}^n$, $v$ is a tangent vector of $S$ at $p$ if there exists an open interval $I \subseteq \mathbb{R}$, $0 \in I$, some $\gamma : I \to \mathbb{R}^n$, with $\gamma$ being $C^1$, $\gamma(I) \subseteq S$, $\gamma(0) = p$, and $\gamma'(0) = v$.

**Definition 24.2** (Tangent Space)**.** The tangent space of $S$ at $p$, $T_P S$, is the set of all tangent vectors of $S$ at $p$.

The tangent space always contains the zero vector, choose a $\gamma$ that always maps to $p$. The tangent plane may not.

**Definition 24.3** (Tangent Plane)**.** Written $p + T_P S = \{p + v : v \in T_P S\}$

TODO: add the lemma!

**Theorem 24.4**

Let $V \subseteq \mathbb{R}^k$ be open, let $F : V \to \mathbb{R}^{n-k}$ be $C^1$. Let $S \subseteq \mathbb{R}^n$ be $F$'s graph $S = \{(x, F(x) : x \in V\}$. For $a \in V, p = (a, F(a))$, have

- $T_p S = \{(w, df_a(w)) : w \in \mathbb{R}^k\}$
- $T_p S = \mathrm{span}\{(e_1, \partial_1 F(a)), \cdots, (e_k, \partial_k F(a))\}$ where $\{e_1, \cdots, e_k\}$ is the standard basis for $\mathbb{R}^k$
- $T_p S$ is a $k$ dimensional subspace of $\mathbb{R}^n$

# §25 Day 25: Smooth Manifolds (Nov 06, 2025)

**Definition 25.1.** Have $k, n \in \mathbb{N}^+$, $k < n$, with $S \subseteq \mathbb{R}^n$ and $p \in S$. $S$ is a $k$-dimensional smooth manifold at $p$, if there exists an open set $U \subseteq \mathbb{R}^n$, $p \in U$ such that $S \cap U$ is a graph of a $C^1$ function $f : V \to \mathbb{R}^{n-k}$, where $V$ is open.

A set $S \subseteq \mathbb{R}^n$ is a $k$-dimensional smooth manifold if $S$ is a $k$-dimensional smooth manifold at every point in $S$.

**Theorem 25.2.** Have $S, S' \subseteq \mathbb{R}^n$, with $p \in S \cap S'$. If there exists an open set $U \subseteq \mathbb{R}^n$ containing $p$ with $S \cap U = S' \cap U$, then $T_p S = TpS'$.

For $v \in T_p S$, notice that there exists some smaller interval $I' \subseteq I$ with $p \in I'$ and $I' \subseteq S'$, since $p \in S \cap S'$ and both sets are open.

**Theorem 25.3**

Let $k, n \in \mathbb{N}^+$ with $k < n$. Have $S \subseteq \mathbb{R}^n$, $p \in S$. If $S$ is a $k$-dimensional manifold at $p$, then $T_p S$ is a $k$-dimensional subspace of $\mathbb{R}^n$.

## §26 Day 26: Diffeomorphism (Nov 09, 2025)

There are 2 big upcoming theorems, being the inverse function theorem and the implicit function theorem. The main motivation of these theorems is that we want to show that more sets are manifolds at some point. But sets can be in a rather complicated form (e.g. implicit form).

**Definition 26.1** (Global Inverse)**.** Let $U \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^n$. The global inverse of $F : U \to V$ is a map denoted $F^{-1} : V \to U$ with $F^{-1} \circ F(u) = u$ for all $u \in U$, and $F \circ F^{-1}(v) = v$ for all $v \in V$.

**Definition 26.2** (Global Diffeomorphism)**.** $F$ is a global diffeomorphism if $F$ is bijective, and both $F$ and its global inverse $F^{-1}$ is $C^1$.

**Theorem 26.3** (Symmetry)**.** Have $F : U \to V$ be bijective. Then $F$ is a diffeomorphism iff $F^{-1}$ is a diffeomorphism.

*Proof.* Suppose $F$ is a diffeomorphism. Then $F^{-1}$ exists, is $C^1$ and is also bijective. We only need to check that $(F^{-1})^{-1}$, the inverse of the inverse of $F$ is $C^1$ . $F^{-1} \circ (F^{-1})^{-1} = I$. Apply $F$ to both sides, giving $(F \circ F^{^{-1}}) \circ (F^{-1})^{-1} = F$, since function composition is associtive, so $(F^{-1})^{-1} = F$. As $F = (F^{-1})^{-1}$ is $C^1$, the global inverse of $F^{-1}$ is $C^1$, thus $F^{-1}$ is a diffeomorphism. The other direction follows from re-labeling (apply the forward direction to $G = F^{-1}$). □

**Theorem 26.4** (Transitivity)**.** Have $U, V, W \subseteq \mathbb{R}^n$ be open. $F : U \to V$ and $G : V \to W$ be diffeomorphisms. Then $G \circ F$ is a diffeomorphism $G \circ F : U \to W$.

**Theorem 26.5**

Have $S \subseteq U$, and $F$ be a diffeomorphism.

- $S$ is open iff $F(S)$ is open
- $S$ is closed iff $F(S)$ is closed
- $S$ is compact iff $F(S)$ is compact
- $S$ is path connected iff $F(S)$ is path connected

**Theorem 26.6**

Let $U, V$ be open subsets of $\mathbb{R}^n$, and $F : U \to V$ be a diffeomorphism. Let $S \subseteq U$, and $p \in S$. Some vector $v \in \mathbb{R}^n$ is a tangent vector of $S$ at $p$ if and only if $dF_p(v) \in \mathbb{R}^n$ is a tangent vector of $F(S)$ at $F(p)$.

We say that diffeomorphisms preserve topological properties and tangency.

**Definition 26.7** (Local Diffeomorphism)**.** Have $A, B \subseteq \mathbb{R}^n$ be open, with $a \in A$. $F : A \to B$ is a local diffeomorphism at $a$ if there exists some $U \subseteq A$, such that $F\big|_U : U \to F(U)$ is a diffeomorphism. $F\big|_G^{-1} : F(U) \to U$ is the local inverse of $F$ at $a$.

# §27 Day 27: Inverse Function Theorem (Nov 11, 2025)

Diffeomorphisms are particularly nice, and there are lots of powerful theorems about diffeomorphisms. This motivates us to find ways to show that a function is a diffeomorphism.

**Theorem 27.1.** Let $U$ and $V$ be open subsets of $\mathbb{R}^n$, and $F : U \to V$ is a diffeomorphism. For every $x \in U$, $DF(x)$ is an invertible $n \times n$ matrix and the jacobian of the inverse $G : V \to U$ satisfies $DG(F(x)) = [DF(x)]^{-1}$

*Proof.* Have that for all $x \in U$, $G(F(x)) = Ix = x$. Since $F, G$ are diffeomorphisms, they are both $C^1$ on their domains, hence both differentiable on their domains by 20.5. The Jacobian of the identity is the identity itself.

$$\begin{aligned} D(G \circ F)(x) &= DG(F(x)) \circ DF(x) && \text{via chain rule} \\ I &= DG(F(x)) \circ DF(x) && \text{jacobian of the identity} \end{aligned}$$

which proves the claim, since we found a (left) inverse for $DF(x)$, hence it must be invertible. $\square$

**Theorem 27.2** (Inverse Function Theorem)

Let $A, B$ be open subsets of $\mathbb{R}^n$, with $a \in A$. Have $F : A \to B$ be $C^1$. If $DF(a)$ is invertible, then $F$ is a local diffeomorphism at $a$.

Combining with 27.1, we get that this is an if and only if relationship, meaning that a function is a local diffeomorphism at some point, if and only if it is $C^1$ and its Jacobian is invertible at said point. You would use this when computing a $C^1$ inverse seems extremely annoying/impossible.

# §28 Day 28: Nonlinear Systems (Nov 14, 2025)

Suppose we have some non-linear system of equations, given by $F : \mathbb{R}^a \to \mathbb{R}^b$.

$$F(x) = 0 \iff \begin{aligned} F_1(x) &= 0 \\ &\vdots \\ F_b(x) &= 0 \end{aligned}$$

Suppose we found the solutions to $P_i \subseteq \mathbb{R}^a$, where $F_i(P_i) = \{0\}$. We can take the intersection $\bigcap P_i$ to get the solutions for $F$. Hence we study the individual components, $F_i : \mathbb{R}^a \to \mathbb{R}$, motivating the upcoming definition.

**Definition 28.1**

Have:
- $U \subseteq \mathbb{R}^n \times \mathbb{R}^k$ be open
- $f : U \to \mathbb{R}$ be a $C^1$ function

For some $(a, b) \in U$, $a \in \mathbb{R}^n, b \in \mathbb{R}^k$, with $f(a, b) = 0$, the equation $f(x, y) = 0$ locally defines $y$ as a $C^1$ function of $x$ near $(a, b)$, if
- Exists open set $V \subseteq \mathbb{R}^n$, $a \in V$
- Exists open set $W \subseteq \mathbb{R}^k$, $b \in W$
- $V \times W \subseteq U$
- Exists $C^1$ function $\phi : V \to W$
- For all $(x, y) \in V \times W$, $f(x, y) = 0$ iff $y = \phi(x)$

The idea is that by shifting $x$ slightly such the new value for $x$ lies in $V$, $\phi$ tells you how to update $y$ such that $f(x, y)$ remains a solution. If $f(x, y) = 0$ locally defines $y$ as a $C^1$ function of $x$ near $(a, b)$, then there would exist infinitely many solutions, as $V$ is open, and the usual topology of $\mathbb{R}^{n+k}$ have all non-empty open sets contain infinitely many points.

To ensure you're on the right track when solving non-linear systems of equations, a checklist of tasks can be as follows:

1. Check that solutions exist in general:
   you now know that the system is solvable
2. Check if infinitely many solutions exist:
   you can't list all solutions, and must attempt to get a general expression
3. Check if you express variables as functions of another:
   if you parametrize every free variable you get the explicit form, which is nice
4. (For non-linear systems with $n + k$ variables and $k$ equations) Check if $k$ variables can be expressed as a $C^1$ function of the other $n$ variables:
   use the implicit function theorem

# §29 Day 29: Implicit Function Theorem (Nov 17, 2025)

We begin by stating some condition on the properties that a function that locally defines some variables in terms of the other ones must have.

**Theorem 29.1.** Let $U \subseteq \mathbb{R}^n \times \mathbb{R}$ be open, $f : U \to \mathbb{R}$ be $C^1$. $(a, b) \in \mathbb{R}^n \times \mathbb{R}$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}$. $f(a, b) = 0$, $f$ is not constant. If $f(x, y) = 0$ locally defines $y$ as a $C^1$ function $\phi : V \to W$ function of $x$ near $(a, b)$, for every point $v \in V$, $w = \phi(v)$, for every $j \in \{1, \cdots, n\}$,

$$\frac{\partial f}{\partial x_j}(v, w) + \frac{\partial f}{\partial y}(v, w)\frac{\partial \phi}{\partial x_j}(v) = 0$$

*Proof.* Let $(x, y) \in \mathbb{R}^{n+1}$. Let $\phi$ be the function that locally defines $y$ as a $C^1$ function of $x$ near $(a, b)$. Define $y = \phi(x)$, have $f(v, \phi(v)) = 0$. Since $f$ is also $C^1$, thus the chain rule applies to $h(x) := f(x, \phi(x))$. Computing,

$$\begin{aligned}
\frac{\partial h}{\partial x_j}(v) &= \frac{\partial h}{\partial x_j}(v, w)\frac{\partial x_j}{\partial x_j}(v) + \left(\sum_{\substack{k=1 \\ k \neq j}}^{n} \frac{\partial f}{\partial x_k}(v, w)\frac{\partial x_k}{\partial x_j}(v)\right) + \frac{\partial f}{\partial y}(v, w)\frac{\partial \phi}{\partial x_j}(v) \\
&= \frac{\partial h}{\partial x_j}(v, w) \cdot 1 + \frac{\partial f}{\partial y}(v, w)\frac{\partial \phi}{\partial x_j}
\end{aligned}$$

Notice that for $k \neq j$, $x_k$ is fixed in the computation of the partial w.r.t. $x_j$, so $\frac{\partial x_k}{\partial x_j}$ is 0 everywhere. Since $V$ is open, for all $v \in V$, $f(v, \phi(v)) = 0$ (constant on $V$), then $\frac{\partial h}{\partial x_j}(v) = 0$, which gives the desired result. □

**Theorem 29.2** (Implicit Function Thereom: Single Var)

Let $U \subseteq \mathbb{R}^n \times \mathbb{R}$ be open, have $f : U \to \mathbb{R}$ be $C^1$. Let $(a, b) \in U$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}$. If $f(a, b) = 0$, and $\frac{\partial f}{\partial y}(a, b) \neq 0$, then the equation $f(x, y) = 0$ locally defines $y$ as a $C^1$ function $\phi$ of $x$, with $(a, b) \in$ dom$(\phi)$.

You don't know how big $\mathrm{dom}(\phi)$ is, and just like the inverse function theorem it is non-constructive.

**Theorem 29.3** (Implicit Function Theorem: Multi Var)

Let $U \subseteq \mathbb{R}^n \times \mathbb{R}^k$ be open. Let $F : U \to \mathbb{R}^k$ be $C^1$. Let $(a, b) \in U$ so $a \in \mathbb{R}^n$, $b \in \mathbb{R}^k$. If $F(a, b) = 0$, and the $k \times k$ matrix $\frac{\partial F}{\partial y}(a, b) = \left[\frac{\partial F_i}{\partial y_j}\right]_{i,j}$ is invertible, then the equation $F(x, y) = 0$ locally defines $y$ as a $\mathbb{R}^k$ valued $C^1$ function $\phi$ of $x$ near $(a, b)$.

The implicit function informally says the following: Assume a solution exists to your nonlinear equation. If you can globally solve an approximate linear equation, then you can locally solve the nonlinear equation.

# §30 Day 30: Smooth Manifolds and Implicit Form (Nov 18, 2025)

TODO: Prove theorem 5.5.3

HINT: Recall Theorem 5.4.6 and 4.5.11

# §31 Day 32: Lagrange Multipliers (Nov 20, 2025)

# §32 Day 33: Optimization with Constraints (Nov 24, 2025)

# §33 Day 33: Second order partial derivatives and the Hessian (Nov 25, 2025)

You can take partial derivatives of partial derivatives, written $\partial_i \partial_j f(a) := \partial_i(\partial_j f)(a)$. If $i \neq j$, we say that the partial is mixed. Otherwise, we say it is pure. We write $f_{xy} = (f_x)_y = \frac{\partial}{\partial y}(\frac{\partial f}{\partial x}) = \frac{\partial f}{\partial y \partial x}$, and that $\frac{\partial}{\partial x}(\frac{\partial f}{\partial x}) = \frac{\partial^2 f}{\partial x^2}$.

**Definition 33.1** ($C^2$)**.** For open $U$, we say that $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is $C^2$ if for $i, j \in \{1, \cdots, n\}$, $\partial_i \partial_j f$ exists and are continuous everywhere in $U$.

$f$ being $C^2$ means that $f$ is $C^1$, since each $\partial_j f$ are $C^1$ themselves on $U$. Being $C^1$ implies differentiability, and differentiability implies continuity. So all $C^2$ functions are also differentiable.

**Theorem 33.2** ($C^2$ Clairaut)**.** If $f$ is $C^2$, then for $i, j \in \{1, \cdots, n\}$, $\partial_i \partial_j f = \partial_j \partial_i f$.

We say that the partials commute. There is a nice proof of this in the textbook, using the single variable MVT.

*Proof.* □

**Definition 33.3** (Hessian)**.** If a real valued $f$ is $C^2$ at some $a \in \mathbb{R}^n$. Define the Hessian of $f$ at $a$, a $n \times n$ matrix given by

$$Hf(a) = [\partial_i \partial_j f(a)]_{i,j} = \begin{pmatrix} \partial_1 \partial_1 f(a) & \cdots & \partial_1 \partial_n f(a) \\ \vdots & \ddots & \vdots \\ \partial_n \partial_1 f(a) & \cdots & \partial_n \partial_n f(a) \end{pmatrix}$$

For $\partial_i \partial_j f(a)$ to be a scalar, we need $f$ to be real-valued.

# §34 Day 34: Higher order partial derivatives (Nov 27, 2025)

The order of some partial derivative $\partial_{i_1} \cdots \partial_{i_k} f$, where each $i_\cdot \in \{1, \cdots, n\}$, is given by $k$.

**Theorem 34.1** (Clairaut)

Let $U \subseteq \mathbb{R}^n$ be open. If $f : U \to \mathbb{R}^m$ is $C^k$, then for any reordering $\{j\}_1^k$ of $\{i\}_1^k$,

$$\partial_{i_1} \cdots \partial_{i_k} f = \partial_{j_1} \cdots \partial_{j_k} f$$

## §35 Day 35: Partitions (Jan 05, 2026)

In single variable calculus, integrals are the area under a curve. $\int_a^b f(x)\,dx$ can be seen as the area between $0$ and $f$. We estimate by dividing up $[a,b]$ into partitions, where usually a finer partition (more pieces) yield a better estimate. In multi-variable calculus, we attempt to do the same.

**Definition 35.1** (Rectangle)**.** A rectangle in $\mathbb{R}^n$ is a set $\mathbb{R} = [a_1, b_1,] \times \cdots \times [a_n, b_n]$, where for each $i$, $a_i < b_i$. The volume is given by $\prod_i (b_i - a_i)$.

**Definition 35.2** (Partition)

An ordered $n$-tuple of sets $P = (P_1, \cdots, P_n)$ is a partition of $R$, if $P_j = \{x_0^{(j)}, \cdots, x_{k_j}^{(j)}\}$ is a partition of the interval $[a_j, b_j]$ for each $j$ The index set of $P$ is the set

$$I = \{(i_1, \cdots, i_n) \in \mathbb{N}^n : 1 \leq i_1 \leq k_1, \cdots, 1 \leq i_n \leq k_n\}$$

The $i$-subrectangle of $P$, with $i = (i_1, \cdots, i_n) \in I$ is given by

$$R_i = [x_{i_1-1}^{(1)}, x_{i_1}^{(1)}] \times \cdots \times [x_{i_n-1}^{(n)}, x_{i_n}^{(n)}]$$

The index set does not contain tuples containing 0.

**Theorem 35.3.** Let $P$ be a partition of some rectangle $R$, with $I$ being the index set of $P$ and $\{R_i : i \in I\}$ be the subrectangles.

- $\bigcup_{i\in I} R_i = R$
- For all $i, j \in I$, if $i \neq j$, $\text{int}(R_i) \cap \int(R_j) = \emptyset$
- $\sum_{i\in I} \text{vol}(R_i) = \text{vol}(R)$

# §36 Day 36: Upper and Lower Sums (Jan 06, 2026)

Let $R$ be a rectangle in $\mathbb{R}^n$, and let $f : R \to \mathbb{R}$ be a bounded function. Let $P$ be a partition of $R$ with index set $I$, with subrectangles $\{R_i : i \in I\}$. Let $P'$ be a refinement of $P$.

**Definition 36.1** (Lower and Upper Sums)**.** The $P$-lower and $P$-upper sums of $f$ are respectively given by

$$L_P(f) = \sum_{i \in I} m_i \operatorname{vol}(R_i), \quad U_P(f) = \sum_{i \in I} M_i \operatorname{vol}(R_i)$$

where for $i \in I$,

$$m_i = \inf\{f(x) : x \in R_i\}, \quad M_i = \sup\{f(x) : x \in R_i\}$$

**Definition 36.2** (Riemann Sum)**.** For each $i \in I$, a sample point $x_i^* \in R_i$ is chosen. The Riemann sum for $f$ with $P$ and these sample points is given by

$$S_P^*(f) = \sum_{i \in I} f(x_i^*) \operatorname{vol}(R_i)$$

**Theorem 36.3**

Properties of partitions include:

- $L_P(f) \leq U_P(f)$
- $L_P(f) \leq L_{P'}(f)$, $U_{P'}(f) \leq U_P(f)$
- For any two partitions $A, B$ of a rectangle, $L_A(f) \leq U_B(f)$
- $U_P(f + g) \leq U_P(f) + U_P(g)$
- For $\lambda > 0$, $U_P(\lambda f) = \lambda U_P(f)$
- $U_P(-f) = -L_P(f)$
- If $f \leq g$ on $R$, then $U_P(f) \leq U_P(g)$
- $S_P^*(f + \lambda g) = S_P^*(f) + \lambda S_P^*(g)$
- If $f \leq g$ on $R$, then $S_P^*(f) \leq S_P^*(g)$

## §37 Day 37: Integration over Rectangles (Jan 07, 2026)

**Definition 37.1.** The lower and upper integral of $f$ on $R$ are defined by

$$\underline{I_R}(f) = \sup_P L_P(f), \quad \overline{I_R}(f) = \inf_P L_P(f)$$

**Theorem 37.2.** If $f : R \to \mathbb{R}$ is a bounded function, then both $\underline{I_R}(f)$ and $\overline{I_R}(f)$ exist, and $\underline{I_R}(f) \leq \overline{I_R}(f)$

**Definition 37.3** (Darboux Integral)**.** For bounded $f : R \to \mathbb{R}$, if $\underline{I_R}(f) = \overline{I_R}(f)$ then $f$ is integrable on $R$ and the integral of $f$ on $R$ is defined by

$$\int_R f\, dV := \underline{I_R}(f) = \overline{I_R}(f)$$

otherwise $f$ is non-integrable.

$dV$ denotes the volume element and is just notation.

TODO: integral properties