Week 2

Lecture 1

Sets, Logic, Metric Spaces and Visualization

---

1. Overview of This Week

In this week, we develop the mathematical tools that we will rely on throughout the course.

We focus primarily on calculus, with an emphasis that is more geometric and visual compared to standard treatments.

We begin with foundational concepts:

- Sets and functions
- Mathematical notation
- Logic
- Graphical visualization of functions

---

2. Basic Sets

We repeatedly use the following fundamental sets.

Real Numbers

$$\mathbb{R}$$

We denote by $\mathbb{R}$ the set of all real numbers.

Positive Real Numbers (Including Zero)

$$\mathbb{R}_+$$

We define $\mathbb{R}_+$ as the set of non-negative real numbers, including 0.

Integers

$$\mathbb{Z}$$

We denote by $\mathbb{Z}$ the set of all integers.

Non-Negative Integers

$$\mathbb{Z}_+$$

We define $\mathbb{Z}_+$ as the set of non-negative integers, including 0.

---

3. Intervals

Closed Interval

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$$

We include both endpoints.

Open Interval

$$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$$

We exclude both endpoints.

---

4. Cartesian Products

We form higher-dimensional sets using Cartesian products.

d-Dimensional Real Space

$$\mathbb{R}^d = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$$

An element of $\mathbb{R}^3$ is:

$$(1, 2, 3)$$

---

d-Dimensional Box

$$[a,b]^d = \{x \in \mathbb{R}^d \mid x_i \in [a,b] \text{ for all } i = 1, \ldots, d\}$$

Each coordinate lies within $[a,b]$.

---

5. Metric Spaces

We define a metric space as a set equipped with a distance function.

Euclidean Distance in $\mathbb{R}^d$

$$d(x,y) = \|x - y\|$$

Where:

$$\|x - y\| = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

This is the standard Euclidean metric.

Open Ball

$$B(x, \epsilon) = \{y \in \mathbb{R}^d \mid d(x,y) < \epsilon\}$$

We exclude the boundary.

Closed Ball

$$\overline{B}(x, \epsilon) = \{y \in \mathbb{R}^d \mid d(x,y) \leq \epsilon\}$$

We include the boundary.

---

6. Sets and Logic

Let V denote a universe.
We define:

- Union: $A \cup B$
- Intersection: $A \cap B$

- Complement: $A^c = V \setminus A$

De Morgan's Laws

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

Example

Let:

$$V = [0, 10]$$

$$A = [2, 5]$$

$$B = [4, 7]$$

Then:

$$A \cup B = [2, 7]$$

$$A \cap B = [4, 5]$$

---

7. Logical Quantifiers

We use the following symbols:

- $\forall$: for all
- $\exists$: there exists
- $\Rightarrow$: implies
- $\Leftrightarrow$: equivalent

---

8. Sequences

A sequence is an ordered collection:

$$x_1, x_2, \ldots$$

Where:

$$x_i \in \mathbb{R}^d$$

Convergence

We say:

$$\lim_{i \to \infty} x_i = x^*$$

If:

$$\forall \epsilon > 0, \exists N \in \mathbb{Z}_+ \text{ such that } x_n \in B(x^*, \epsilon) \text{ for all } n \geq N$$

This means that eventually all points lie inside every ϵ-ball around x^*.

---

9. Vector Spaces

A vector space V satisfies:
If:

$$u, v \in V$$

and:

$$\alpha, \beta \in \mathbb{R}$$

Then:

$$\alpha u + \beta v \in V$$

Dot Product

For $x, y \in \mathbb{R}^d$:

$$x \cdot y = x^\top y = \sum_{i=1}^{d} x_i y_i$$

Norm

$$\|x\|^2 = x \cdot x = \sum_{i=1}^{d} x_i^2$$

Orthogonality

We say x and y are orthogonal if:

$$x^\top y = 0$$

10. Functions

A function:

$$f : A \to B$$

- A = domain
- B = codomain

1-Dimensional Function

$$f : \mathbb{R} \to \mathbb{R}$$

Example:

$$f(x) = x^2$$

Graph:

$$G_f = \{(x, f(x)) \mid x \in \mathbb{R}\}$$

d-Dimensional Function

$$f : \mathbb{R}^d \to \mathbb{R}$$

Graph:

$$G_f = \{(x, f(x)) \mid x \in \mathbb{R}^d\}$$

Where:

$$G_f \subset \mathbb{R}^{d+1}$$

11. Contour Plots (2D Functions)

For:

$$f : \mathbb{R}^2 \to \mathbb{R}$$

Example:

$$f(x_1, x_2) = x_1 + x_2$$

Contours are defined by:

$$f(x_1, x_2) = c$$

For various constants c.

Another Example

$$f(x_1, x_2) = x_1 x_2$$

Contours form hyperbolas.

Circular Example

$$f(x_1, x_2) = x_1^2 + x_2^2$$

Contours are circles.

Summary

In this lecture, we introduced:

- Core sets and interval notation
- Metric spaces and Euclidean distance
- Logical operations and quantifiers
- Sequences and convergence
- Vector spaces, dot products and orthogonality
- Functions and their graphical representation
- Contour plots and heat maps

These tools form the mathematical foundation required for the rest of the course.

****

Lecture 2

Univariate Calculus: Continuity and Differentiability

Continuity of Functions

In this lecture, we restrict ourselves to **real-valued one-dimensional functions**:

$$f : \mathbb{R} \to \mathbb{R}$$

Definition — Continuity at a Point

We say that a function $f : \mathbb{R} \to \mathbb{R}$ is **continuous at a point** $x^\star \in \mathbb{R}$ if for **all sequences** $\{x_i\}_{i=1}^\infty$ converging to $x^\star$, we have:

$$\lim_{i \to \infty} x_i = x^\star \;\Rightarrow\; \lim_{i \to \infty} f(x_i) = f(x^\star)$$

Equivalently, we write:

$$\lim_{x \to x^\star} f(x) = f(x^\star)$$

Example 1 — Continuous Function

Let us consider:

$$f(x) = x^2$$

We test continuity at $x^\star = 2$.
Let us take a sequence converging to 2:

$$x_i = 3,\ 2.5,\ 2.25,\ \ldots$$

Then:

$$f(x_i) = 9,\ 6.25,\ (2.25)^2,\ \ldots$$

We observe that:

$$\lim_{i \to \infty} x_i = 2 \quad \text{and} \quad \lim_{i \to \infty} f(x_i) = 4 = f(2)$$

Hence, $f(x) = x^2$ is continuous at $x^\star = 2$.
In fact, we know that x^2 is continuous everywhere.

Example 2 — Discontinuous Function

Let us define:

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

We test continuity at $x^\star = 0$.

Sequence 1 (approaching from the right)

$$x_i = 1,\ \tfrac{1}{2},\ \tfrac{1}{4},\ \dots$$

Then:

$$f(x_i) = 1,\ 1,\ 1,\ \dots$$

Thus,

$$\lim_{i \to \infty} f(x_i) = 1$$

Sequence 2 (approaching from the left)

$$x_i = -1,\ -\tfrac{1}{2},\ -\tfrac{1}{4},\ \dots$$

Then:

$$f(x_i) = -1,\ -1,\ -1,\ \dots$$

Thus,

$$\lim_{i \to \infty} f(x_i) = -1$$

Since the limits disagree and are not equal to $f(0) = 0$, the function is **not continuous at 0**.

Example 3 — Piecewise Function

Let us define:

$$f(x) = \begin{cases} 2x + 1 & x > 1 \\ 3 & x \le 1 \end{cases}$$

We observe that:

- From the right at $x = 1$: $2(1) + 1 = 3$
- From the left at $x = 1$: value is 3

Since both agree with $f(1) = 3$, we conclude that the function is continuous at $x = 1$.

Example 4 — $f(x) = \frac{1}{x}$

Let us consider:

$$f(x) = \frac{1}{x}$$

Take the sequence:

$$x_i = 1,\ \tfrac{1}{2},\ \tfrac{1}{4},\ \ldots$$

Then:

$$f(x_i) = 1,\ 2,\ 4,\ 8,\ \ldots$$

This diverges.
Hence $f(x)$ is **not continuous at 0**.
However, it is continuous on any domain that excludes 0.

Example 5 — Oscillatory Discontinuity

Let us consider:

$$f(x) = \cos\left(\frac{1}{x}\right)$$

Using:

$$x_i = 1,\ \tfrac{1}{2},\ \tfrac{1}{4},\ \ldots$$

we get:

$$f(x_i) = \cos(1),\ \cos(2),\ \cos(4),\ \cos(8),\ \ldots$$

This sequence does not converge.
Hence the function is not continuous at 0.

Differentiability of Functions

Definition — Differentiability at a Point

We say that $f : \mathbb{R} \to \mathbb{R}$ is **differentiable at** $x^\star$ if the limit

$$\lim_{x \to x^\star} \frac{f(x) - f(x^\star)}{x - x^\star}$$

exists.
If this limit exists, we denote it by:

$$f'(x^\star)$$

Key Implication

If f is **not continuous** at $x^\star$, then f is **not differentiable** at $x^\star$.
However, the converse is false: a function can be continuous but not differentiable.

Example — $f(x) = |x|$

Let:

$$f(x) = |x|$$

Consider:

$$x_i = 1,\ \tfrac{1}{2},\ \tfrac{1}{4},\ \dots$$

Then:

$$\frac{f(x_i) - f(0)}{x_i} = \frac{x_i}{x_i} = 1$$

Now consider:

$$x_i = -1,\ -\tfrac{1}{2},\ -\tfrac{1}{4},\ \dots$$

Then:

$$\frac{f(x_i) - f(0)}{x_i} = \frac{-x_i}{x_i} = -1$$

Since the limits differ, the derivative does not exist at 0.
Thus $|x|$ is **continuous but not differentiable at 0**.

Piecewise Example — Not Continuous

Let:

$$f(x) = \begin{cases} 4x + 2 & x \geq 2 \\ 2x + 8 & x < 2 \end{cases}$$

Left and right limits differ at $x = 2$.
Hence f is not continuous and therefore not differentiable at 2.

Piecewise Example — Continuous but Not Differentiable

Let:

$$f(x) = \begin{cases} 4x + 2 & x \geq 2 \\ 2x + 6 & x < 2 \end{cases}$$

This function is continuous at 2, but:

$$\lim_{x \to 2^+} \frac{f(x) - f(2)}{x - 2} = 4$$

$$\lim_{x \to 2^-} \frac{f(x) - f(2)}{x - 2} = 2$$

Since the limits differ, f is not differentiable at 2.

Alternate Expression for the Derivative

We can rewrite the derivative as:

$$f'(x^\star) = \lim_{h \to 0} \frac{f(x^\star + h) - f(x^\star)}{h}$$

Geometric Interpretation

We interpret:

$$\frac{f(x^\star + h) - f(x^\star)}{h}$$

as the slope of the secant line between:

$$(x^\star, f(x^\star)) \quad \text{and} \quad (x^\star + h, f(x^\star + h))$$

As $h \to 0$, the secant line approaches the tangent line.
Thus:

$$f'(x^\star)$$

represents the **slope of the tangent line** to the curve at $x^\star$.
With this, we complete our recap of continuity and differentiability in one dimension.

```
********************************************************************************
```

Lecture 3

Univariate Calculus: Continuity and Differentiability

Continuity of Functions

In this lecture, we restrict ourselves to **real-valued one-dimensional functions**:

$$f : \mathbb{R} \to \mathbb{R}$$

Definition — Continuity at a Point

We say that a function $f : \mathbb{R} \to \mathbb{R}$ is **continuous at a point** $x^\star \in \mathbb{R}$ if for **all sequences** $\{x_i\}_{i=1}^{\infty}$ converging to $x^\star$, we have:

$$\lim_{i \to \infty} x_i = x^\star \;\Rightarrow\; \lim_{i \to \infty} f(x_i) = f(x^\star)$$

Equivalently, we write:

$$\lim_{x \to x^\star} f(x) = f(x^\star)$$

Example 1 — Continuous Function

Let us consider:

$$f(x) = x^2$$

We test continuity at $x^\star = 2$.
Let us take a sequence converging to 2:

$$x_i = 3,\ 2.5,\ 2.25,\ \dots$$

Then:

$$f(x_i) = 9,\ 6.25,\ (2.25)^2,\ \dots$$

We observe that:

$$\lim_{i \to \infty} x_i = 2 \quad \text{and} \quad \lim_{i \to \infty} f(x_i) = 4 = f(2)$$

Hence, $f(x) = x^2$ is continuous at $x^\star = 2$.
In fact, we know that x^2 is continuous everywhere.

Example 2 — Discontinuous Function

Let us define:

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

We test continuity at $x^\star = 0$.

Sequence 1 (approaching from the right)

$$x_i = 1,\ \tfrac{1}{2},\ \tfrac{1}{4},\ \dots$$

Then:

$$f(x_i) = 1,\ 1,\ 1,\ \dots$$

Thus,

$$\lim_{i \to \infty} f(x_i) = 1$$

Sequence 2 (approaching from the left)

$$x_i = -1,\ -\tfrac{1}{2},\ -\tfrac{1}{4},\ \dots$$

Then:

$$f(x_i) = -1,\ -1,\ -1,\ \dots$$

Thus,

$$\lim_{i \to \infty} f(x_i) = -1$$

Since the limits disagree and are not equal to $f(0) = 0$, the function is **not continuous at 0**.

Example 3 — Piecewise Function

Let us define:

$$f(x) = \begin{cases} 2x+1 & x > 1 \\ 3 & x \leq 1 \end{cases}$$

We observe that:

- From the right at $x = 1$: $2(1) + 1 = 3$
- From the left at $x = 1$: value is 3

Since both agree with $f(1) = 3$, we conclude that the function is continuous at $x = 1$.

Example 4 — $f(x) = \frac{1}{x}$

Let us consider:

$$f(x) = \frac{1}{x}$$

Take the sequence:

$$x_i = 1,\ \tfrac{1}{2},\ \tfrac{1}{4},\ \ldots$$

Then:

$$f(x_i) = 1,\ 2,\ 4,\ 8,\ \ldots$$

This diverges.
Hence $f(x)$ is **not continuous at 0**.
However, it is continuous on any domain that excludes 0.

Example 5 — Oscillatory Discontinuity

Let us consider:

$$f(x) = \cos\left(\frac{1}{x}\right)$$

Using:

$$x_i = 1,\ \tfrac{1}{2},\ \tfrac{1}{4},\ \ldots$$

we get:

$$f(x_i) = \cos(1),\ \cos(2),\ \cos(4),\ \cos(8),\ \dots$$

This sequence does not converge.
Hence the function is not continuous at 0.

Differentiability of Functions

Definition — Differentiability at a Point

We say that $f : \mathbb{R} \to \mathbb{R}$ is **differentiable at** $x^\star$ if the limit

$$\lim_{x \to x^\star} \frac{f(x) - f(x^\star)}{x - x^\star}$$

exists.
If this limit exists, we denote it by:

$$f'(x^\star)$$

Key Implication

If f is **not continuous** at $x^\star$, then f is **not differentiable** at $x^\star$.
However, the converse is false: a function can be continuous but not differentiable.

Example — $f(x) = |x|$

Let:

$$f(x) = |x|$$

Consider:

$$x_i = 1,\ \tfrac{1}{2},\ \tfrac{1}{4},\ \dots$$

Then:

$$\frac{f(x_i) - f(0)}{x_i} = \frac{x_i}{x_i} = 1$$

Now consider:

$$x_i = -1,\ -\tfrac{1}{2},\ -\tfrac{1}{4},\ \dots$$

Then:

$$\frac{f(x_i) - f(0)}{x_i} = \frac{-x_i}{x_i} = -1$$

Since the limits differ, the derivative does not exist at 0.
Thus $|x|$ is **continuous but not differentiable at 0**.

Piecewise Example — Not Continuous

Let:

$$f(x) = \begin{cases} 4x + 2 & x \geq 2 \\ 2x + 8 & x < 2 \end{cases}$$

Left and right limits differ at $x = 2$.
Hence f is not continuous and therefore not differentiable at 2.

Piecewise Example — Continuous but Not Differentiable

Let:

$$f(x) = \begin{cases} 4x + 2 & x \geq 2 \\ 2x + 6 & x < 2 \end{cases}$$

This function is continuous at 2, but:

$$\lim_{x \to 2^+} \frac{f(x) - f(2)}{x - 2} = 4$$

$$\lim_{x \to 2^-} \frac{f(x) - f(2)}{x - 2} = 2$$

Since the limits differ, f is not differentiable at 2.

Alternate Expression for the Derivative

We can rewrite the derivative as:

$$f'(x^\star) = \lim_{h \to 0} \frac{f(x^\star + h) - f(x^\star)}{h}$$

Geometric Interpretation

We interpret:

$$\frac{f(x^\star + h) - f(x^\star)}{h}$$

as the slope of the secant line between:

$$(x^\star, f(x^\star)) \quad \text{and} \quad (x^\star + h, f(x^\star + h))$$

As $h \to 0$, the secant line approaches the tangent line.

Thus:

$$f'(x^\star)$$

represents the **slope of the tangent line** to the curve at $x^\star$.

With this, we complete our recap of continuity and differentiability in one dimension.

```
*********************************************************************************
```

Lecture 4

Univariate Calculus – Derivatives and Linear Approximations

1. Derivative of a Differentiable Function

Let us consider a differentiable function

$$f : \mathbb{R} \to \mathbb{R}.$$

We define the derivative of f at a point $x^\star$ as

$$f'(x^\star) = \lim_{x \to x^\star} \frac{f(x) - f(x^\star)}{x - x^\star}.$$

We observe that this definition involves a limit, so we cannot directly manipulate the expression algebraically without care.
However, if we restrict ourselves to values of x that are very close to $x^\star$, we may treat

$$\frac{f(x) - f(x^\star)}{x - x^\star} \approx f'(x^\star) \quad (\text{for } x \text{ near } x^\star).$$

Rearranging, we obtain

$$f(x) \approx f(x^\star) + f'(x^\star)(x - x^\star), \quad \text{for } x \approx x^\star.$$

---

2. Linear Approximation

We define the linear approximation of f around $x^\star$ as

$$L_{x^\star}[f](x) = f(x^\star) + f'(x^\star)(x - x^\star).$$

Thus, we write

$$f(x) \approx L_{x^\star}[f](x), \quad \text{around } x = x^\star.$$

We note that:

- $f(x^\star)$ is a constant.
- $f'(x^\star)$ is also a constant.
- The right-hand side is linear in x.
  Therefore, $L_{x^\star}[f]$ is a linear function of x that approximates f near $x^\star$.

---

3. Example: Quadratic Function

Let us consider

$$f(x) = x^2.$$

We choose

$$x^\star = 1.$$

We compute:

$$f'(x) = 2x,$$

so

$$f'(1) = 2.$$

Now we construct the linear approximation:

$$L_1[f](x) = f(1) + f'(1)(x - 1).$$

Since

$$f(1) = 1,$$

we obtain

$$L_1[f](x) = 1 + 2(x - 1) = 2x - 1.$$

Thus,

$$f(x) \approx 2x - 1, \quad \text{around } x = 1.$$

We interpret this as follows:

- The graph of $y = 2x - 1$ is a straight line.
- This line touches the curve $y = x^2$ at the point $(1, 1)$.
- Near $x = 1$, the line closely approximates the curve.
- As we move farther from 1, the approximation deteriorates.

---

4. Tangent Line Interpretation

We observe that the graph of the linear approximation

$$G_{L_{x^\star}[f]}$$

is a subset of $\mathbb{R}^2$.
We say that this line is tangent to the graph of f, denoted G_f, at the point

$$(x^\star, f(x^\star)).$$

Thus, the linear approximation corresponds geometrically to the tangent line to the curve at that point.

---

5. Classical Linear Approximations

(i) $\sin x$ around $x^\star = 0$

Let

$$f(x) = \sin x.$$

We compute

$$f'(x) = \cos x, \quad f'(0) = 1, \quad f(0) = 0.$$

Thus,

$$\sin x \approx 0 + 1(x - 0) = x, \quad \text{around } x = 0.$$

(ii) e^x around $x^\star = 0$

Let

$$f(x) = e^x.$$

We compute

$$f'(x) = e^x, \quad f'(0) = 1, \quad f(0) = 1.$$

Thus,

$$e^x \approx 1 + x, \quad \text{around } x = 0.$$

(iii) $\log(1 + x)$ around $x^\star = 0$

Let

$$f(x) = \log(1 + x).$$

We compute

$$f'(x) = \frac{1}{1 + x}, \quad f'(0) = 1, \quad f(0) = 0.$$

Thus,

$$\log(1 + x) \approx x, \quad \text{around } x = 0.$$

(iv) $(1 + x)^r$ around $x^\star = 0$

Let

$$f(x) = (1 + x)^r,$$

where r is a constant.
We compute

$$f'(x) = r(1 + x)^{r-1}, \quad f'(0) = r, \quad f(0) = 1.$$

Thus,

$$(1+x)^r \approx 1 + rx, \quad \text{around } x = 0.$$

---

6. Application: Approximating 0.99^7

We observe that

$$0.99^7 = (1 - 0.01)^7.$$

Using the approximation

$$(1+x)^r \approx 1 + rx \quad \text{for small } x,$$

we set

$$x = -0.01, \quad r = 7.$$

Thus,

$$(0.99)^7 \approx 1 + 7(-0.01) = 1 - 0.07 = 0.93.$$

Hence, among the options

- 0.95
- 0.93
- 0.91
- 0.9
  we conclude that

  $$0.93$$

  is the closest approximation.

Core Insight

We summarize the key principle:
For a differentiable function,

$$f(x) \approx f(x^\star) + f'(x^\star)(x - x^\star), \quad \text{near } x = x^\star.$$

We recognize that linear approximation simplifies complicated functions into linear forms locally, and this idea forms the foundation of optimization and machine learning.

```
********************************************************************************
```

Lecture 5

Applications and Advanced Rules of Derivatives

1. Higher Order Approximations

We have previously defined the **linear approximation** of a differentiable function $f : \mathbb{R} \to \mathbb{R}$ around a point $x^\star$ as

$$f(x) \approx f(x^\star) + f'(x^\star)(x - x^\star)$$

We now observe that there is nothing fundamentally special about stopping at the first derivative. We can incorporate higher order derivatives to obtain better approximations.

The **quadratic approximation** of f around $x^\star$ is given by

$$f(x) \approx f(x^\star) + f'(x^\star)(x - x^\star) + \frac{1}{2} f''(x^\star)(x - x^\star)^2$$

We emphasize that both approximations are valid when

$$x \approx x^\star$$

The quadratic approximation is generally more accurate than the linear approximation, but it introduces additional computational complexity.

Example 1: Exactness for Quadratic Functions

Let us consider

$$f(x) = x^2$$

We compute

$$f'(x) = 2x$$

$$f''(x) = 2$$

Using the quadratic approximation around an arbitrary $x^\star$, we obtain

$$x^2 \approx (x^\star)^2 + 2x^\star(x - x^\star) + \frac{1}{2} \cdot 2(x - x^\star)^2$$

Simplifying,

$$x^2 = (x^\star)^2 + 2x^\star(x - x^\star) + (x - x^\star)^2$$

which exactly equals x^2.
We conclude that for quadratic functions, the quadratic approximation recovers the function exactly.

Example 2: Quadratic Approximation of e^x Around 0

We consider

$$f(x) = e^x$$

We compute

$$f(0) = 1, \quad f'(x) = e^x, \quad f'(0) = 1, \quad f''(0) = 1$$

The quadratic approximation around 0 becomes

$$e^x \approx 1 + x + \frac{x^2}{2} \quad \text{for } x \approx 0$$

We recognize this as the first three terms of the Taylor expansion of e^x.

Application: Approximating $(1.1)^7$

We wish to approximate

$$(1.1)^7 = (1 + 0.1)^7$$

Let

$$f(x) = (1 + x)^7$$

We compute

$$f'(x) = 7(1 + x)^6$$

$$f''(x) = 42(1 + x)^5$$

At $x = 0$,

$$f(0) = 1, \quad f'(0) = 7, \quad f''(0) = 42$$

Using quadratic approximation at $x = 0.1$,

$$f(0.1) \approx 1 + 7(0.1) + \frac{1}{2} \cdot 42(0.1)^2$$

$$= 1 + 0.7 + 21(0.01)$$

$$= 1 + 0.7 + 0.21$$

$$= 1.91$$

We observe that this value is significantly closer to the true value than the linear approximation

$$1 + 7(0.1) = 1.7$$

---

2. Product Rule via Linear Approximation

Let

$$f(x) = g(x)h(x)$$

We approximate both g and h around $x = 0$:

$$g(x) \approx g(0) + g'(0)x$$

$$h(x) \approx h(0) + h'(0)x$$

Multiplying,

$$f(x) \approx (g(0) + g'(0)x)(h(0) + h'(0)x)$$

Expanding,

$$= g(0)h(0) + x[g'(0)h(0) + h'(0)g(0)] + x^2 g'(0)h'(0)$$

Ignoring the quadratic term for linear approximation,

$$f(x) \approx f(0) + x[g'(0)h(0) + h'(0)g(0)]$$

By matching with

$$f(x) \approx f(0) + f'(0)x$$

we obtain the **product rule**

$$f'(0) = g'(0)h(0) + h'(0)g(0)$$

Generalizing,

$$f'(x) = g'(x)h(x) + g(x)h'(x)$$

---

3. Chain Rule via Linear Approximation

Let

$$f(x) = g(h(x))$$

We first approximate

$$h(x) \approx h(0) + h'(0)x$$

Then,

$$g(h(x)) \approx g(h(0)) + g'(h(0))[h(x) - h(0)]$$

Substituting the approximation of $h(x)$,

$$f(x) \approx g(h(0)) + g'(h(0))h'(0)x$$

Matching with

$$f(x) \approx f(0) + f'(0)x$$

we obtain the **chain rule**

$$f'(0) = g'(h(0))h'(0)$$

In general,

$$\frac{d}{dx} g(h(x)) = g'(h(x))h'(x)$$

---

4. Linear Approximation Examples

Example 1

Approximate

$$\frac{e^{3x}}{\sqrt{1+x}}$$

around $x = 0$.
We use

$$e^{3x} \approx 1 + 3x$$

$$(1+x)^{-1/2} \approx 1 - \frac{x}{2}$$

Multiplying,

$$\frac{e^{3x}}{\sqrt{1+x}} \approx (1+3x)\left(1 - \frac{x}{2}\right)$$

Ignoring quadratic terms,

$$\approx 1 + \frac{5}{2}x \quad \text{for } x \approx 0$$

Example 2

Approximate

$$e^{\sqrt{1+x}}$$

around $x = 1$.
We compute

$$f(1) = e^{\sqrt{2}}$$

$$f'(x) = e^{\sqrt{1+x}} \cdot \frac{1}{2\sqrt{1+x}}$$

$$f'(1) = \frac{e^{\sqrt{2}}}{2\sqrt{2}}$$

Thus the linear approximation is

$$e^{\sqrt{1+x}} \approx e^{\sqrt{2}} + \frac{e^{\sqrt{2}}}{2\sqrt{2}}(x-1) \quad \text{for } x \approx 1$$

---

5. Critical Points, Maxima, Minima, Saddle Points

The linear approximation of f around $x^\star$ is

$$L_{x^\star}[f](x) = f(x^\star) + f'(x^\star)(x - x^\star)$$

If

$$f'(x^\star) = 0$$

then

$$L_{x^\star}[f](x) = f(x^\star)$$

which means the approximation is constant.
We define

$$f'(x^\star) = 0 \iff x^\star \text{ is a critical point of } f$$

In one dimension, a critical point may correspond to:

- a local minimum
- a local maximum
- a saddle point
  These points are fundamental in machine learning because optimization problems reduce to finding points where

$$f'(x) = 0$$

Closing Remark

We conclude that linear approximation is the central computational tool underlying:

- product rule
- chain rule
- Taylor approximations
- optimization
- critical point analysis

This completes the univariate calculus component. We next transition to multivariate calculus.

```
*****************************************************************************************
*
```

Lecture 6

Multivariate Calculus: Lines, Hyperplanes, Partial Derivatives, and Gradients

1. From Univariate to Multivariate Functions

We now generalize our setting from univariate calculus to multivariate calculus.
Previously, we studied functions of the form

$$f : \mathbb{R} \to \mathbb{R}$$

We now consider multivariate functions of the form

$$f : \mathbb{R}^d \to \mathbb{R}$$

Here:

- The input is a d-dimensional vector.
- The output remains a real number.
- We continue to focus on differential calculus.
  Before studying derivatives in this setting, we first understand the geometry of $\mathbb{R}^d$.

---

2. Geometry in $\mathbb{R}^d$

2.1 Lines in $\mathbb{R}^d$

A line in $\mathbb{R}^d$ is a subset of $\mathbb{R}^d$.

Definition 1 — Line Through a Point Along a Direction

Given:

- A point $u \in \mathbb{R}^d$
- A direction vector $v \in \mathbb{R}^d$

We define the line through u along v as:

$$\{x \in \mathbb{R}^d \mid x = u + \alpha v,\ \alpha \in \mathbb{R}\}$$

We vary α over all real numbers to trace the entire line.

Definition 2 — Line Through Two Points

Given two distinct points $u, u' \in \mathbb{R}^d$, we define the line through them as:

$$\{x \in \mathbb{R}^d \mid x = u + \alpha(u' - u),\ \alpha \in \mathbb{R}\}$$

Equivalently,

$$x = (1 - \alpha)u + \alpha u'$$

We observe:

- The line through u and u'
- The line through u along direction $u' - u$
- The line through u' along direction $u - u'$
  all define the same subset of $\mathbb{R}^d$.

2.2 Hyperplanes in $\mathbb{R}^d$

A hyperplane in $\mathbb{R}^d$ is a $(d-1)$-dimensional subset of $\mathbb{R}^d$.

Definition — Hyperplane Normal to a Vector

Given:

- A vector $w \in \mathbb{R}^d$
- A scalar $b \in \mathbb{R}$

We define the hyperplane as:

$$\{x \in \mathbb{R}^d \mid w^T x = b\}$$

In expanded form:

$$\left\{x \in \mathbb{R}^d \mid \sum_{i=1}^{d} w_i x_i = b\right\}$$

The vector w is normal (perpendicular) to the hyperplane.

---

3. Points vs Vectors vs Tuples

In $\mathbb{R}^d$, we represent:

- Points (locations)
- Vectors (directions)
- Tuples (data representations)
  all using d-dimensional coordinate arrays.

Geometrically:

- A point represents a location.
- A vector represents a direction and magnitude.
  Algebraically, both are elements of $\mathbb{R}^d$.
  Context determines interpretation.

---

4. Partial Derivatives

Let us consider a function:

$$f : \mathbb{R}^d \to \mathbb{R}$$

Let $v \in \mathbb{R}^d$.

4.1 Definition of Partial Derivative

The partial derivative with respect to the i-th coordinate is defined as:

$$\frac{\partial f}{\partial x_i}(v) = \lim_{\alpha \to 0} \frac{f(v + \alpha e_i) - f(v)}{\alpha}$$

where:

- e_i is the i-th coordinate vector.
- e_i has 1 in the i-th position and 0 elsewhere.

This definition generalizes the univariate derivative.
We vary only one coordinate and keep the others fixed.

4.2 Example

Let:

$$f(x_1, x_2) = x_1^2 + x_2^2$$

Then:

$$\frac{\partial f}{\partial x_1} = 2x_1$$

$$\frac{\partial f}{\partial x_2} = 2x_2$$

Evaluated at $v = (v_1, v_2)$:

$$\frac{\partial f}{\partial x_1}(v) = 2v_1$$

$$\frac{\partial f}{\partial x_2}(v) = 2v_2$$

---

5. Gradient

We collect all partial derivatives into a single vector.

5.1 Derivative with Respect to a Vector

We define:

$$\frac{\partial f}{\partial x}(v) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(v) & \frac{\partial f}{\partial x_2}(v) & \cdots & \frac{\partial f}{\partial x_d}(v) \end{bmatrix}$$

This is a row vector.

5.2 Gradient

The gradient is defined as:

$$\nabla f(v) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(v) \\ \frac{\partial f}{\partial x_2}(v) \\ \vdots \\ \frac{\partial f}{\partial x_d}(v) \end{bmatrix}$$

This is a column vector.
The gradient is simply the transpose of the derivative row vector.

5.3 Example — Quadratic Function

For:

$$f(x_1, x_2) = x_1^2 + x_2^2$$

We obtain:

$$\nabla f(v) = \begin{bmatrix} 2v_1 \\ 2v_2 \end{bmatrix}$$

5.4 Example — Linear Function

Let:

$$f(x_1, x_2, x_3) = x_1 + 2x_2 + 3x_3$$

Then:

$$\nabla f(v) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

The gradient is constant.
We observe that linear functions have constant gradients.

---

6. Conceptual Interpretation

We note:

- A partial derivative measures variation along a coordinate direction.
- The gradient collects all directional rates of change.
- The gradient will later connect to optimization.
- Critical points occur when:

  $$\nabla f(v) = 0$$

  These points are candidates for minima, maxima, or saddle points.
  We now have the foundational geometric and differential tools required for multivariate calculus, which we will use extensively in machine learning for optimization and modeling.

```
*****************************************************************************************
```

Lecture 7

Gradient Interpretations and Higher-Order Approximations

Multivariate Linear Approximation

We now extend the idea of linear approximation from functions
$f : \mathbb{R} \to \mathbb{R}$
to functions
$f : \mathbb{R}^d \to \mathbb{R}$.
In one dimension, we approximate around $x^\star$ as:

$$f(x) \approx f(x^\star) + f'(x^\star)(x - x^\star)$$

We now generalize this to higher dimensions.
Let $v \in \mathbb{R}^d$ and let x be close to v.
We approximate:

$$f(x) \approx f(v) + \nabla f(v)^\top (x - v)$$

Expanding the gradient into components:

$$f(x) \approx f(v) + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(v)(x_i - v_i)$$

We denote this approximation as:

$$L_v[f](x) = f(v) + \nabla f(v)^\top (x - v)$$

This approximation is valid when:

$$x \approx v$$

Derivation via Coordinate-wise Approximation (2D Case)

Let $f : \mathbb{R}^2 \to \mathbb{R}$.
We approximate first in the x_1 direction:

$$f(y_1, v_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_1}(v)(y_1 - v_1)$$

Then in the x_2 direction:

$$f(v_1, y_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_2}(v)(y_2 - v_2)$$

If both coordinates change simultaneously, we add contributions:

$$f(y_1, y_2) \approx f(v_1, v_2) + \frac{\partial f}{\partial x_1}(v)(y_1 - v_1) + \frac{\partial f}{\partial x_2}(v)(y_2 - v_2)$$

Which compactly becomes:

$$f(y) \approx f(v) + \nabla f(v)^\top (y - v)$$

Example — Linear Approximation of a Quadratic

Let:

$$f(x_1, x_2) = x_1^2 + x_2^2$$

We approximate around:

$$v = (6, 2)$$

We compute:

$$f(v) = 6^2 + 2^2 = 40$$

Gradient:

$$\nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

So:

$$\nabla f(v) = \begin{bmatrix} 12 \\ 4 \end{bmatrix}$$

Linear approximation:

$$L_v[f](x) = 40 + \begin{bmatrix} 12 & 4 \end{bmatrix} \begin{bmatrix} x_1 - 6 \\ x_2 - 2 \end{bmatrix}$$

Simplifying:

$$\begin{aligned} L_v[f](x) &= 40 + 12(x_1 - 6) + 4(x_2 - 2) \\ &= 12x_1 + 4x_2 - 40 \end{aligned}$$

Valid when:

$$(x_1, x_2) \approx (6, 2)$$

Tangent Plane Interpretation

For $f : \mathbb{R}^d \to \mathbb{R}$, the graph lies in $\mathbb{R}^{d+1}$.
The graph of $L_v[f]$ is a plane.
That plane is tangent to the graph of f at:

$$(v, f(v))$$

Thus:
The gradient determines the tangent plane.

Contour Interpretation

Consider the level set:

$$\{x \in \mathbb{R}^d : f(x) = f(v)\}$$

We show:

$$\nabla f(v)$$

is perpendicular to this contour.
Proof via linear approximation:
Level set of linear approximation:

$$L_v[f](x) = f(v)$$

Substitute:

$$f(v) + \nabla f(v)^\top (x - v) = f(v)$$

Which gives:

$$\nabla f(v)^\top x = \nabla f(v)^\top v$$

This is the equation of a hyperplane:

$$w^\top x = b$$

Thus:

$$\nabla f(v)$$

is normal to the contour.

Directional Derivative

Directional derivative of f at v along direction u:

$$D_u f(v) = \lim_{\alpha \to 0} \frac{f(v + \alpha u) - f(v)}{\alpha}$$

Using linear approximation:

$$f(v + \alpha u) \approx f(v) + \nabla f(v)^\top (\alpha u)$$

So:

$$D_u f(v) = \nabla f(v)^\top u$$

Thus:
Directional derivative equals gradient dotted with direction.

Cauchy–Schwarz Inequality

For vectors $a, b \in \mathbb{R}^d$:

$$-\|a\|\|b\| \leq a^\top b \leq \|a\|\|b\|$$

Equality holds when:

- $a = \alpha b$, $\alpha > 0$ (upper bound)
- $a = \alpha b$, $\alpha < 0$ (lower bound)

Direction of Steepest Ascent

We seek unit vector u that maximizes:

$$D_u f(v) = \nabla f(v)^\top u$$

Subject to:

$$\|u\| = 1$$

By Cauchy–Schwarz:
Maximum occurs when:

$$u = \frac{\nabla f(v)}{\|\nabla f(v)\|}$$

Thus:
Gradient gives direction of steepest ascent.
Steepest descent direction:

$$-\nabla f(v)$$

Descent Directions

Set of descent directions:

$$\{u \in \mathbb{R}^d : \nabla f(v)^\top u < 0\}$$

These directions reduce function value.

Higher-Order Approximation

Linear approximation:

$$f(x) \approx f(v) + \nabla f(v)^\top (x - v)$$

Quadratic approximation:

$$f(x) \approx f(v) + \nabla f(v)^\top (x - v) + \frac{1}{2}(x - v)^\top \nabla^2 f(v)(x - v)$$

Where:

$$\nabla^2 f(v)$$

is the Hessian matrix (a $d \times d$ matrix).

Critical Points

If f is minimized at v, then:

$$\nabla f(v) = 0$$

Similarly for maxima.
Points satisfying:

$$\nabla f(v) = 0$$

are called **critical points**.
This condition is called the **first-order necessary condition for optimality**.
Critical points may be:

- Local minima
- Local maxima
- Saddle points

```
********************************************************************************
```
