02: Differentiating With Respect to Anything (The Total Derivative)

Chapter Goal: To use the Partial Derivatives we just learned and combine them with the Chain Rule to create a powerful new concept: the **Total Derivative**.

---

1. Background: Sharpening Partial Derivative Skills

The video begins with a slightly more complex partial derivative example to train our "muscles".

- **Function:** $f(x, y, z) = \sin(x)e^{yz^2}$
- **Thought Process (always the same):** When differentiating with respect to one variable, "freeze" all the others.
- $\frac{\partial f}{\partial x}$:
  - e^{yz^2} is treated as a constant. The derivative of $\sin(x)$ is $\cos(x)$.
  - **Result:** $\frac{\partial f}{\partial x} = \cos(x)e^{yz^2}$
- $\frac{\partial f}{\partial y}$:
  - $\sin(x)$ is treated as a constant. The derivative of e^{stuff} is e^{stuff} times the derivative of the "stuff" (Chain Rule). The derivative of yz^2 with respect to `y` is z^2.
  - **Result:** $\frac{\partial f}{\partial y} = \sin(x)e^{yz^2} \cdot z^2$
- $\frac{\partial f}{\partial z}$:
  - Similar to `y`. The derivative of yz^2 with respect to `z` is $y \cdot 2z = 2yz$.
  - **Result:** $\frac{\partial f}{\partial z} = \sin(x)e^{yz^2} \cdot 2yz$

---

2. The New Idea: What If All Inputs are Interconnected?

- **New Scenario:** Imagine if `x`, `y`, and `z` are not actually independent variables. Imagine they all depend on **one single underlying variable**, let's say time (`t`).
  - $x(t) = t - 1$
  - $y(t) = t^2$
  - $z(t) = 1/t$
- **Meaning:** `x`, `y`, and `z` are actually the coordinates of a particle moving through space over time `t`.
- **The Key Question:**

> "How fast does the value of the function `f` change as time `t` passes?"

- We want to find the **total derivative**, $\frac{df}{dt}$.

---

3. "Aha!" Moment: The Total Derivative as an Extension of the Chain Rule

- **The Problem:** Finding $\frac{df}{dt}$ could be very messy if we substitute all the `t` expressions into `f` first.
- **The Elegant Solution (Total Derivative):** We can use the "domino effect" logic from the Chain Rule. The total change in `f` (`df`) is the sum of all the changes caused by each intermediate variable (`x` , `y` , `z`).

> Total change in `f` =
> (Change in `f` caused by `t` through the `x` path)
>
> - (Change in `f` caused by `t` through the `y` path)
> - (Change in `f` caused by `t` through the `z` path)

- **The Total Derivative Formula (Multivariate Chain Rule):**
  If we divide everything by `dt` :

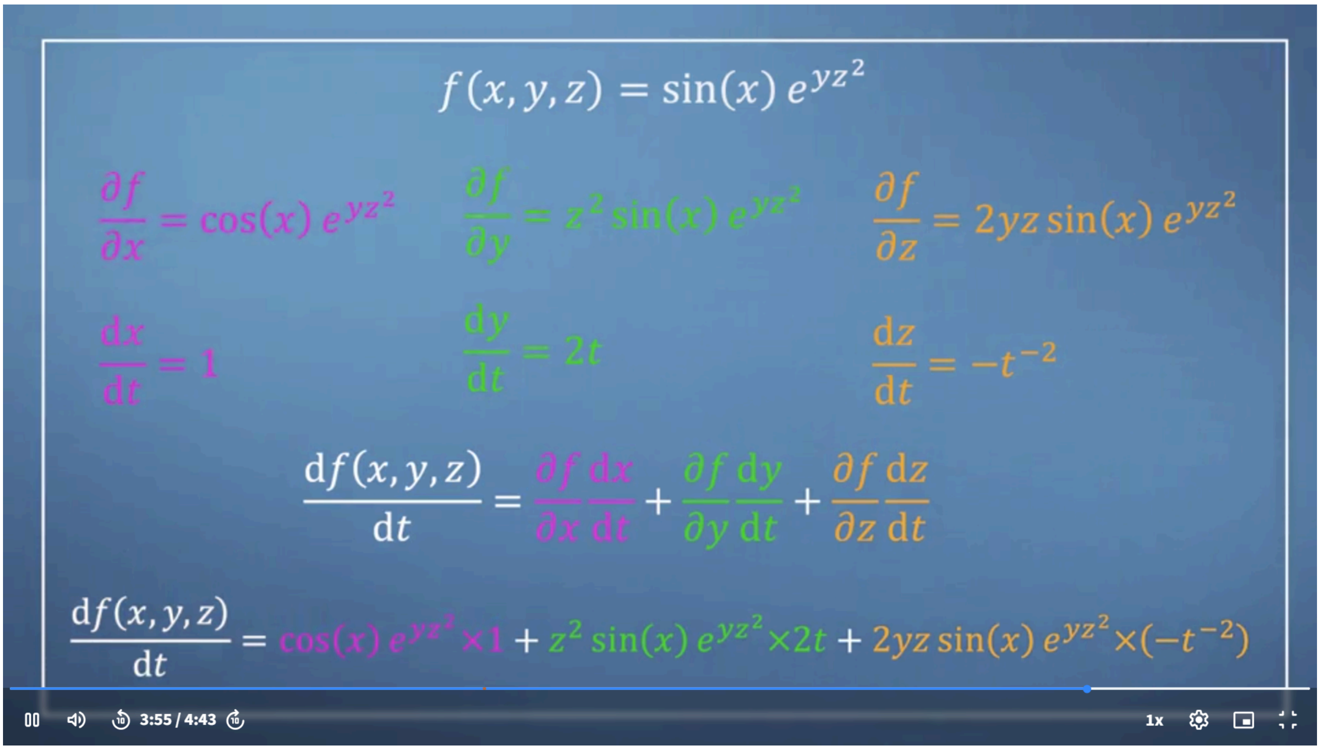
> $$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}$$

- **Leibniz Notation Intuition:**
  Visually, the intermediate terms (`dx` , `dy` , `dz`) look like they can "cancel out", leaving `df/dt` in all paths.

$$\frac{df}{dt} = \frac{\partial f}{\partial \cancel{x}}\frac{\cancel{dx}}{dt} + \frac{\partial f}{\partial \cancel{y}}\frac{\cancel{dy}}{dt} + \frac{\partial f}{\partial \cancel{z}}\frac{\cancel{dz}}{dt}$$

---

4. Applying the Formula to the Example



Now we just need to collect all the "ingredients" we need:

- **Partial Derivatives (already calculated):** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$.
- **Ordinary Derivatives (easy to calculate):**
  - $\frac{dx}{dt} = 1$
  - $\frac{dy}{dt} = 2t$
  - $\frac{dz}{dt} = -t^{-2}$
- **Combine Everything:** Plug all these ingredients into the Total Derivative formula.

$$\frac{df}{dt} = (\cos(x)e^{yz^2}) \cdot (1) + (z^2 \sin(x)e^{yz^2}) \cdot (2t) + (2yz \sin(x)e^{yz^2}) \cdot (-t^{-2})$$

- Although the result looks like a "monster", the video shows that after substituting `x, y, z` back in terms of `t` and simplifying, the result becomes very neat.

$$f(x,y,z) = \sin(x)\, e^{yz^2}$$

$$\frac{df(x,y,z)}{dt} = \cos(x)\, e^{yz^2} \times 1 + z^2 \sin(x)\, e^{yz^2} \times 2t + 2yz \sin(x)\, e^{yz^2} \times (-t^{-2})$$

$$x = t - 1\,; \quad y = t^2; \quad z = \frac{1}{t}$$

$$\frac{df(x,y,z)}{dt} = \cos(t-1)\, e + t^{-2} \sin(t-1)\, e \times 2t + 2t \sin(t-1)\, e \times (-t^{-2})$$

$$\frac{df(x,y,z)}{dt} = \cos(t-1)\, e + 2t^{-1} \sin(t-1)\, e - 2t^{-1} \sin(t-1)\, e$$

$$\frac{df(x,y,z)}{dt} = \cos(t-1)\, e$$

---

5. Key Message

- The **Total Derivative** is the "super version" of the Chain Rule for functions with multiple inputs that all depend on a single base variable.
- It is a very powerful tool because it allows us to calculate a total rate of change by breaking it down into the contributions from each intermediate path, which are often easier to calculate separately.
- This is the fundamental concept behind **Backpropagation** in Neural Networks.

---

6. Concrete Analogy: Measuring "Stress" While Mountain Climbing

The concept of a "Total Derivative" can feel very abstract if not connected to a real-world problem. Let's use a new, more "feelable" analogy to understand its usefulness.

Imagine you are a mountain climber. Your level of **"Stress" (`S`)** is a function that depends on three things:

> `Stress = f(Altitude, Temperature, Wind_Speed)`

Now, imagine you are climbing. As time (`t`) passes, all three of these variables are changing:

- **Altitude (`h`)** is constantly increasing $\to h(t)$
- **Temperature (`T`)** is constantly decreasing $\to T(t)$

- **Wind Speed (`v`)** is fluctuating unpredictably $\rightarrow v(t)$

The Key Question:

> "How fast is my Stress level changing **right now**?"

We want to find $\frac{dS}{dt}$ (the **Total Derivative** of Stress).

Why Do We Need the Total Derivative?

The change in your Stress is a combination of three effects happening simultaneously:

1. The effect of you getting higher (air getting thinner).
2. The effect of the temperature getting colder.
3. The effect of the wind changing.

The Total Derivative is a way to calculate this combined impact by analyzing each "path" separately.
`dS/dt = (Change in Stress due to Altitude) + (Change in Stress due to Temp) + (Change in Stress due to Wind)`

Thinking with Partial Derivatives (The Chain Rule)

- **Path 1: The Altitude Effect**
  - "How fast does Stress change with respect to Altitude?" $\rightarrow$ This is $\frac{\partial S}{\partial h}$ (a partial derivative). It's your body's "sensitivity" to altitude.
  - "How fast is Altitude changing with respect to time?" $\rightarrow$ This is $\frac{dh}{dt}$ (your vertical climbing speed).
  - **Contribution from the Altitude Path:** $\frac{\partial S}{\partial h} \cdot \frac{dh}{dt}$
- **Path 2: The Temperature Effect**
  - "How fast does Stress change with respect to Temperature?" $\rightarrow \frac{\partial S}{\partial T}$.
  - "How fast is Temperature changing with respect to time?" $\rightarrow \frac{dT}{dt}$.
  - **Contribution from the Temperature Path:** $\frac{\partial S}{\partial T} \cdot \frac{dT}{dt}$
- **Path 3: The Wind Effect**
  - "How fast does Stress change with respect to Wind Speed?" $\rightarrow \frac{\partial S}{\partial v}$.
  - "How fast is Wind Speed changing with respect to time?" $\rightarrow \frac{dv}{dt}$.
  - **Contribution from the Wind Path:** $\frac{\partial S}{\partial v} \cdot \frac{dv}{dt}$

Putting It All Together

The Total Derivative formula tells us that to get the total rate of change of Stress, we just sum the contributions from each path:

$$\frac{dS}{dt} = \frac{\partial S}{\partial h}\frac{dh}{dt} + \frac{\partial S}{\partial T}\frac{dT}{dt} + \frac{\partial S}{\partial v}\frac{dv}{dt}$$

What is This Useful For?

- **Problem Decomposition:** It allows us to break down a very complex problem ("how does stress change?") into several simpler sub-problems that are easier to measure:
  1. Measure the body's sensitivity to each factor ($\frac{\partial S}{\partial h}$, etc.).
  2. Measure how fast each factor is changing ($\frac{dh}{dt}$, etc.).
- **Contribution Analysis:** By looking at each term in the sum, we can analyze which factor is having the biggest impact.
  - Maybe your climbing speed $\frac{dh}{dt}$ is small, but your sensitivity to altitude $\frac{\partial S}{\partial h}$ is huge. So, even though you're climbing slowly, your stress is rising dramatically.
  - Maybe the temperature change $\frac{dT}{dt}$ is huge (a storm suddenly hits), making the second term the one that dominates your change in stress.

Application in Machine Learning (Backpropagation):

- **"Stress"** is the **"Error"** or **"Loss"** of the model.
- **"Altitude, Temperature, Wind"** are the **weights** in the layers of a Neural Network.
- **"Time"** is the training process.

Backpropagation is essentially the application of the Total Derivative (Multivariate Chain Rule) to calculate $\frac{\partial \text{Loss}}{\partial \text{Weight}}$ for every single weight in a very deep network. It tells us "how much each tiny knob contributed to the final total error," so we know which knobs to turn and by how much.

Conclusion:
The Total Derivative is a fundamental tool for understanding how an interconnected system changes over time. It gives us a way to track and measure the contribution of each component to the total change.

Tags: #mml-specialization #multivariate-calculus #partial-derivatives #total-derivative #chain-rule #backpropagation